Thierry Roncalli

Handbook of Financial Risk Management

Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

Handbook of Financial Risk Management

CHAPMAN & HALL/CRC
Financial Mathematics Series

Aims and scope:

The field of financial mathematics forms an ever-expanding slice of the financial sector. This series aims to capture new developments and summarize what is known over the whole spectrum of this field. It will include a broad range of textbooks, reference works and handbooks that are meant to appeal to both academics and practitioners. The inclusion of numerical code and concrete real-world examples is highly encouraged.

Series Editors

M.A.H. Dempster
Centre for Financial Research
Department of Pure Mathematics and Statistics
University of Cambridge

Dilip B. Madan
Robert H. Smith School of Business
University of Maryland

Rama Cont
Department of Mathematics
Imperial College

Interest Rate Modeling

Theory and Practice, 2nd Edition
Lixin Wu

Metamodeling for Variable Annuities

Guojun Gan and Emiliano A. Valdez

Modeling Fixed Income Securities and Interest Rate Options

Robert A. Jarrow

Financial Modelling in Commodity Markets

Viviana Fanelli

Introductory Mathematical Analysis for Quantitative Finance

Daniele Ritelli, Giulia Spaletta

Handbook of Financial Risk Management

Thierry Roncalli

Optional Processes

Stochastic Calculus and Applications
Mohamed Abdelghani, Alexander Melnikov

For more information about this series please visit: *https://www.crcpress.com/Chapman-and-HallCRC-Financial-Mathematics-Series/book-series/CHFINANCMTH*

Handbook of Financial Risk Management

Thierry Roncalli
University of Evry


CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-138-50187-4 (Hardback)

Library of Congress Cataloging-in-Publication Data

Names: Roncalli, Thierry, author.
Title: Handbook of financial risk management / Thierry Roncalli.
Description: Boca Raton : CRC Press, 2020. | Series: Chapman and Hall/CRC financial mathematics series | Includes bibliographical references and index.
Identifiers: LCCN 2019059891 | ISBN 9781138501874 (hardback) | ISBN 9781315144597 (ebook)
Subjects: LCSH: Financial risk management. | Risk management--Mathematical models.
Classification: LCC HD61 .R663 2020 | DDC 658.15/5--dc23
LC record available at https://lccn.loc.gov/2019059891

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface xxi

List of Symbols and Notations xxv

1 Introduction 1
1.1 The need for risk management . . . 1
1.1.1 Risk management and the financial system . . . 1
1.1.2 The development of financial markets . . . 3
1.1.3 Financial crises and systemic risk . . . 7
1.2 Financial regulation . . . 12
1.2.1 Banking regulation . . . 13
1.2.2 Insurance regulation . . . 22
1.2.3 Market regulation . . . 25
1.2.4 Systemic risk . . . 26
1.3 Financial regulation overview . . . 28
1.3.1 List of supervisory authorities . . . 28
1.3.2 Timeline of financial regulation . . . 29

I Risk Management in the Financial Sector 35

2 Market Risk 37
2.1 Regulatory framework . . . 37
2.1.1 The Basel I/II framework . . . 38
2.1.1.1 Standardized measurement method . . . 38
2.1.1.2 Internal model-based approach . . . 45
2.1.2 The Basel III framework . . . 52
2.1.2.1 Standardized approach . . . 53
2.1.2.2 Internal model-based approach . . . 57
2.2 Statistical estimation methods of risk measures . . . 61
2.2.1 Definition . . . 61
2.2.1.1 Coherent risk measures . . . 61
2.2.1.2 Value-at-risk . . . 64
2.2.1.3 Expected shortfall . . . 65
2.2.1.4 Estimator or estimate? . . . 65
2.2.2 Historical methods . . . 66
2.2.2.1 The order statistic approach . . . 67
2.2.2.2 The kernel approach . . . 71
2.2.3 Analytical methods . . . 72
2.2.3.1 Derivation of the closed-form formula . . . 72
2.2.3.2 Linear factor models . . . 75
2.2.3.3 Volatility forecasting . . . 80
2.2.3.4 Extension to other probability distributions . . . 84

2.2.4 Monte Carlo methods . . . 90
2.2.5 The case of options and derivatives . . . 92
2.2.5.1 Identification of risk factors . . . 93
2.2.5.2 Methods to calculate VaR and ES risk measures . . . 94
2.2.5.3 Backtesting . . . 102
2.2.5.4 Model risk . . . 102
2.3 Risk allocation . . . 104
2.3.1 Euler allocation principle . . . 105
2.3.2 Application to non-normal risk measures . . . 108
2.3.2.1 Main results . . . 108
2.3.2.2 Calculating risk contributions with historical and simulated scenarios . . . 109
2.4 Exercises . . . 116
2.4.1 Calculating regulatory capital with the Basel I standardized measurement method . . . 116
2.4.2 Covariance matrix . . . 117
2.4.3 Risk measure . . . 118
2.4.4 Value-at-risk of a long/short portfolio . . . 119
2.4.5 Value-at-risk of an equity portfolio hedged with put options . . . 119
2.4.6 Risk management of exotic options . . . 120
2.4.7 P&L approximation with Greek sensitivities . . . 121
2.4.8 Calculating the non-linear quadratic value-at-risk . . . 121
2.4.9 Risk decomposition of the expected shortfall . . . 123
2.4.10 Expected shortfall of an equity portfolio . . . 123
2.4.11 Risk measure of a long/short portfolio . . . 123
2.4.12 Kernel estimation of the expected shortfall . . . 124

3 Credit Risk 125
3.1 The market of credit risk . . . 125
3.1.1 The loan market . . . 125
3.1.2 The bond market . . . 126
3.1.2.1 Statistics of the bond market . . . 128
3.1.2.2 Bond pricing . . . 131
3.1.3 Securitization and credit derivatives . . . 137
3.1.3.1 Credit securitization . . . 137
3.1.3.2 Credit default swap . . . 141
3.1.3.3 Basket default swap . . . 151
3.1.3.4 Collateralized debt obligations . . . 155
3.2 Capital requirement . . . 159
3.2.1 The Basel I framework . . . 160
3.2.2 The Basel II standardized approach . . . 162
3.2.2.1 Standardized risk weights . . . 162
3.2.2.2 Credit risk mitigation . . . 165
3.2.3 The Basel II internal ratings-based approach . . . 168
3.2.3.1 The general framework . . . 168
3.2.3.2 The credit risk model of Basel II . . . 169
3.2.3.3 The IRB formulas . . . 176
3.2.4 The Basel III revision . . . 181
3.2.4.1 The standardized approach . . . 181
3.2.4.2 The internal ratings-based approach . . . 184
3.2.5 The securitization framework . . . 185

3.2.5.1 Overview of the approaches . . . 185
3.2.5.2 Internal ratings-based approach . . . 186
3.2.5.3 External ratings-based approach . . . 188
3.2.5.4 Standardized approach . . . 188
3.3 Credit risk modeling . . . 190
3.3.1 Exposure at default . . . 190
3.3.2 Loss given default . . . 191
3.3.2.1 Definition . . . 191
3.3.2.2 Stochastic modeling . . . 193
3.3.2.3 Economic modeling . . . 200
3.3.3 Probability of default . . . 201
3.3.3.1 Survival function . . . 201
3.3.3.2 Transition probability matrix . . . 206
3.3.3.3 Structural models . . . 214
3.3.4 Default correlation . . . 220
3.3.4.1 The copula model . . . 220
3.3.4.2 The factor model . . . 225
3.3.4.3 Estimation methods . . . 227
3.3.4.4 Dependence and credit basket derivatives . . . 234
3.3.5 Granularity and concentration . . . 241
3.3.5.1 Difference between fine-grained and concentrated portfolios 241
3.3.5.2 Granularity adjustment . . . 245
3.4 Exercises . . . 247
3.4.1 Single- and multi-name credit default swaps . . . 247
3.4.2 Risk contribution in the Basel II model . . . 248
3.4.3 Calibration of the piecewise exponential model . . . 249
3.4.4 Modeling loss given default . . . 250
3.4.5 Modeling default times with a Markov chain . . . 251
3.4.6 Continuous-time modeling of default risk . . . 252
3.4.7 Derivation of the original Basel granularity adjustment . . . 253
3.4.8 Variance of the conditional portfolio loss . . . 255

4 Counterparty Credit Risk and Collateral Risk 257
4.1 Counterparty credit risk . . . 257
4.1.1 Definition . . . 258
4.1.2 Modeling the exposure at default . . . 260
4.1.2.1 An illustrative example . . . 260
4.1.2.2 Measuring the counterparty exposure . . . 264
4.1.2.3 Practical implementation for calculating counterparty exposure . . . 265
4.1.3 Regulatory capital . . . 266
4.1.3.1 Internal model method . . . 268
4.1.3.2 Non-internal model methods (Basel II) . . . 269
4.1.3.3 SA-CCR method (Basel III) . . . 270
4.1.4 Impact of wrong way risk . . . 275
4.1.4.1 An example . . . 275
4.1.4.2 Calibration of the α factor . . . 276
4.2 Credit valuation adjustment . . . 278
4.2.1 Definition . . . 279
4.2.1.1 Difference between CCR and CVA . . . 279
4.2.1.2 CVA, DVA and bilateral CVA . . . 279

4.2.1.3 Practical implementation for calculating CVA 282
4.2.2 Regulatory capital 283
4.2.2.1 The 2010 version of Basel III 284
4.2.2.2 The 2017 version of Basel III 285
4.2.3 CVA and wrong/right way risk 289
4.3 Collateral risk 293
4.3.1 Definition 293
4.3.2 Capital allocation 294
4.4 Exercises 300
4.4.1 Impact of netting agreements in counterparty credit risk 300
4.4.2 Calculation of the effective expected positive exposure 301
4.4.3 Calculation of the capital charge for counterparty credit risk 302
4.4.4 Calculation of CVA and DVA measures 302
4.4.5 Approximation of the CVA for an interest rate swap 303
4.4.6 Risk contribution of CVA with collateral 303

5 Operational Risk **305**
5.1 Definition of operational risk 305
5.2 Basel approaches for calculating the regulatory capital 307
5.2.1 The basic indicator approach 308
5.2.2 The standardized approach 308
5.2.3 Advanced measurement approaches 310
5.2.4 Basel III approach 311
5.3 Loss distribution approach 312
5.3.1 Definition 312
5.3.2 Parametric estimation 315
5.3.2.1 Estimation of the loss severity distribution 315
5.3.2.2 Estimation of the loss frequency distribution 321
5.3.3 Calculating the capital charge 327
5.3.3.1 Monte Carlo approach 327
5.3.3.2 Analytical approaches 331
5.3.3.3 Aggregation issues 336
5.3.4 Incorporating scenario analysis 339
5.3.4.1 Probability distribution of a given scenario 339
5.3.4.2 Calibration of a set of scenarios 340
5.3.5 Stability issue of the LDA model 342
5.4 Exercises 342
5.4.1 Estimation of the loss severity distribution 342
5.4.2 Estimation of the loss frequency distribution 343
5.4.3 Using the method of moments in operational risk models 344
5.4.4 Calculation of the Basel II required capital 345
5.4.5 Parametric estimation of the loss severity distribution 345
5.4.6 Mixed Poisson process 346

6 Liquidity Risk **347**
6.1 Market liquidity 347
6.1.1 Transaction cost versus volume-based measures 348
6.1.1.1 Bid-ask spread 348
6.1.1.2 Trading volume 349
6.1.1.3 Liquidation ratio 350
6.1.1.4 Liquidity ordering 352

6.1.2 Other liquidity measures . . . 353
6.1.3 The liquidity-adjusted CAPM . . . 355
6.2 Funding liquidity . . . 357
6.2.1 Asset liability mismatch . . . 357
6.2.2 Relationship between market and funding liquidity risks . . . 358
6.3 Regulation of the liquidity risk . . . 360
6.3.1 Liquidity coverage ratio . . . 360
6.3.1.1 Definition . . . 360
6.3.1.2 Monitoring tools . . . 364
6.3.2 Net stable funding ratio . . . 365
6.3.2.1 Definition . . . 365
6.3.2.2 ASF and RSF factors . . . 366
6.3.3 Leverage ratio . . . 367

7 Asset Liability Management Risk 369
7.1 General principles of the banking book risk management . . . 369
7.1.1 Definition . . . 370
7.1.1.1 Balance sheet and income statement . . . 370
7.1.1.2 Accounting standards . . . 373
7.1.1.3 Role and importance of the ALCO . . . 375
7.1.2 Liquidity risk . . . 376
7.1.2.1 Definition of the liquidity gap . . . 376
7.1.2.2 Asset and liability amortization . . . 377
7.1.2.3 Dynamic analysis . . . 385
7.1.2.4 Liquidity hedging . . . 392
7.1.3 Interest rate risk in the banking book . . . 393
7.1.3.1 Introduction on IRRBB . . . 393
7.1.3.2 Interest rate risk principles . . . 396
7.1.3.3 The standardized approach . . . 396
7.1.4 Other ALM risks . . . 402
7.1.4.1 Currency risk . . . 402
7.1.4.2 Credit spread risk . . . 403
7.2 Interest rate risk . . . 404
7.2.1 Duration gap analysis . . . 405
7.2.1.1 Relationship between Macaulay duration and modified duration . . . 405
7.2.1.2 Relationship between the duration gap and the equity duration . . . 407
7.2.1.3 An illustration . . . 408
7.2.1.4 Immunization of the balance sheet . . . 409
7.2.2 Earnings-at-risk . . . 410
7.2.2.1 Income gap analysis . . . 410
7.2.2.2 Net interest income . . . 412
7.2.2.3 Hedging strategies . . . 419
7.2.3 Simulation approach . . . 420
7.2.4 Funds transfer pricing . . . 421
7.2.4.1 Net interest and commercial margins . . . 422
7.2.4.2 Computing the internal transfer rates . . . 425
7.3 Behavioral options . . . 427
7.3.1 Non-maturity deposits . . . 427
7.3.1.1 Static and dynamic modeling . . . 428

7.3.1.2 Behavioral modeling . . . 431
7.3.2 Prepayment risk . . . 437
7.3.2.1 Factors of prepayment . . . 438
7.3.2.2 Structural models . . . 439
7.3.2.3 Reduced-form models . . . 440
7.3.2.4 Statistical measure of prepayment . . . 446
7.3.3 Redemption risk . . . 447
7.3.3.1 The funding risk of term deposits . . . 447
7.3.3.2 Modeling the early withdrawal risk . . . 448
7.4 Exercises . . . 449
7.4.1 Constant amortization of a loan . . . 449
7.4.2 Computation of the amortization functions $\mathbf{S}(t,u)$ and $\mathbf{S}^{\star}(t,u)$ . . . 450
7.4.3 Continuous-time analysis of the constant amortization mortgage (CAM) . . . 450
7.4.4 Valuation of non-maturity deposits . . . 451
7.4.5 Impact of prepayment on the amortization scheme of the CAM . . . 452

8 Systemic Risk and Shadow Banking System 453
8.1 Defining systemic risk . . . 453
8.1.1 Systemic risk, systematic risk and idiosyncratic risk . . . 454
8.1.2 Sources of systemic risk . . . 456
8.1.2.1 Systematic shocks . . . 456
8.1.2.2 Propagation mechanisms . . . 458
8.1.3 Supervisory policy responses . . . 459
8.1.3.1 A new financial regulatory structure . . . 460
8.1.3.2 A myriad of new standards . . . 461
8.2 Systemic risk measurement . . . 463
8.2.1 The supervisory approach . . . 463
8.2.1.1 The G-SIB assessment methodology . . . 463
8.2.1.2 Identification of G-SIIs . . . 466
8.2.1.3 Extension to NBNI SIFIs . . . 466
8.2.2 The academic approach . . . 467
8.2.2.1 Marginal expected shortfall . . . 467
8.2.2.2 Delta conditional value-at-risk . . . 469
8.2.2.3 Systemic risk measure . . . 471
8.2.2.4 Network measures . . . 474
8.3 Shadow banking system . . . 478
8.3.1 Definition . . . 478
8.3.2 Measuring the shadow banking . . . 478
8.3.2.1 The broad (or MUNFI) measure . . . 478
8.3.2.2 The narrow measure . . . 481
8.3.3 Regulatory developments of shadow banking . . . 485
8.3.3.1 Data gaps . . . 485
8.3.3.2 Mitigation of interconnectedness risk . . . 486
8.3.3.3 Money market funds . . . 486
8.3.3.4 Complex shadow banking activities . . . 487

II Mathematical and Statistical Tools 489

9 Model Risk of Exotic Derivatives 491
9.1 Basics of option pricing . . . 491
9.1.1 The Black-Scholes model . . . 491
9.1.1.1 The general framework . . . 491
9.1.1.2 Application to European options . . . 492
9.1.1.3 Principle of dynamic hedging . . . 495
9.1.1.4 The implied volatility . . . 508
9.1.2 Interest rate risk modeling . . . 513
9.1.2.1 Pricing zero-coupon bonds with the Vasicek model . . . 514
9.1.2.2 The calibration issue of the yield curve . . . 516
9.1.2.3 Caps, floors and swaptions . . . 518
9.1.2.4 Change of numéraire and equivalent martingale measure . 519
9.1.2.5 The HJM model . . . 522
9.1.2.6 Market models . . . 525
9.2 Volatility risk . . . 529
9.2.1 The uncertain volatility model . . . 530
9.2.1.1 Formulation of the partial differential equation . . . 530
9.2.1.2 Computing lower and upper pricing bounds . . . 531
9.2.1.3 Application to ratchet options . . . 532
9.2.2 The shifted log-normal model . . . 535
9.2.2.1 The fixed-strike parametrization . . . 537
9.2.2.2 The floating-strike parametrization . . . 538
9.2.2.3 The forward parametrization . . . 539
9.2.2.4 Mixture of SLN distributions . . . 540
9.2.2.5 Application to binary, corridor and barrier options . . . 541
9.2.3 Local volatility model . . . 546
9.2.3.1 Derivation of the forward equation . . . 546
9.2.3.2 Duality between local volatility and implied volatility . . . 549
9.2.3.3 Dupire model in practice . . . 551
9.2.3.4 Application to exotic options . . . 559
9.2.4 Stochastic volatility models . . . 560
9.2.4.1 General analysis . . . 560
9.2.4.2 Heston model . . . 562
9.2.4.3 SABR model . . . 568
9.2.5 Factor models . . . 575
9.2.5.1 Linear and quadratic Gaussian models . . . 575
9.2.5.2 Dynamics of risk factors under the forward probability measure . . . 577
9.2.5.3 Pricing caplets and swaptions . . . 577
9.2.5.4 Calibration and practice of factor models . . . 578
9.3 Other model risk topics . . . 580
9.3.1 Dividend risk . . . 581
9.3.1.1 Understanding the impact of dividends on option prices . 581
9.3.1.2 Models of discrete dividends . . . 582
9.3.2 Correlation risk . . . 583
9.3.2.1 The two-asset case . . . 583
9.3.2.2 The multi-asset case . . . 587
9.3.2.3 The copula method . . . 589
9.3.3 Liquidity risk . . . 591

9.4 Exercises . . . 593
9.4.1 Option pricing and martingale measure . . . 593
9.4.2 The Vasicek model . . . 593
9.4.3 The Black model . . . 594
9.4.4 Change of numéraire and Girsanov theorem . . . 594
9.4.5 The HJM model and the forward probability measure . . . 596
9.4.6 Equivalent martingale measure in the Libor market model . . . 597
9.4.7 Displaced diffusion option pricing . . . 598
9.4.8 Dupire local volatility model . . . 599
9.4.9 The stochastic normal model . . . 599
9.4.10 The quadratic Gaussian model . . . 601
9.4.11 Pricing two-asset basket options . . . 602

10 Statistical Inference and Model Estimation 603
10.1 Estimation methods . . . 603
10.1.1 Linear regression . . . 603
10.1.1.1 Least squares estimation . . . 604
10.1.1.2 Relationship with the conditional normal distribution . . . 608
10.1.1.3 The intercept problem . . . 610
10.1.1.4 Coefficient of determination . . . 611
10.1.1.5 Extension to weighted least squares regression . . . 612
10.1.2 Maximum likelihood estimation . . . 614
10.1.2.1 Definition of the estimator . . . 614
10.1.2.2 Asymptotic distribution . . . 616
10.1.2.3 Statistical inference . . . 618
10.1.2.4 Some examples . . . 620
10.1.2.5 EM algorithm . . . 621
10.1.3 Generalized method of moments . . . 628
10.1.3.1 Method of moments . . . 628
10.1.3.2 Extension to the GMM approach . . . 631
10.1.3.3 Simulated method of moments . . . 635
10.1.4 Non-parametric estimation . . . 637
10.1.4.1 Non-parametric density estimation . . . 637
10.1.4.2 Non-parametric regression . . . 641
10.2 Time series modeling . . . 644
10.2.1 ARMA process . . . 644
10.2.1.1 The VAR(1) process . . . 644
10.2.1.2 Extension to ARMA models . . . 646
10.2.2 State space models . . . 647
10.2.2.1 Specification and estimation of state space models . . . 647
10.2.2.2 Some applications . . . 650
10.2.3 Cointegration and error correction models . . . 655
10.2.3.1 Nonstationarity and spurious regression . . . 655
10.2.3.2 The concept of cointegration . . . 657
10.2.3.3 Error correction model . . . 659
10.2.3.4 Estimation of cointegration relationships . . . 660
10.2.4 GARCH and stochastic volatility models . . . 664
10.2.4.1 GARCH models . . . 664
10.2.4.2 Stochastic volatility models . . . 667
10.2.5 Spectral analysis . . . 670
10.2.5.1 Fourier analysis . . . 670

10.2.5.2 Definition of the spectral density function . . . 672
10.2.5.3 Frequency domain localization . . . 673
10.2.5.4 Main properties . . . 675
10.2.5.5 Statistical estimation in the frequency domain . . . 683
10.2.5.6 Extension to multidimensional processes . . . 687
10.2.5.7 Some applications . . . 692
10.3 Exercises . . . 704
10.3.1 Probability distribution of the t-statistic in the case of the linear regression model . . . 704
10.3.2 Linear regression without a constant . . . 705
10.3.3 Linear regression with linear constraints . . . 705
10.3.4 Maximum likelihood estimation of the Poisson distribution . . . 706
10.3.5 Maximum likelihood estimation of the exponential distribution . . . 707
10.3.6 Relationship between the linear regression and the maximum likelihood method . . . 707
10.3.7 The Gaussian mixture model . . . 707
10.3.8 Parameter estimation of diffusion processes . . . 707
10.3.9 The Tobit model . . . 708
10.3.10 Derivation of Kalman filter equations . . . 709
10.3.11 Steady state of time-invariant state space model . . . 710
10.3.12 Kalman information filter versus Kalman covariance filter . . . 711
10.3.13 Granger representation theorem . . . 712
10.3.14 Probability distribution of the periodogram . . . 712
10.3.15 Spectral density function of structural time series models . . . 713
10.3.16 Spectral density function of some processes . . . 714

11 Copulas and Dependence Modeling 715
11.1 Canonical representation of multivariate distributions . . . 715
11.1.1 Sklar's theorem . . . 715
11.1.2 Expression of the copula density . . . 716
11.1.3 Fréchet classes . . . 717
11.1.3.1 The bivariate case . . . 718
11.1.3.2 The multivariate case . . . 719
11.1.3.3 Concordance ordering . . . 719
11.2 Copula functions and random vectors . . . 722
11.2.1 Countermonotonicity, comonotonicity and scale invariance property 722
11.2.2 Dependence measures . . . 724
11.2.2.1 Concordance measures . . . 724
11.2.2.2 Linear correlation . . . 727
11.2.2.3 Tail dependence . . . 730
11.3 Parametric copula functions . . . 731
11.3.1 Archimedean copulas . . . 732
11.3.1.1 Definition . . . 732
11.3.1.2 Properties . . . 733
11.3.1.3 Two-parameter Archimedean copulas . . . 734
11.3.1.4 Extension to the multivariate case . . . 734
11.3.2 Normal copula . . . 735
11.3.3 Student's t copula . . . 737
11.4 Statistical inference and estimation of copula functions . . . 741
11.4.1 The empirical copula . . . 741
11.4.2 The method of moments . . . 743

11.4.3 The method of maximum likelihood . . . 745
11.5 Exercises . . . 748
11.5.1 Gumbel logistic copula . . . 748
11.5.2 Farlie-Gumbel-Morgenstern copula . . . 748
11.5.3 Survival copula . . . 749
11.5.4 Method of moments . . . 749
11.5.5 Correlated loss given default rates . . . 749
11.5.6 Calculation of correlation bounds . . . 750
11.5.7 The bivariate Pareto copula . . . 751

12 Extreme Value Theory **753**
12.1 Order statistics . . . 753
12.1.1 Main properties . . . 753
12.1.2 Extreme order statistics . . . 755
12.1.3 Inference statistics . . . 758
12.1.4 Extension to dependent random variables . . . 760
12.2 Univariate extreme value theory . . . 762
12.2.1 Fisher-Tippet theorem . . . 763
12.2.2 Maximum domain of attraction . . . 765
12.2.2.1 MDA of the Gumbel distribution . . . 767
12.2.2.2 MDA of the Fréchet distribution . . . 767
12.2.2.3 MDA of the Weibull distribution . . . 768
12.2.2.4 Main results . . . 769
12.2.3 Generalized extreme value distribution . . . 770
12.2.3.1 Definition . . . 770
12.2.3.2 Estimating the value-at-risk . . . 771
12.2.4 Peak over threshold . . . 773
12.2.4.1 Definition . . . 773
12.2.4.2 Estimating the expected shortfall . . . 775
12.3 Multivariate extreme value theory . . . 777
12.3.1 Multivariate extreme value distributions . . . 778
12.3.1.1 Extreme value copulas . . . 778
12.3.1.2 Deheuvels-Pickands representation . . . 779
12.3.2 Maximum domain of attraction . . . 781
12.3.3 Tail dependence of extreme values . . . 783
12.4 Exercises . . . 783
12.4.1 Uniform order statistics . . . 783
12.4.2 Order statistics and return period . . . 784
12.4.3 Extreme order statistics of exponential random variables . . . 784
12.4.4 Extreme value theory in the bivariate case . . . 784
12.4.5 Maximum domain of attraction in the bivariate case . . . 785

13 Monte Carlo Simulation Methods **787**
13.1 Random variate generation . . . 787
13.1.1 Generating uniform random numbers . . . 787
13.1.2 Generating non-uniform random numbers . . . 789
13.1.2.1 Method of inversion . . . 790
13.1.2.2 Method of transformation . . . 793
13.1.2.3 Rejection sampling . . . 794
13.1.2.4 Method of mixtures . . . 798
13.1.3 Generating random vectors . . . 799

13.1.3.1 Method of conditional distributions . . . 799
13.1.3.2 Method of transformation . . . 802
13.1.4 Generating random matrices . . . 807
13.1.4.1 Orthogonal and covariance matrices . . . 807
13.1.4.2 Correlation matrices . . . 809
13.1.4.3 Wishart matrices . . . 813
13.2 Simulation of stochastic processes . . . 813
13.2.1 Discrete-time stochastic processes . . . 814
13.2.1.1 Correlated Markov chains . . . 814
13.2.1.2 Time series . . . 815
13.2.2 Univariate continuous-time processes . . . 818
13.2.2.1 Brownian motion . . . 818
13.2.2.2 Geometric Brownian motion . . . 819
13.2.2.3 Ornstein-Uhlenbeck process . . . 819
13.2.2.4 Stochastic differential equations without an explicit solution . . . 820
13.2.2.5 Poisson processes . . . 824
13.2.2.6 Jump-diffusion processes . . . 827
13.2.2.7 Processes related to Brownian motion . . . 828
13.2.3 Multivariate continuous-time processes . . . 833
13.2.3.1 Multidimensional Brownian motion . . . 833
13.2.3.2 Multidimensional geometric Brownian motion . . . 833
13.2.3.3 Euler-Maruyama and Milstein schemes . . . 835
13.3 Monte Carlo methods . . . 838
13.3.1 Computing integrals . . . 838
13.3.1.1 A basic example . . . 838
13.3.1.2 Theoretical framework . . . 839
13.3.1.3 Extension to the calculation of mathematical expectations 841
13.3.2 Variance reduction . . . 844
13.3.2.1 Antithetic variates . . . 844
13.3.2.2 Control variates . . . 850
13.3.2.3 Importance sampling . . . 856
13.3.2.4 Other methods . . . 860
13.3.3 MCMC methods . . . 870
13.3.3.1 Gibbs sampling . . . 871
13.3.3.2 Metropolis-Hastings algorithm . . . 873
13.3.3.3 Sequential Monte Carlo methods and particle filters . . . 878
13.3.4 Quasi-Monte Carlo simulation methods . . . 880
13.4 Exercises . . . 885
13.4.1 Simulating random numbers using the inversion method . . . 885
13.4.2 Simulating random numbers using the transformation method . . . 886
13.4.3 Simulating random numbers using rejection sampling . . . 887
13.4.4 Simulation of Archimedean copulas . . . 889
13.4.5 Simulation of conditional random variables . . . 889
13.4.6 Simulation of the bivariate Normal copula . . . 890
13.4.7 Computing the capital charge for operational risk . . . 890
13.4.8 Simulating a Brownian bridge . . . 891
13.4.9 Optimal importance sampling . . . 891

14 Stress Testing and Scenario Analysis **893**
14.1 Stress test framework . . . 894
14.1.1 Definition . . . 894
14.1.1.1 General objective . . . 894
14.1.1.2 Scenario design and risk factors . . . 894
14.1.1.3 Firm-specific versus supervisory stress testing . . . 896
14.1.2 Methodologies . . . 899
14.1.2.1 Historical approach . . . 899
14.1.2.2 Macroeconomic approach . . . 900
14.1.2.3 Probabilistic approach . . . 903
14.2 Quantitative approaches . . . 904
14.2.1 Univariate stress scenarios . . . 904
14.2.2 Joint stress scenarios . . . 906
14.2.2.1 The bivariate case . . . 906
14.2.2.2 The multivariate case . . . 908
14.2.3 Conditional stress scenarios . . . 909
14.2.3.1 The conditional expectation solution . . . 909
14.2.3.2 The conditional quantile solution . . . 912
14.2.4 Reverse stress testing . . . 916
14.2.4.1 Mathematical computation of reverse stress testing . . . 917
14.2.4.2 Practical solutions . . . 919
14.3 Exercises . . . 919
14.3.1 Construction of a stress scenario with the GEV distribution . . . 919
14.3.2 Conditional expectation and linearity . . . 920
14.3.3 Conditional quantile and linearity . . . 920

15 Credit Scoring Models **923**
15.1 The method of scoring . . . 923
15.1.1 The emergence of credit scoring . . . 923
15.1.1.1 Judgmental credit systems versus credit scoring systems . . . 923
15.1.1.2 Scoring models for corporate bankruptcy . . . 924
15.1.1.3 New developments . . . 924
15.1.2 Variable selection . . . 925
15.1.2.1 Choice of the risk factors . . . 925
15.1.2.2 Data preparation . . . 927
15.1.2.3 Variable selection . . . 932
15.1.3 Score modeling, validation and follow-up . . . 938
15.1.3.1 Cross-validation approach . . . 938
15.1.3.2 Score modeling . . . 941
15.1.3.3 Score follow-up . . . 942
15.2 Statistical methods . . . 943
15.2.1 Unsupervised learning . . . 944
15.2.1.1 Clustering . . . 944
15.2.1.2 Dimension reduction . . . 948
15.2.2 Parametric supervised methods . . . 958
15.2.2.1 Discriminant analysis . . . 958
15.2.2.2 Binary choice models . . . 969
15.2.3 Non-parametric supervised methods . . . 974
15.2.3.1 k-nearest neighbor classifier . . . 974
15.2.3.2 Neural networks . . . 975
15.2.3.3 Support vector machines . . . 989

15.2.3.4 Model averaging . . . 1001
15.3 Performance evaluation criteria and score consistency . . . 1008
15.3.1 Shannon entropy . . . 1009
15.3.1.1 Definition and properties . . . 1009
15.3.1.2 Application to scoring . . . 1010
15.3.2 Graphical methods . . . 1011
15.3.2.1 Performance curve, selection curve and discriminant curve 1012
15.3.2.2 Some properties . . . 1013
15.3.3 Statistical methods . . . 1015
15.3.3.1 Kolmogorov-Smirnov test . . . 1015
15.3.3.2 Gini coefficient . . . 1016
15.3.3.3 Choice of the optimal cut-off . . . 1018
15.4 Exercises . . . 1021
15.4.1 Elastic net regression . . . 1021
15.4.2 Cross-validation of the ridge linear regression . . . 1022
15.4.3 K-means and the Lloyd's algorithm . . . 1023
15.4.4 Derivation of the principal component analysis . . . 1024
15.4.5 Two-class separation maximization . . . 1024
15.4.6 Maximum likelihood estimation of the probit model . . . 1025
15.4.7 Computation of feed-forward neural networks . . . 1025
15.4.8 Primal and dual problems of support vector machines . . . 1026
15.4.9 Derivation of the AdaBoost algorithm as the solution of the additive logit model . . . 1029
15.4.10 Weighted estimation . . . 1029

Conclusion **1031**

A Technical Appendix **1033**
A.1 Numerical analysis . . . 1033
A.1.1 Linear algebra . . . 1033
A.1.1.1 Eigendecomposition . . . 1033
A.1.1.2 Generalized eigendecomposition . . . 1034
A.1.1.3 Schur decomposition . . . 1034
A.1.2 Approximation methods . . . 1035
A.1.2.1 Spline functions . . . 1035
A.1.2.2 Positive definite matrix approximation . . . 1036
A.1.2.3 Numerical integration . . . 1037
A.1.2.4 Finite difference methods . . . 1041
A.1.3 Numerical optimization . . . 1046
A.1.3.1 Quadratic programming problem . . . 1046
A.1.3.2 Non-linear unconstrained optimization . . . 1047
A.1.3.3 Sequential quadratic programming algorithm . . . 1049
A.1.3.4 Dynamic programming in discrete time with finite states . 1049
A.2 Statistical and probability analysis . . . 1051
A.2.1 Probability distributions . . . 1051
A.2.1.1 The Bernoulli distribution . . . 1051
A.2.1.2 The binomial distribution . . . 1051
A.2.1.3 The geometric distribution . . . 1052
A.2.1.4 The Poisson distribution . . . 1052
A.2.1.5 The negative binomial distribution . . . 1052
A.2.1.6 The gamma distribution . . . 1052

A.2.1.7 The beta distribution . . . 1053
A.2.1.8 The noncentral chi-squared distribution . . . 1053
A.2.1.9 The exponential distribution . . . 1054
A.2.1.10 The normal distribution . . . 1054
A.2.1.11 The Student's t distribution . . . 1055
A.2.1.12 The log-normal distribution . . . 1055
A.2.1.13 The Pareto distribution . . . 1056
A.2.1.14 The generalized extreme value distribution . . . 1056
A.2.1.15 The generalized Pareto distribution . . . 1057
A.2.1.16 The skew normal distribution . . . 1057
A.2.1.17 The skew t distribution . . . 1059
A.2.1.18 The Wishart distribution . . . 1060
A.2.2 Special results . . . 1060
A.2.2.1 Affine transformation of random vectors . . . 1060
A.2.2.2 Change of variables . . . 1061
A.2.2.3 Relationship between density and quantile functions . . . 1062
A.2.2.4 Conditional expectation in the case of the normal distribution . . . 1062
A.2.2.5 Calculation of a useful integral function in credit risk models . . . 1063
A.3 Stochastic analysis . . . 1064
A.3.1 Brownian motion and Wiener process . . . 1065
A.3.2 Stochastic integral . . . 1066
A.3.3 Stochastic differential equation and Itô's lemma . . . 1067
A.3.3.1 Existence and uniqueness of a stochastic differential equation . . . 1067
A.3.3.2 Relationship with diffusion processes . . . 1068
A.3.3.3 Itô calculus . . . 1068
A.3.3.4 Extension to the multidimensional case . . . 1069
A.3.4 Feynman-Kac formula . . . 1070
A.3.5 Girsanov theorem . . . 1072
A.3.6 Fokker-Planck equation . . . 1072
A.3.7 Reflection principle and stopping times . . . 1073
A.3.8 Some diffusion processes . . . 1074
A.3.8.1 Geometric Brownian motion . . . 1074
A.3.8.2 Ornstein-Uhlenbeck process . . . 1075
A.3.8.3 Cox-Ingersoll-Ross process . . . 1076
A.3.8.4 Multidimensional processes . . . 1077
A.4 Exercises . . . 1078
A.4.1 Discrete-time random process . . . 1078
A.4.2 Properties of Brownian motion . . . 1078
A.4.3 Stochastic integral for random step functions . . . 1078
A.4.4 Power of Brownian motion . . . 1079
A.4.5 Exponential of Brownian motion . . . 1079
A.4.6 Exponential martingales . . . 1080
A.4.7 Existence of solutions to stochastic differential equations . . . 1080
A.4.8 Itô calculus and stochastic integration . . . 1080
A.4.9 Solving a PDE with the Feynman-Kac formula . . . 1081
A.4.10 Fokker-Planck equation . . . 1081
A.4.11 Dynamic strategy based on the current asset price . . . 1081
A.4.12 Strong Markov property and maximum of Brownian motion . . . 1082

A.4.13 Moments of the Cox-Ingersoll-Ross process 1083
A.4.14 Probability density function of Heston and SABR models 1083
A.4.15 Discrete dynamic programming 1084
A.4.16 Matrix computation 1084

Subject Index **1121**

Author Index **1135**

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface

Teaching risk management in finance

This book is a handbook for students of Master's in finance, who want to learn risk management. It corresponds to the lecture notes of my course "*Risk Management & Financial Regulation*" at the University of Paris Saclay. This title highlights the role of financial regulation. Indeed, it appears that financial regulation is an important component to understand the practice of risk management in finance. This is particularly true in the banking sector, but it is also valid in other financial sectors. At first sight, it may be curious to teach for example the standards developed by the Basel Committee. They are freely available and any student may consult them. However, the regulation is so complex and the documentation produced is so abundant that students (but also professionals) may be lost when they want to have an overview on a specific topic or seek particular information. Therefore, I consider that the primary role of a course in risk management is to understand in general terms the financial regulation and be able to navigate between the various regulatory standards. This is all the more important that financial regulation is everywhere since the 2008 Global Financial Crisis (GFC). Today, most of the resources of a risk management department within a bank are dedicated to the regulation, and this is also the case of big projects. Understanding risk management requires them to know the regulation. Nevertheless, teaching risk management cannot be limited to the study of the regulation. Another important component of risk management is risk measurement. This requires having a statistical model for calculating the probability of a loss. A brief review shows that there are many risk models from the simplest to the most complicated because there are many types of risk and many risk factors. Moreover, the modeling of risk factors is not an easy task and requires making assumptions, and the complexity of a model can increase with the likelihood of these assumptions[1]. Therefore, the second role of a course in risk management is to distinguish between the mathematical models of risk measurement and study those that are actually used by professionals. From an academic point of view, some models may appear to be outdated or old-fashioned. However, they can continue to be used by risk managers for many reasons: more robust, easier to calibrate, etc. For example, the most important risk measurement model is certainly the historical value-at-risk. This is why it is important to choose the right models to study. A handbook cannot be a comprehensive catalogue of risk management methods. But it must present the most frequently used models and the essential mathematical tools in order to help the Master student when he will be faced with reality and situations that will require a more complex modeling.

[1] However, a complex model does not mean that the assumptions are more realistic.

About this book

These lecture notes are divided into two parts. After an introductory chapter presenting the main concepts of risk management and an overview of the financial regulation, the first part is dedicated to the risk management in the banking sector and is made up of seven chapters: market risk, credit risk, counterparty credit risk, operational risk, liquidity risk, asset liability management risk and systemic risk. I begin with the market risk, because it allows to introduce naturally the concept of risk factor, describe what a risk measure is and define the risk allocation approach. For each chapter, I present the corresponding regulatory framework and the risk management tools. I continue with five chapters that are mainly focused on the banking sector. However, even if these six chapters are dedicated to the banking sector, these materials also establish the basics of risk management in other financial sectors. They are the common language that is shared by all risk managers in finance. This first part ends with a eighth chapter on systemic risk and shadow banking system. In particular, this chapter supplements the introductory chapter and shows that the risk regulation culture has affected the other non-banking financial sectors such as asset management, insurance, pension funds and market infrastructure. The second part of these lecture notes develops the mathematical and statistical tools used in risk management. It contains seven chapters: model risk of exotic derivatives, statistical inference and model estimation, copula functions, extreme value theory, Monte Carlo simulation, stress testing methods and credit scoring models. Each chapter of these lecture notes is extensively illustrated by numerical examples and contains also tutorial exercises. Finally, a technical appendix completes the lecture notes and contains some important elements on numerical analysis.

The writing of these lecture notes started in April 2015 and is the result of twenty years of academic courses. When I began to teach risk management, a large part of my course was dedicated to statistical tools. Over the years, financial regulation became however increasingly important. I am convinced that risk management is now mainly driven by the regulation, not by the progress of the mathematical models. The writing of this book has benefited from the existing materials of my French book called "*La Gestion des Risques Financiers*". Nevertheless, the structure of the two books is different, because my previous book only concerned market, credit and operational risk before Basel III. Some years ago, I decided to extend the course to other financial sectors, especially insurance, asset management and market infrastructure. In fact, it appears that the quantitative methods of risk management are the same across the different financial areas even if each sector presents its particular aspects. But they differ mainly by the regulation, not by the mathematical tools. The knowledge of the different regulations is not an easy task for students. However, it is necessary if one would like to understand what the role of risk management is in financial institutions in the present-day world. Moreover, reducing the practice of risk management to the assimilation of the regulation rules is not sufficient. The sound understanding of the financial products and the mathematical models are essential to know where the risks are. This is why some parts of this book can be difficult because risk management is today complex in finance. A companion book, which contains the solutions of the tutorial exercises, is available in order to facilitate learning and knowledge assimilation at the following internet web page:

http://www.thierry-roncalli.com/RiskManagementBook.html

Acknowledgments

I taught risk management in various French universities: ENSAE (1999-2000), ENSAI (2000-2004), University of Paris-Dauphine (2001-2002), Telecom Sudparis (2015-2019) and University of Évry (2003-2019). I would like to thank the different people who helped me to organize these academic courses, in the chronological order: Nathalie Pistre, Hélyette Geman, Jean Lainé, Christophe Villa, Marc-Arthur Diaye, Monique Jeanblanc, Xavier Fairise, Christian de Peretti, Jérôme Glachant, Gaëlle Le Fol, Stéphane Crepey, Juliana Caicedo-Llano, Zhun Peng, Enareta Kurtbegu, Michel Guillard, Sun Yu, Wojciech Pieczynski, Agathe Guilloux, Fabien Tripier and Irinah Ratsimbazafy. I also want to acknowledge all the Master's students, who have participated in my risk management classes.

I also thank all my former colleagues of the *Groupe de Recherche Opérationnelle* (GRO) at Crédit Lyonnais and Crédit Agricole between 1999 and 2005, and especially Fatima Aïssaoui, Nicolas Baud, Abdelkader Bousabaa, Sophie Charbonnier, Anne Chartrain, Arnaud Costinot, Sophie Coutant, Paul Demey, Valdo Durrleman, Antoine Frachot, Pierre Georges, Pascal Gonnet, Sakda Hoeung, Wenjie Hou, Damien Jacomy, Jean-Frédéric Jouanin, David Kurtz, Ludivine Labarre, Thibault Lagache, Florence L'Hostis, Étienne Marot, Jean-Paul Marrero, Sloane Martin, Pierre Martineu, Laurent Michel, Amrane Mokdad, Olivier Moudoulaud, Ashkan Nikeghbali, Laure Philippon, Thomas Pignard, Vincent Porte, Grégory Rapuch, Andrée-Lise Rémy, Gaël Riboulet, Céline Roget, Éric Salomon and Bruno Vial. It was during the seven years I was in this research team that I learned the practice of risk management. The writing of this handbook is then highly influenced by the materials and models developed at GRO[2].

I am also grateful to Mohammed El Mendili, Joan Gonzalvez, Paul Jusselin, Nazar Kostyuchyk, Abdelaziz Lamaaizi, Marina Marena, Jules Roche, Théo Roncalli and Jiali Xu for having read some parts of this book and tested some exercises.

Last but not least, I express my deep gratitude to Théo, Eva, Sarah, Lucie and Nathalie for their support and encouragement during the writing of these lecture notes.

Paris, July 2019 Thierry Roncalli

[2] In particular, the chapter dedicated to the ALM owes a great deal to the works of Paul Demey, Antoine Frachot and Gaël Riboulet while the chapter on credit scoring is influenced by the unpublished lecture notes of Pierre Georges and Etienne Marot.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

List of Symbols and Notations

Symbol Description

$\times$	Arithmetic multiplication
$\cdot$	Scalar, vector and matrix multiplication
$*$	Convolution
$\circ$	Hadamard product: $(x \circ y)_i = x_i y_i$
$\otimes$	Kronecker product $A \otimes B$
$\lvert\mathcal{E}\rvert$	Cardinality of the set $\mathcal{E}$
$\prec$	Concordance ordering
$\langle x, x' \rangle$	Inner product of x and x'
$\mathbf{1}$	Vector of ones
$\mathbb{1}\{\mathcal{A}\}$	The indicator function is equal to 1 if $\mathcal{A}$ is true, 0 otherwise
$\mathbb{1}_{\mathcal{A}}\{x\}$	The characteristic function is equal to 1 if $x \in \mathcal{A}$, 0 otherwise
$\mathbf{0}$	Vector of zeros
$(A_{i,j})$	Matrix A with entry $A_{i,j}$ in row i and column j
A^{-1}	Inverse of the matrix A
$A^{1/2}$	Square root of the matrix A
$A^\top$	Transpose of the matrix A
A^+	Moore-Penrose pseudo-inverse of the matrix A
b	Vector of weights $(b_1, \ldots, b_n)$ for the benchmark b
$B_t(T)$	Price of the zero-coupon bond at time t for the maturity T
$B(t,T)$	Alternative form of $B_t(T)$
$\mathcal{B}(p)$	Bernoulli distribution with parameter p
$\mathcal{B}(n,p)$	Binomial distribution with parameter n and p
β_i	Beta of asset i with respect to portfolio w
$\beta_i(w)$	Another notation for the symbol β_i
$\beta(w \mid b)$	Beta of portfolio w when the benchmark is b
$\mathcal{B}(\alpha,\beta)$	Beta distribution with parameter α and β
$\mathfrak{B}(\alpha,\beta)$	Beta function defined as $\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$
$\mathfrak{B}(x;\alpha,\beta)$	Incomplete beta function $\int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$
$\boldsymbol{c}$	Coupon rate of the CDS premium leg
$\mathbb{C}$ (or ρ)	Correlation matrix
$\mathfrak{C}$	OTC contract
$\mathbf{C}(u_1,u_2)$	Copula function
$\mathcal{C}$	Set of copula functions
$\mathcal{C}(i)$	Mapping function
$\breve{\mathbf{C}}(u_1,u_2)$	Survival copula
$\mathbf{C}^-$	Fréchet lower bound copula
$\mathbf{C}^\perp$	Product copula
$\mathbf{C}^+$	Fréchet upper bound copula
$\boldsymbol{\mathcal{C}}_t$	Price of the call option at time t
$C(t_m)$	Coupon paid at time t_m
$\mathbb{C}_n(\rho)$	Constant correlation matrix of dimension n with $\rho_{i,j} = \rho$
$\mathrm{CE}(t_0)$	Current exposure at time t_0
$\mathrm{cov}(X)$	Covariance of the random vector X
$\chi^2(\nu)$	Chi-squared distribution with ν degrees of freedom
D	Covariance matrix of idiosyncratic risks
$\mathcal{D}(t)$	Liquidity duration of the new production
$\mathcal{D}^\star(t)$	Liquidity duration of the production stock
$\mathbf{D}_k(x)$	Debye function
$\det(A)$	Determinant of the matrix A
$\mathrm{diag}\, v$	Diagonal matrix with elements $(v_1, \ldots, v_n)$
$\boldsymbol{\Delta}_t$	Delta of the option at time t
Δ_h	Difference operator $\Delta_h V_t = V_t - V_{t-h}$ with lag h

$\Delta\,\mathrm{CoVaR}_i$ Delta CoVaR of institution i
Δt_m Time interval $t_m = t_{m-1}$
$\delta_x(y)$ Dirac delta function
$\mathbf{e}_i$ The value of the vector is 1 for the row i and 0 elsewhere
$\mathbb{E}[X]$ Mathematical expectation of the random variable X
$\mathcal{E}(\lambda)$ Exponential probability distribution with parameter λ
$e(t)$ Potential future exposure at time t
$\mathrm{EE}(t)$ Expected exposure at time t
$\mathrm{EEE}(t)$ Effective expected exposure at time t
$\mathrm{EEPE}(0;t)$ Effective expected positive exposure for the time period $[0,t]$
$\mathrm{EnE}(t)$ Risk-neutral expected negative exposure at time t
$\mathrm{EpE}(t)$ Risk-neutral expected positive exposure at time t
$\mathrm{EPE}(0;t)$ Expected positive exposure for the time period $[0,t]$
$\mathrm{ES}_\alpha(w)$ Expected shortfall of portfolio w at the confidence level α
$\exp(A)$ Exponential of the matrix A
$f(x)$ Probability density function
$f_{i:n}(x)$ Probability density function of the order statistic $X_{i:n}$
$f_y(\lambda)$ Spectral density function of the stochastic process y_t
$\mathbf{F}(x)$ Cumulative distribution function
$\mathbf{F}_{i:n}(x)$ Cumulative distribution function of the order statistic $X_{i:n}$
$\mathbf{F}^{-1}(\alpha)$ Quantile function
$\mathbf{F}^{n\star}$ n-fold convolution of the probability distribution $\mathbf{F}$ with itself
$\mathcal{F}$ Vector of risk factors $(\mathcal{F}_1,\ldots,\mathcal{F}_m)$
$\mathcal{F}_j$ Risk factor j
$\mathcal{F}_t$ Filtration
$f_t(T)$ Instantaneous forward rate at time t for the maturity T
$f(t,T)$ Alternative form of $f_t(T)$
$F_t(T_1,T_2)$ Forward interest rate at time t for the period $[T_1,T_2]$
$F(t,T_1,T_2)$ Alternative form of $F_t(T_1,T_2)$
$\mathfrak{F}(\nu_1,\nu_2)$ Fisher-Snedecor distribution with parameters ν_1 and ν_2
$\mathcal{G}(p)$ Geometric distribution with parameter p
$\mathcal{G}(\alpha)$ Standard gamma distribution with parameter α
$\mathcal{G}(\alpha,\beta)$ Gamma distribution with parameters α and β
γ_1 Skewness
γ_2 Excess kurtosis
$\mathbf{\Gamma}_t$ Gamma of the option at time t
$\Gamma(\alpha)$ Gamma function defined as $\int_0^\infty t^{\alpha-1}e^{-t}\,\mathrm{d}t$
$\gamma(\alpha,x)$ Lower incomplete gamma function defined as $\int_0^x t^{\alpha-1}e^{-t}\,\mathrm{d}t$
$\Gamma(\alpha,x)$ Upper incomplete gamma function defined as $\int_x^\infty t^{\alpha-1}e^{-t}\,\mathrm{d}t$
$\mathcal{GEV}(\mu,\sigma,\xi)$ GEV distribution with parameters μ, σ and ξ
$\mathcal{GPD}(\sigma,\xi)$ Generalized Pareto distribution with parameters σ and ξ
h Holding period
$\boldsymbol{h}$ Kernel or smoothing parameter
$\mathbf{H}^-$ Lower half-space
$\mathbf{H}^+$ Upper half-space
$\mathcal{H}$ Hyperplane
$H(X)$ Shannon entropy of X
$H(X,Y)$ Cross-entropy of X and Y
$H(Y \mid X)$ Conditional entropy of Y with respect to X
i Asset (or component) i
I_n Identity matrix of dimension n
$I(X,Y)$ Mutual information of X and Y
$\mathcal{I}(\theta)$ Information matrix
$\mathcal{IB}(x;\alpha,\beta)$ Regularized incomplete beta function
$\mathcal{J}(\theta)$ Fisher information matrix
$\boldsymbol{\mathcal{K}}$ Regulatory capital
$\mathcal{K}(x,x')$ Kernel function of x and x'
$\boldsymbol{\ell}(\theta)$ Log-likelihood function with θ the vector of parameters to estimate
$\boldsymbol{\ell}_t$ Log-likelihood function for the observation t
L Lag operator: $Ly_t = y_{t-1}$
L or $L(w)$ Loss of portfolio w
$\mathcal{L}(x;\lambda)$ Lagrange function, whose Lagrange multiplier is λ
$\ln A$ Logarithm of the matrix A
$\mathcal{LG}(\alpha,\beta)$ Log-gamma distribution with parameters α and β
$\mathcal{LL}(\alpha,\beta)$ Log-logistic distribution with parameters α and β
$\mathcal{LN}(\mu,\sigma^2)$ Log-normal distribution with parameters μ and σ
λ Parameter of exponential survival times
$\lambda(t)$ Hazard function

λ^- Lower tail dependence
λ^+ Upper tail dependence
$\mathbf{\Lambda}(x)$ Gumbel distribution
$\Lambda(t)$ Markov generator
$\mathrm{MDA}(\mathbf{G})$ Maximum domain of attraction of the extreme value distribution $\mathbf{G}$
MES_i Marginal expected shortfall of institution i
$\mathrm{MPE}_\alpha(0;t)$ Maximum peak exposure for the time period $[0,t]$ with a confidence level α
$\mathcal{MR}_i$ Marginal risk of asset i
MtM Mark-to-market of the portfolio
μ Vector of expected returns $(\mu_1,\ldots,\mu_n)$
μ_i Expected return of asset i
μ_m Expected return of the market portfolio
$\hat{\mu}$ Empirical mean
$\mu(w)$ Expected return of portfolio w
$\mu(X)$ Mean of the random vector X
$\mu_m(X)$ m-th centered moment of the random vector X
$\mu'_m(X)$ m-th moment of the random vector X
$\mathcal{N}(\mu,\sigma^2)$ Normal distribution with mean μ and standard deviation σ
$\mathcal{N}(\mu,\Sigma)$ Multivariate normal distribution with mean μ and covariance matrix Σ
n_S Number of scenarios or simulations
$N(t)$ Poisson counting process for the time interval $[0,t]$
$N(t_1;t_2)$ Poisson counting process for the time interval $[t_1,t_2]$
$\mathcal{NB}(r,p)$ Negative binomial distribution with parameters r and p
Ω Covariance matrix of risk factors
P Markov transition matrix
$\mathbb{P}$ Historical probability measure
$P(\Sigma)$ Cholesky decomposition of Σ
$\mathcal{P}(\lambda)$ Poisson distribution with parameter λ
$p(k)$ Probability mass function of an integer-valued random variable
$\boldsymbol{\mathcal{P}}_t$ Price of the put option at time t
$\mathcal{P}(\alpha,x_-)$ Pareto distribution with parameters α and x_-
$\mathcal{P}(\alpha,\theta)$ Pareto distribution with parameters α and θ
$\mathrm{PE}_\alpha(t)$ Peak exposure at time t with a confidence level α
$PV_t(\mathcal{L})$ Present value of the leg $\mathcal{L}$
Π or $\Pi(w)$ P&L of the portfolio w
$\phi(x)$ Probability density function of the standardized normal distribution
$\phi_2(x_1,x_2;\rho)$ Probability density function of the bivariate normal distribution with correlation ρ
$\phi_n(x;\Sigma)$ Probability density function of the multivariate normal distribution with covariance matrix Σ
$\Phi(x)$ Cumulative density function of the standardized normal distribution
$\Phi^{-1}(\alpha)$ Inverse of the cdf of the standardized normal distribution
$\Phi_2(x_1,x_2;\rho)$ Cumulative density function of the bivariate normal distribution with correlation ρ
$\Phi_n(x;\Sigma)$ Cumulative density function of the multivariate normal distribution with covariance matrix Σ
$\mathbf{\Phi}_\alpha(x)$ Fréchet distribution
$\mathbf{\Psi}_\alpha(x)$ Weibull distribution
$\varphi_X(t)$ Characteristic function of the random variable X
$q_\alpha(n_S)$ Integer part of αn_S
$q_{\bar{\alpha}}(n_S)$ Integer part of $(1-\alpha)n_S$
$\mathbb{Q}$ Risk-neutral probability measure
$\mathbb{Q}_T$ Forward probability measure
$\mathfrak{R}(t)$ Rating of the entity at time t
r Return of the risk-free asset
R Vector of asset returns $(R_1,\ldots,R_n)$
R_i Return of asset i
$R_{i,t}$ Return of asset i at time t
$R_{m,t}$ Return of the market portfolio at time t
$R(w)$ Return of portfolio w
$\mathcal{R}(w)$ Risk measure of portfolio w
$\mathcal{R}(L)$ Risk measure of loss L
$\mathcal{R}(\Pi)$ Risk measure of P&L Π
$R_t(T)$ Zero-coupon rate at time t for the maturity T
$\mathcal{RC}_i$ Risk contribution of asset i
$\mathcal{RC}_i^\star$ Relative risk contribution of asset i
$\boldsymbol{\mathcal{R}}$ Recovery rate
RPV_{01} Risky PV01

ρ (or $\mathbb{C}$) Correlation matrix of asset returns
$\rho_{i,j}$ Correlation between asset returns i and j
$\rho(x,y)$ Correlation between portfolios x and y
$\mathcal{s}$ Credit spread
$\mathbf{S}(x)$ Survival function
$\mathbb{S}$ Stress scenario
S_t Price of the underlying asset at time t
$\mathbf{S}_t(T)$ Survival function of T at time t
$\mathbf{S}(t,u)$ Amortization function of the new production
$\mathbf{S}^\star(t,u)$ Amortization function of the production stock
$\mathcal{S}(y_t)$ Stationary form of the process y_t
SES_i Systemic expected shortfall of institution i
$\mathcal{SN}(\xi,\Omega,\eta)$ Skew normal distribution
SRISK_i Systemic risk contribution of institution i
$\mathcal{ST}(\xi,\Omega,\eta,\nu)$ Skew t distribution
$SV_t(\mathcal{L})$ Stochastic discounted value of the leg $\mathcal{L}$
Σ Covariance matrix
$\hat{\Sigma}$ Empirical covariance matrix
σ_i Volatility of asset i
σ_m Volatility of the market portfolio
$\tilde{\sigma}_i$ Idiosyncratic volatility of asset i
$\hat{\sigma}$ Empirical volatility
$\sigma(w)$ Volatility of portfolio w
$\sigma(X)$ Standard deviation of the random variable X
t_ν Student's t distribution with ν degrees of freedom
$t_n(\Sigma,\nu)$ Multivariate Student's t distribution with ν degrees of freedom and covariance matrix Σ
$\mathbf{t}(x;\nu)$ Probability density function of the univariate t distribution with number of degrees of freedom ν
$\mathbf{t}_n(x;\Sigma,\nu)$ Probability density function of the multivariate t distribution with parameters Σ and ν
$\mathbf{t}_2(x_1,x_2;\rho,\nu)$ Probability density function of the bivariate t distribution with parameters ρ and ν
T Maturity date
$\mathbf{T}(x;\nu)$ Cumulative density function of the univariate t distribution with number of degrees of freedom ν
$\mathbf{T}^{-1}(\alpha;\nu)$ Inverse of the cdf of the Student's t distribution with ν the number of degrees of freedom
$\mathbf{T}_n(x;\Sigma,\nu)$ Cumulative density function of the multivariate t distribution with parameters Σ and ν
$\mathbf{T}_2(x_1,x_2;\rho,\nu)$ Cumulative density function of the bivariate t distribution with parameters ρ and ν
$\mathcal{T}$ return period
$\text{tr}(A)$ Trace of the matrix A
θ Vector of parameters
$\hat{\theta}$ Estimator of θ
$\boldsymbol{\Theta}_t$ Theta of the option at time t
$\boldsymbol{\tau}$ Default time
τ Time to maturity $T-t$
$\mathcal{U}_{[a,b]}$ Uniform distribution between a and b
$\text{var}(X)$ Variance of the random variable X
$\text{VaR}_\alpha(w)$ Value-at-risk of portfolio w at the confidence level α
$\boldsymbol{\upsilon}_t$ Vega of the option t
w Vector of weights $(w_1,\ldots,w_n)$ for portfolio w
w_i Weight of asset i in portfolio w
$W(t)$ Wiener process
X Random variable
x^+ Maximum value between x and 0
$X_{i:n}$ i^{th} order statistic of a sample of size n
y_t Discrete-time stochastic process
$\boldsymbol{y}$ Yield to maturity

Abbreviations

ABCP Asset-backed commercial paper
ABS Asset-backed security
ADF Augmented Dickey-Fuller unit root test
ADV Average daily volume
AER Annual equivalent rate
AFME Association for Financial Markets in Europe
AFS Available-for-sale
AIC Akaike information criterion
AIFMD Alternative investment fund managers directive

AIRB Advanced internal ratings-based approach (credit risk)
ALCO ALM committee
ALM Asset liability management
AM-CVA Advanced method (credit valuation adjustment)
AMA Advanced measurement approaches (operational risk)
AMF Autorité des Marchés Financiers
AMLF ABCP money market mutual fund liquidity facility
AR Autoregressive process
ARCH Autoregressive conditional heteroscedasticity process
ARMA Autoregressive moving average process
AT1 Additional tier 1
ATM At-the-money (option)
BA-CVA Basic approach (credit valuation adjustment)
BAC Binary asset-or-nothing call option
BaFin Bundesanstalt für Finanzdienstleistungsaufsicht
BAP Binary asset-or-nothing put option
BCBS Basel Committee on Banking Supervision
BCC Binary cash-or-nothing call option
BCP Binary cash-or-nothing put option
BCVA Bilateral CVA
BD Broker-dealer
BFGS Broyden-Fletcher-Goldfarb-Shanno algorithm
BGD Batch gradient descent
BIA Basic indicator approach (operational risk)
BIS Bank for International Settlements
BLUE Best linear unbiased estimator
BoJ Bank of Japan
BS Black-Scholes model
BSM Basic structural model
BUE Best unbiased estimator
CAD Capital adequacy directive
CAM Constant amortization mortgage
CaR Capital-at-risk
CB Conservation buffer (CET1)
CBO Collateralized bond obligation
CCB Countercyclical capital buffer (CET1)
CCF Credit conversion factor
CCP Central counterparty clearing house
CCR Counterparty credit risk
CDF Cumulative distribution function
CDO Collateralized debt obligation
CDS Credit default swap
CDT Credit default tranche
CDX Credit default index
CE Current exposure
CEM Current exposure method (CCR)
CET1 Common equity tier 1
CFH Cash flow hedge
CFI Captive financial institution
CFO Chief financial officer
CGFS Committee on the Global Financial System
CIR Cox-Ingersoll-Ross process
CISC Constant inter-sector correlation model
CLO Collateralized loan obligation
CMBS Commercial mortgage-backed security
CMO Collateralized mortgage obligation
CoVaR Conditional value-at-risk
CP Consultation paper
CPM Constant payment mortgage
CPR Conditional prepayment rate
CRA Credit rating agency
CRD Capital requirements directive
CRM Comprehensive risk measure
CRO Chief risk officer
CRR Capital requirements regulation
CSRBB Credit spread risk in the banking book
CVA Credit valuation adjustment
DF Dickey-Fuller unit root test
DFAST Dodd-Frank Act stress testing
DFP Davidon-Fletcher-Powell algorithm
DFT Discrete Fourier transform
DGAP Duration gap
DIC Down-and-in call option
DIP Down-and-in put option
DOC Down-and-out call option
DOP Down-and-out put option
DP Dynamic programming
DRC Default risk capital
DV01 Dollar value of a one basis point decrease in interest rates
DVA Debit valuation adjustment
EAD Exposure at default
EaR Earnings-at-risk
EAR Effective annual rate
EBA European Banking Authority

ECB	European Central Bank
ECM	Error correction model
ECRA	External credit risk assessment
EE	Expected exposure
EEE	Effective expected exposure
EEPE	Effective expected positive exposure
EL	Expected loss
EMIR	European market infrastructure regulation
ENE	Expected negative exposure
EPE	Expected positive exposure
ERBA	External ratings-based approach
ES	Expected shortfall
ESMA	European Securities and Markets Authority
ETF	Exchange traded fund
EV	Economic value
EVaR	Economic value-at-risk
EVE	Economic value of equity
EVT	Extreme value theory
FASB	Financial Accounting Standards Board
FBA	Fall-back approach
FC	Finance company
FDIC	Federal Deposit Insurance Corporation
FDML	Frequency domain maximum likelihood
FFT	Fast Fourier transform
FHFA	Federal Housing Finance Agency
FICO	Fair Isaac Corporation score
FIR	Finite impulse response filter
FIRB	Foundation internal ratings-based approach (credit risk)
FNMA	Fannie Mae
FRA	Forward rate agreement
FRB	Board of Governors of the Federal Reserve System
FRTB	Fundamental review of the trading book
FSAP	Financial sector assessment program
FSB	Financial Stability Board
FtD	First-to-default swap
FTP	Funds transfer pricing
FV	Fair value
FVA	Founding valuation adjustment
FVH	Fair value hedge
FVOCI	Fair value through other comprehensive income
FVTPL	Fair value through profit and loss
FWN	Fractional white noise
GAAP	Generally accepted accounting principles (US)
GARCH	Generalized autoregressive conditional heteroscedasticity process
GBM	Geometric Brownian motion
GCV	Generalized cross-validation
GEV	Generalized extreme value distribution
GFC	Global Financial Crisis (2008)
GMM	Generalized method of moments
GMM	Gaussian mixture model
GNMA	Ginnie Mae
GPD	Generalized Pareto distribution
HELOC	Home equity line of credit
HF	Hedge fund
HFT	Held-for-trading
HJM	Heath-Jarrow-Morton model
HLA	Higher loss absorbency
HPP	Homogeneous Poisson process
HQLA	High-quality liquid assets
HTM	Held-to-maturity
HY	High yield entity
IAIS	International Association of Insurance Supervisors
IAS	International accounting standards
ICAAP	Internal capital adequacy assessment process
ICP	Insurance Core Principles
ICPF	Insurance companies and pension funds
IF	Investment fund
IFG	Infinitely fine-grained portfolio
IFRS	International financial reporting standards
IG	Investment grade entity
ILAAP	Internal liquidity adequacy assessment process
IMA	Internal model-based approach (market risk)
IMCC	Internally modelled capital charge (Basel III)
IMF	International Monetary Fund
IMM	Internal model method (counterparty credit risk)
IOSCO	International Organization of Securities Commissions
IPP	Integration by parts
IRB	Internal ratings-based approach (credit risk)

IRRBB	Interest rate risk in the banking book
IRC	Incremental risk charge
IRS	Interest rate swap
ISDA	International Swaps and Derivatives Association
ITM	In-the-money (option)
JTD	Jump-to-default
KF	Kalman filter
KIC	Knock-in call option
KIP	Knock-in put option
KOC	Knock-out call option
KOP	Knock-out put option
KPSS	Kwiatkowski-Phillips-Schmidt-Shin stationary test
KRI	Key risk indicator
L&R	Loans and receivables
LAD	Least absolute deviation estimator
LCG	Linear congruential generator
LCR	Liquidity coverage ratio
LDA	Loss distribution approach (operational risk)
LDA	Linear discriminant analysis
LDCE	Loss data collection exercise
LEE	Loan equivalent exposure
LGD	Loss given default
LL	Local level model
LLT	Local linear trend model
LMM	Libor market model
LTA	Look-through approach
LtD	Last-to-default swap
LTI	Linear time-invariant filter
LTV	Loan-to-value ratio
M	Effective maturity
MA	Moving average process
MBS	Mortgage-backed security
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MCR	Minimum capital requirement
MDA	Maximum domain of attraction
MDB	Multilateral development bank
MES	Marginal expected shortfall
MEV	Multivariate extreme value
MF	Mutual fund
MGD	Mini-batch gradient descent
MiFID	Markets in financial instruments directive
MiFIR	Markets in financial instruments regulation
ML	Maximum likelihood
MLE	Maximum likelihood estimator
MM	Method of moments
MMF	Money market fund
MPE	Maximum peak exposure
MPOR	Margin period of risk
MPP	Mixed Poisson process
MSMVE	Min-stable multivariate exponential distribution
MtM	Mark-to-market
MUNFI	Monitoring universe of non-bank financial intermediation
NHPP	Non-homogeneous Poisson process
NIH	Net investment hedge
NII	Net interest income
NIM	Net interest margin
NIS	Net interest spread
NMD	Non-maturity deposit
NMF	Non-negative matrix factorization
NN	Neural network
NOW	Negotiable order of withdrawal
NQD	Negative quadrant dependence
NSFR	Net stable funding ratio
OCC	Office of the Comptroller of the Currency
ODE	Ordinary differential equation
OFI	Other financial intermediary
OLS	Ordinary least squares
ORSA	Own risk and solvency assessment
OTC	Over-the-counter
OTM	Out-of-the-money (option)
OTS	Office of Thrift Supervision
OU	Ornstein-Uhlenbeck process
P&L	Profit and loss
PCA	Principal component analysis
PD	Probability of default
PDE	Partial differential equation
PDF	Probability density function
PE	Peak exposure
PFE	Potential future exposure
PLA	Profit and loss attribution (Basel III)
PMF	Probability mass function
POT	Peak over threshold
PP	Phillips-Perron unit root test
PQD	Positive quadrant dependence
PRESS	Predicted residual error sum of squares
PSE	Public sector entity
PV01	Present value of one bp
QDA	Quadratic discriminant analysis
QIS	Quantitative impact study
QMC	Quasi-Monte Carlo

QP	Quadratic programming
RBC	Risk-based capital (US insurance)
REIT	Real estate investment trust
RFET	Risk factor eligibility test (Basel III)
RLS	Recursive least squares
RMBS	Residential mortgage-backed security
ROE	Return-on-equity
RRAO	Residual risk add-on
RW	Risk weight
RWA	Risk-weighted asset
RWR	Right way risk
SA	Standardized approach (credit risk)
SA-CCR	Standardized approach (counterparty credit risk)
SA-CVA	Standardized approach (credit valuation adjustment)
SA-TB	Standardized approach for the trading book (market risk)
SABR	Stochastic alpha-beta-rho model
SBE	Shadow banking entity
SBS	Shadow banking system
SCR	Solvency capital requirement
SCRA	Standardized credit risk approach
SDE	Stochastic differential equation
SES	Systemic expected shortfall
SFT	Securities financing transaction
SFV	Structured finance vehicle
SGD	Stochastic gradient descent
SIFI	Systemically important financial institution
SIFMA	Securities Industry and Financial Markets Association
SIR	Sampling importance resampling
SIS	Sequential importance sampling
SIV	Structured investment vehicle
SLA	Single loss approximation
SLN	Shifted log-normal model
SM-CCR	Standardized method (counterparty credit risk)
SM-CVA	Standardized method (credit valuation adjustment)
SMC	Sequential Monte Carlo
SME	Small and medium-sized enterprises
SMM	Standardized measurement method (market risk)
SMM	Swap market model
SN	Skew normal distribution
SPV	Special purpose vehicle
SQP	Sequential quadratic programming
SRC	Specific risk charge
SREP	Supervisory review and evaluation process
SRISK	Systemic risk contribution
SRP	Supervisory review process
SSFA	Simplified supervisory formula approach
SSM	Single supervisory mechanism
SSM	State space model
ST	Skew t distribution
STC	Simple, transparent and comparable (securitization)
StD	Second-to-default swap
SVaR	Stressed value-at-risk
SVI	Stochastic volatility inspired
SVM	Support vector machine
T1	Tier 1
T2	Tier 2
TC	Trust company
TDML	Time domain maximum likelihood
TDRR	Term deposit redemption ratio
TLAC	Total loss absorbing capacity
TSA	The standardized approach (operational risk)
UCITS	Undertakings for collective investment in transferable securities (directive)
UCVA	Unilateral CVA
UDVA	Unilateral DVA
UL	Unexpected loss
UIC	Up-and-in call option
UIP	Up-and-in put option
UOC	Up-and-out call option
UOP	Up-and-out put option
UVM	Uncertain volatility model
VaR	Value-at-risk
VAR	Vector autoregressive process
VARMA	Vector autoregressive moving average process
VECM	Vector error correction model
WAL	Weighted average life
WLS	Weighted least squares
WWR	Wrong way risk
XO	Crossover (or sub-investment grade) entity

Other scientific conventions

YYYY-MM-DD	We use the international standard date notation where YYYY is the year in the usual Gregorian calendar, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31.
USD (or $)	US dollar
EUR (or €)	Euro
KUSD	One thousand dollars
$1 mn/bn/tn	One million/billion/trillion dollars
bp	Basis point or 0.01%

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 1

Introduction

The idea that risk management creates value is largely accepted today. However, this has not always been the case in the past, especially in the financial sector (Stulz, 1996). Rather, it has been a long march marked by a number of decisive steps. In this introduction, we present an outline of the most important achievements from a historical point of view. We also give an overview of the current financial regulation, which is a cornerstone in financial risk management.

1.1 The need for risk management

The need for risk management is the title of the first section of the leadership book by Jorion (2007), who shows that risk management can be justified at two levels. At the firm level, risk management is essential for identifying and managing business risk. At the industry level, risk management is a central factor for understanding and preventing systemic risk. In particular, this second need is the '*raison d'être*' of the financial regulation itself.

1.1.1 Risk management and the financial system

The concept of risk management has evolved considerably since its creation, which is believed to be in the early fifties[1]. In November 1955, Wayne Snider gave a lecture entitled '*The Risk Manager*' where he proposed creating an integrated department responsible for risk prevention in the insurance industry (Snider, 1956). Some months later, Gallagher (1956) published an article to outline the most important principles of risk management and to propose the hiring of a full-time risk manager in large companies. For a long time, risk management was systematically associated with insurance management, both from a practical point of view and a theoretical point of view. For instance, the book of Mehr and Hedges (1963) is largely dedicated to the field of insurance with very few applications to other industries. This is explained by the fact that the collective risk model[2] has helped to apply the mathematical and statistical tools for measuring risk in insurance companies since 1930. A new discipline known as actuarial science has been developed at the same time outside the other sciences and has supported the generalization of risk management in the insurance industry.

Simultaneously, risk became an important field of research in economics and finance. Indeed, Arrow (1964) made an important step by extending the Arrow-Debreu model of general equilibrium in an uncertain environment[3]. In particular, he showed the importance

[1] See Crockford (1982) or Snider (1991) for a retrospective view on the risk management development.
[2] It is also known as the ruin theory or the compound Poisson risk model.
[3] This paper was originally presented in 1952 and was also published in Cahiers du CNRS (1953).

of hedging and introduced the concept of payoff. By developing the theory of optimal allocation for a universe of financial securities, Markowitz (1952) pointed out that the risk of a financial portfolio can be diversified. These two concepts, hedging and diversification, together with insurance, are the main pillars of modern risk management. These concepts will be intensively used by academics in the 1960s and 1970s. In particular, Black and Scholes (1973) showed the interconnection between hedging and pricing problems. Their work had a strong impact on the development of equity, interest rates, currency and commodity derivatives, which are today essential for managing the risk of financial institutions. With the Markowitz model, a new era had begun in portfolio management and asset pricing. First, Sharpe (1964) showed how risk premia are related to non-diversifiable risks and developed the first asset pricing model. Then, Ross (1976) extended the CAPM model of Sharpe and highlighted the role of risk factors in arbitrage pricing theory. These academic achievements will support the further development of asset management, financial markets and investment banking.

In commercial and retail banking, risk management was not integrated until recently. Even though credit scoring models have existed since the fifties, they were rather designed for consumer lending, especially credit cards. When banks used them for loans and credit issuances, they were greatly simplified and considered as a decision-making tool, playing a minor role in the final decision. The underlying idea was that the banker knew his client better than a statistical model could. However, Banker Trust introduced the concept of risk-adjusted return on capital or RAROC under the initiative of Charles Sanford in the late 1970s for measuring risk-adjusted profitability. Gene Guill mentions a memorandum dated February 1979 by Charles Sanford to the head of bank supervision at the Federal Reserve Board of New York that helps to understand the RAROC approach:

> "*We agree that one bank's book equity to assets ratio has little relevance for another bank with a different mix of businesses. Certain activities are inherently riskier than others and more risk capital is required to sustain them. The truly scarce resource is equity, not assets, which is why we prefer to compare and measure businesses on the basis of return on equity rather than return on assets*" (Guill, 2009, page 10).

RAROC compares the expected return to the economic capital and has become a standard model for combining performance management and risk management. Even if RAROC is a global approach for allocating capital between business lines, it has been mainly used as a credit scoring model. Another milestone was the development of credit portfolio management when Vasicek (1987) adapted the structural default risk approach of Merton (1974) to model the loss distribution of a loan portfolio. He then jointly founded KMV Corporation with Stephen Kealhofer and John McQuown, which specializes in quantitative credit analysis tools and is now part of Moody's Analytics.

In addition to credit risk, commercial and retail banks have to manage interest rate and liquidity risks, because their primary activity is to do asset, liquidity and maturity transformations. Typically, a commercial bank has long-term and illiquid assets (loans) and short-term and liquid liabilities (deposits). In such a situation, a bank faces a loss risk that can be partially hedged. This is the role of asset liability management(ALM). But depositors also face a loss risk that is virtually impossible to monitor and manage. Consequently, there is an information asymmetry between banks and depositors.

In the banking sector, the main issue centered therefore around the deposit insurance. How can we protect depositors against the failure of the bank? The 100% reserve proposal by Fisher (1935) required banks to keep 100% of demand deposit accounts in cash or government-issued money like bills. Diamond and Dybvig (1983) argued that the mixing

policy of liquid and illiquid assets can rationally produce systemic risks, such as bank runs. A better way to protect the depositors is to create a deposit insurance guaranteed by the government. According to the Modigliani-Miller theorem on capital structure[4], this type of government guarantee implied a higher cost of equity capital. Since the eighties, this topic has been highly written about (Admati and Hellwig, 2014). Moreover, banks also differ from other companies, because they create money. Therefore, they are at the heart of the monetary policy. These two characteristics (implicit guarantee and money creation) imply that banks have to be regulated and need regulatory capital. This is all the more valid with the huge development of financial innovations, which has profoundly changed the nature of the banking system and the risk.

1.1.2 The development of financial markets

The development of financial markets has a long history. For instance, the Chicago Board of Trade (CBOT) listed the first commodity futures contract in 1864 (Carlton, 1984). Some authors even consider that the first organized futures exchange was the Dojima Rice Market in Osaka in the 18th century (Schaede, 1989). But the most important breakthrough came in the seventies with two major financial innovations. In 1972, the Chicago Mercantile Exchange (CME) launched currency futures contracts after the US had decided to abandon the fixed exchange rate system of Bretton Woods (1946). The oil crisis of 1973 and the need to hedge currency risk have considerably helped in the development of this market. After commodity and currency contracts, interest rate and equity index futures have consistently grown. For instance, US Treasury bond, S&P 500, German Bund, and Euro Stoxx 50 futures were first traded in 1977, 1982, 1988 and 1998 respectively. Today, the Bund futures contract is the most traded product in the world.

The second main innovation in the seventies concerned option contracts. The CBOT created the Chicago Board of Options (CBOE) in 1973, which was the first exchange specialized in listed stock call options. The same year, Black and Scholes (1973) published their famous formula for pricing a European option. It has been the starting point of the intensive development of academic research concerning the pricing of financial derivatives and contingent claims. The works of Fisher Black, Myron Scholes and Robert Merton[5] are all the more significant in that they consider the pricing problem in terms of risk hedging. Many authors had previously found a similar pricing formula, but Black and Scholes introduced the revolutionary concept of the hedging portfolio. In their model, they derived the corresponding dynamic trading strategy to hedge the option contract, and the option price is therefore equivalent to the cost of the hedging strategy. Their pricing method had a great influence on the development of the derivatives market and more exotic options, in particular path-dependent options[6].

Whereas the primary goal of options is to hedge a directional risk, they will be largely used as underlying assets of investment products. In 1976, Hayne Leland and Mark Rubinstein developed the portfolio insurance concept, which allows for investing in risky assets while protecting the capital of the investment. In 1980, they founded LOR Associates, Inc. with John O'Brien and proposed structured investment products to institutional investors (Tufano and Kyrillos, 1995). They achieved very rapid growth until the 1987 stock market

[4] Under some (unrealistic) assumptions, Modigliani and Miller (1958) showed that the market value of a firm is not affected by how that firm is financed (by issuing stock or debt). They also established that the cost of equity is a linear function of the firm's leverage measured by its debt/equity ratio.

[5] As shown by Bernstein (1992), the works of Black and Scholes cannot be dissociated from the research of Merton (1973). This explains why they both received the 1997 Nobel Prize in Economics for their option pricing model.

[6] See Box 1 for more information about the rise of exotic options.

crash[7], and were followed by Wells Fargo, J.P. Morgan and Chase Manhattan as well as other investment banks. This period marks the start of financial engineering applied to structured products and the development of popular trading strategies, such as constant proportion portfolio insurance (CPPI) and option based portfolio insurance (OBPI). Later, they will be extensively used for designing retail investment products, especially capital guaranteed products.

Box 1

Evolution of financial innovations

1864	Commodity futures
1970	Mortgage-backed securities
1971	Equity index funds
1972	Foreign currency futures
1973	Stock options
1977	Put options
1979	Over-the-counter currency options
1980	Currency swaps
1981	Interest rate swaps
1982	Equity index futures
1983	Equity index options
	Interest rate caps/floors
	Collateralized mortgage obligations
1985	Swaptions
	Asset-backed securities
1987	Path-dependent options (Asian, look-back, etc.)
	Collateralized debt obligations
1992	Catastrophe insurance futures and options
1993	Captions/floortions
	Exchange-traded funds
1994	Credit default swaps
1996	Electricity futures
1997	Weather derivatives
2004	Volatility index futures
2006	Leveraged and inverse ETFs
2008	Green bonds
2009	Crypto currencies

Source: Jorion (2007) and author's research.

After options, the next great innovation in risk management was the swap. In a swap contract, two counterparties exchange a series of cash flows of one financial instrument for those of another financial instrument. For instance, an interest rate swap (IRS) is an exchange of interest rate cash flows from a fixed rate to a floating rate or between two floating

[7] In fact, portfolio insurance was blamed by the Brady Commission report (1988) for the stock market crash of October 1987. See for instance Leland and Rubinstein (1988), Shiller (1987), Gennotte and Leland (1990) and Jacklin *et al.* (1992) for a discussions about the impact of portfolio insurance on the October 1987 crash.

rates. Swaps have become an important tool for managing balance sheets, in particular interest rate and currency risks in the banking book. The original mechanism of cash flow exchanges has been extended to other instruments and underlying assets: inflation-indexed bonds, stocks, equity indices, commodities, etc. But one of the most significant advances in financial innovations was the creation of credit default swaps (CDS) in the mid-nineties, and more generally credit derivatives. In the simplest case, the cash flows depend on the default of a loan, a bond or a company. We refer then to single-name instruments. Otherwise, they depend on credit events or credit losses of a portfolio (multi-name instruments). However, the development of credit derivatives was made possible thanks to securitization. This is a process through which assets are pooled in a portfolio and securities representing interests in the portfolio are issued. Securities backed by mortgages are called mortgage-backed securities (MBS), while those backed by other types of assets are asset-backed securities (ABS).

Derivatives are traded either in organized markets or in over-the-counter markets (OTC). In organized exchanges, the contracts are standardized and the transactions are arranged by the clearing house, which is in charge of clearing and settlement. By contrast, in OTC markets, the contracts are customized and the trades are done directly by the two counterparties. This implies that OTC trades are exposed to the default risk of the participants. The location of derivatives trades depends on the contract:

Contract	Futures	Forward	Option	Swap
On-exchange	✓		✓	
Off-exchange		✓	✓	✓

For instance, the only difference between futures and forward contracts is that futures are traded in organized markets whereas forwards are traded over-the-counter. Contrary to options which are negotiated in both markets, swaps are mainly traded OTC. In Table 1.1, we report the outstanding amount of exchange-traded derivatives concerning futures and options published by the Bank for International Settlements (2019). In December 2018, their notional amount is equal to \$94.8 tn, composed of \$39.0 tn in futures (41.2%) and \$55.7 tn in options (58.8%). For each instrument, we indicate the split between interest rates and currencies[8]. We notice that exchange-traded derivatives on interest rates are the main contributor. The evolution of the total notional amount is reported in Figure 1.1. The size of exchange-traded derivative markets has grown rapidly since 2000, peaking in June 2007 with an aggregated amount of \$86.6 tn. This trend ended with the financial crisis since we observe a decrease between 2007 and 2016. This is only recently that the outstanding amount of exchange-traded derivatives exceeds the 2007 figure.

Statistics[9] concerning OTC derivative markets are given in Table 1.2. These markets are between six and ten times bigger than exchange-traded markets in terms of outstanding amount (Figure 1.3). In June 2018, the aggregated amount of forwards, swaps and options is equal to \$594.8 tn. Contrary to exchange-traded derivative markets, the notional outstanding amount of OTC derivative markets continues to increase after the crisis period, but declines recently since 2014 (Figure 1.2). In terms of instruments, swaps dominate and represent 65.0% of the total. Like in exchange-traded markets, the main asset class remains fixed income. We also notice the impact of the 2008 financial crisis on credit default swaps,

[8]The BIS decided in September 2015 to discontinue the compilation of equity index exchange-traded derivatives statistics. This is why these statistics do not include the equity index futures and options. In December 2014, equity index futures and options represented 11.1% of exchange-traded derivatives.

[9]In order to compute these statistics, we have done some assumptions because we don't have a perfect granularity of the data. For equity and commodity buckets, we don't have the split between forwards and swaps. We allocate 50% of the amount in each category. We also attribute the full amount of credit derivatives to the swap bucket.

TABLE 1.1: Notional outstanding amount of exchange-traded derivatives

	2004	2007	2010	2014	2018
Futures	42.6%	37.9%	34.1%	44.4%	41.2%
Interest rate	99.4%	99.3%	99.2%	99.1%	99.3%
Short-term	94.7%	94.0%	94.9%	93.6%	92.6%
Long-term	5.3%	6.0%	5.1%	6.4%	7.4%
Currency	0.6%	0.7%	0.8%	0.9%	0.7%
Options	57.4%	62.1%	65.9%	55.6%	58.8%
Interest rate	99.8%	99.7%	99.6%	99.6%	99.8%
Short-term	98.2%	98.6%	98.9%	97.7%	98.3%
Long-term	1.9%	1.4%	1.1%	2.3%	1.7%
Currency	0.2%	0.3%	0.4%	0.5%	0.3%
Total (in $ tn)	43.0	71.5	62.3	57.6	94.8

Source: Bank for International Settlements (2019) and author's calculations.

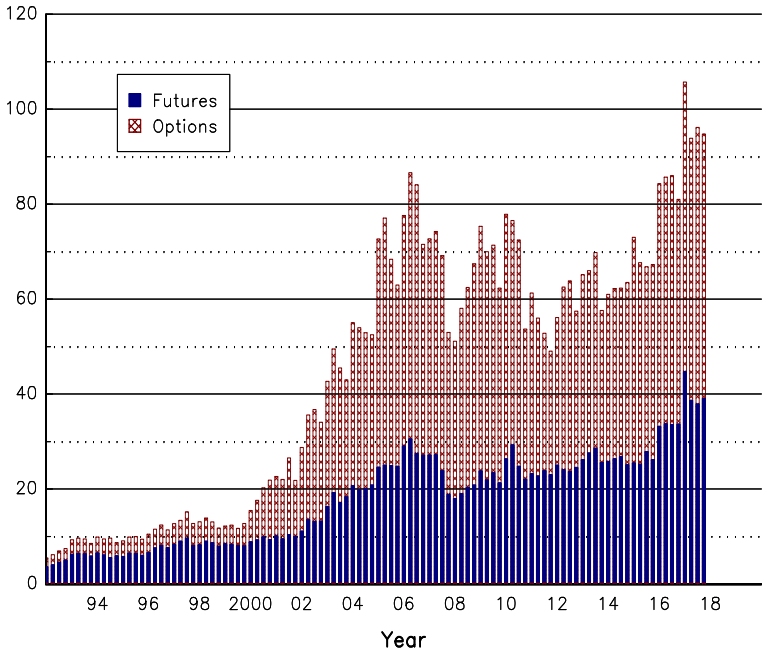


FIGURE 1.1: Notional outstanding amount of exchange-traded derivatives (in $ tn)

Source: Bank for International Settlements (2019) and author's calculations.

which represented more than 10% of the OTC derivative markets in December 2007. Ten years after, they represent less than 2.0% of these markets.

TABLE 1.2: Notional outstanding amount of OTC derivatives

	2004	2007	2010	2014	2018
Forwards	12.9%	11.8%	15.4%	20.2%	24.0%
Swaps	71.1%	73.3%	73.2%	69.4%	65.0%
Options	15.9%	14.9%	11.4%	10.3%	10.8%
Unallocated	0.1%	0.0%	0.0%	0.1%	0.1%
Currency	13.4%	11.4%	11.3%	13.1%	16.1%
Interest rate	79.5%	73.8%	81.9%	82.8%	80.9%
Equity	2.0%	1.6%	1.0%	1.1%	1.2%
Commodity	0.6%	1.6%	0.6%	0.3%	0.4%
Credit	4.5%	11.6%	5.2%	2.7%	1.4%
Unallocated	0.1%	0.0%	0.0%	0.0%	0.0%
Total (in $ tn)	258.6	585.9	601.0	627.8	594.8

Source: Bank for International Settlements (2019) and author's calculations.

Whereas the notional outstanding amount is a statistic to understand the size of the derivatives markets, the risk and the activity of these markets may be measured by the gross market value and the turnover:

- The *gross market value* of outstanding derivatives contracts represents "*the cost of replacing all outstanding contracts at market prices prevailing on the reporting date. It corresponds to the maximum loss that market participants would incur if all counterparties failed to meet their contractual payments and the contracts were replaced at current market prices*" (Bank for International Settlements, 2014).
- The *turnover* is defined as "*the gross value of all new deals entered into during a given period, and is measured in terms of the nominal or notional amount of the contracts. It provides a measure of market activity, and can also be seen as a rough proxy for market liquidity.*" (Bank for International Settlements, 2014).

In June 2018, the gross market value is equal to $10.3 tn for OTC derivatives. It is largely lower than the figure of $34.9 tn in December 2008. This decrease is explained by less complexity in derivatives, but also by a lower volatility regime. For OTC derivatives, it is difficult to measure a turnover, because the contracts are not standardized. This statistic is more pertinent for exchange-traded markets. In December 2018, the daily average turnover is equal to $8.1 tn for futures contacts and $1.8 tn for options. This means that each day, almost $10 tn of new derivative exposures are negotiated in exchange-traded markets. The consequence of this huge activity is a growing number of financial losses for banks and financial institutions (Reinhart and Rogoff, 2009).

1.1.3 Financial crises and systemic risk

A financial institution generally faces five main risks: (1) market risk, (2) credit risk, (3) counterparty credit risk, (4) operational risk and (5) liquidity risk. Market risk is the risk of losses due to changes in financial market prices. We generally distinguish four major types

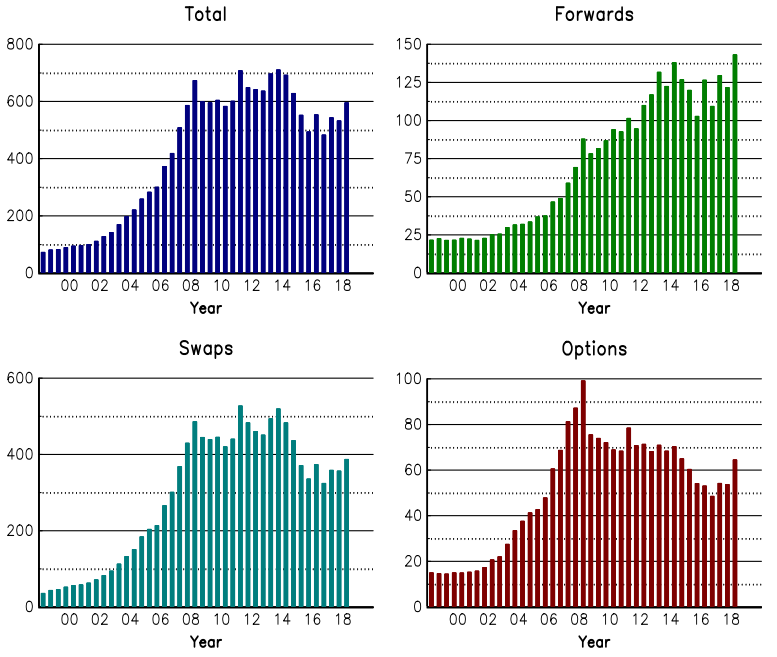


FIGURE 1.2: Notional outstanding amount of OTC derivatives (in $ tn)

Source: Bank for International Settlements (2019).

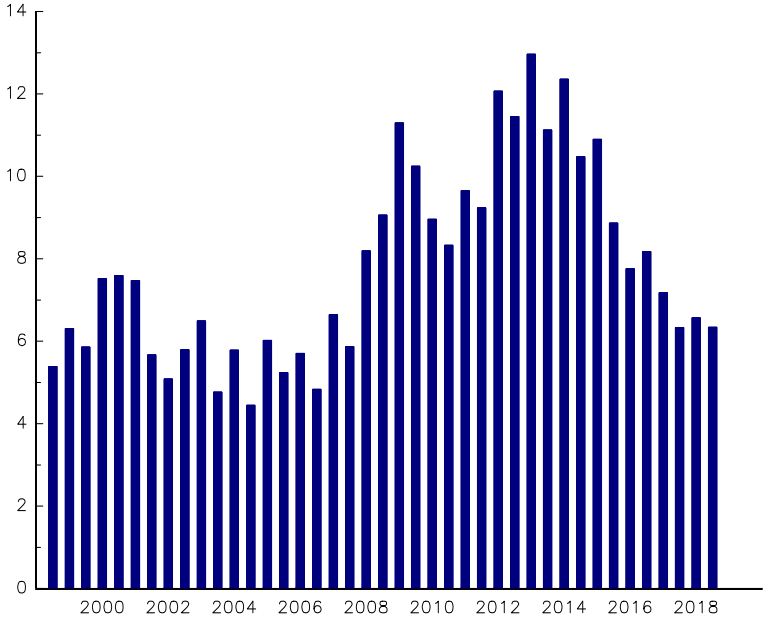


FIGURE 1.3: Ratio OTC derivatives/exchange-traded derivatives

Source: Bank for International Settlements (2019).

of market risk: equity risk, interest rate risk, currency risk and commodity risk. These risks are present in trading activities, but they also affect all activities that use financial assets. Credit risk is the risk of losses due to the default of a counterparty to fulfill its contractual obligations, that is to make its required payments. It principally concerns debt transactions such as loans and bonds. Counterparty credit risk is another form of credit risk, but concerns the counterparty of OTC transactions. Examples include swaps and options, security lending or repo transactions. Operational risk is the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events. Examples of operational risk are frauds, natural disasters, business disruption, rogue trading, etc. Finally, liquidity risk is the risk of losses resulting from the failure of the financial institution to meet its obligations on time. This definition corresponds more to funding liquidity, but liquidity risk also concerns market liquidity, which is the cost to buy or sell assets on the market.

Box 2

An history of financial losses

1974	Herstatt Bank: $620 mn (foreign exchange trading)
1994	Metallgesellschaft: $1.3 bn (oil futures)
1994	Orange County: $1.8 bn (reverse repo)
1994	Procter & Gamble: $160 mn (ratchet swap)
1995	Barings Bank: $1.3 bn (stock index futures)
1997	Natwest: $127 mn (swaptions)
1998	LTCM: $4.6 bn (liquidity crisis)
2001	Dexia Bank: $270 mn (corporate bonds)
2006	Amaranth Advisors: $6.5 bn (gaz forward contracts)
2007	Morgan Stanley: $9.0 bn (credit derivatives)
2008	Société Générale: $7.2 bn (rogue trading)
2008	Madoff: $65 bn (fraud)
2011	UBS: $2.0 bn (rogue trading)
2012	JPMorgan Chase: $5.8 bn (credit derivatives)

Source: Jorion (2007) and author's research.

In Box 2, we have reported some famous financial losses. Most of them are related to the market risk or the operational risk[10]. In this case, these losses are said to be idiosyncratic because they are specific to a financial institution. Idiosyncratic risk is generally opposed to systemic risk: systemic risk refers to the system whereas idiosyncratic risk refers to an entity of the system. For instance, the banking system may collapse, because many banks may be affected by a severe common risk factor and may default at the same time. In financial theory, we generally make the assumption that idiosyncratic and common risk factors are independent. However, there exist some situations where idiosyncratic risk may affect the system itself. It is the case of large financial institutions, for example the default of big banks. In this situation, systemic risk refers to the propagation of a single bank distressed risk to the other banks.

[10] We have excluded the credit risk losses due to the 2008 global financial crisis. Even if the true cost of this crisis will never be known, it is very high, certainly larger than $10 tn.

The case of Herstatt Bank is an example of an idiosyncratic risk that could result in a systemic risk. Herstatt Bank was a privately German bank. On 26 June 1974, the German Banking Supervisory Office withdrew Herstatt's banking licence after finding that the bank's foreign exchange exposures amounted to three times its capital (BCBS, 2014d). This episode of settlement risk caused heavy losses to other banks, adding a systemic dimension to the individual failure of Herstatt Bank. In response to this turmoil, the central bank governors of the G10 countries established the Basel Committee on Banking Supervision at the end of 1974 with the aim to enhance the financial stability at the global level.

Even if the default of a non-financial institution is a dramatic event for employees, depositors, creditors and clients, the big issue is its impact on the economy. Generally, the failure of a company does not induce a macro-economic stress and is well located to a particular sector or region. For instance, the decade of the 2000s had faced a lot of bankruptcies, e.g. Pacific Gas and Electric Company (2001), Enron (2001), WorldCom (2002), Arthur Andersen (2002), Parmalat (2003), US Airways (2004), Delta Air Lines (2005), Chrysler (2009), General Motors (2009) and LyondellBasell (2009). However, the impact of these failures was contained within the immediate environment of the company and was not spread to the rest of the economy.

In the financial sector, the issue is different because of the interconnectedness between the financial institutions and the direct impact on the economy. And the issue is especially relevant that the list of bankruptcies in finance is long including, for example: Barings Bank (1995); HIH Insurance (2001); Conseco (2002); Bear Stearns (2008), Lehman Brothers (2008); Washington Mutual (2008); DSB Bank (2008). The number of banking and insurance distresses is even more impressive, for example: Northern Rock (2007); Countrywide Financial (2008); Indy Mac Bank (2008); Fannie Mae/Freddie Mac (2008); Merrill Lynch (2008); AIG (2008); Wachovia (2008); Depfa Bank (2008); Fortis (2009); Icelandic banks (2008-2010); Dexia (2011). In Figure 1.4, we report the number of bank failures computed by the Federal Deposit Insurance Corporation (FDIC), the organization in charge of insuring depositors in the US. We can clearly identify three periods of massive defaults[11]: 1935-1942, 1980-1994 and 2008-2014. Each period corresponds to a banking crisis[12] and lasts long because of delayed effects. Whereas the 1995–2007 period is characterized by a low default rate with no default in 2005–2006, there is a significant number of bank defaults these last years (517 defaults between 2008 and 2014).

The Lehman Brothers collapse is a case study for understanding the systemic risk. Lehman Brothers filed for Chapter 11 bankruptcy protection on 15 September 2008 after incurring heavy credit and market risk losses implied by the US subprime mortgage crisis. The amount of losses is generally estimated to be about \$600 bn, because Lehman Brothers had at this time \$640 bn in assets and \$620 bn in debt. However, the cost for the system is far greater than this figure. On equity markets, about \$10 tn went missing in October 2008. The post-Lehman Brothers default period (from September to December 2008) is certainly one of the most extreme liquidity crisis experienced since many decades. This forced central banks to use unconventional monetary policy measures by implementing quantitative easing (QE) programmes. For instance, the Fed now holds more than five times the amount of securities it had prior before September 2008. The collapse of Lehman Brothers had a huge impact on the banking industry, but also on the asset management industry. For instance, four days after the Lehman Brothers bankruptcy, the US government extended temporary guarantee on money market funds. At the same time, the hedge fund industry suffered a lot because of the stress on the financial markets, but also because Lehman Brothers served as prime broker for many hedge funds.

[11] We define these periods when the yearly number of defaults is larger than 15.
[12] They are the Great Depression, the savings and loan crisis of the 1980s and the subprime crisis.

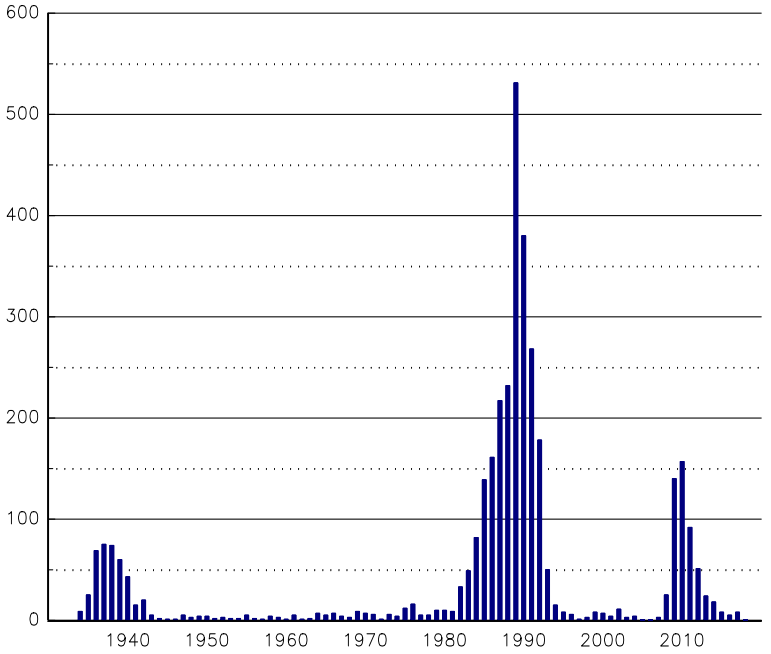


FIGURE 1.4: Number of bank defaults in the US

Source: Federal Deposit Insurance Corporation, Historical Statistics on Banking – Failures & Assistance Transactions, www.fdic.gov/bank/individual/failed.

The 2008 Global Financial Crisis also demonstrated that banks are not the only layer of systemic risk. In fact, a systemic risk implies that the entire financial system is seriously affected, but also participates to the creation of this risk:

> "*[...] there are both old and new components in both the origins and the propagation of the subprime shock. Old components include government financial subsidies for bearing risk, accommodative monetary policy, and adverse selection facilitated by asymmetric information. New components include the central role of agency problems in asset management, the ability of financial institutions to raise new capital from external sources, the activist role of the United States Treasury Department and Federal Reserve, and improvements in U.S. financial system diversification resulting from deregulation, consolidation, and globalization*" (Calomiris, 2009, page 6).

This implies that all financial components, and not only the banking system, can potentially be a source of systemic risk. This is why the bankruptcy of a financial institution cannot be compared to the bankruptcy of a corporate company. Nevertheless, because of the nature of the systemic risk, it is extremely difficult to manage it directly. This explains that the financial supervision is principally a micro-prudential regulation at the firm level. This is only recently that it has been completed by macro-prudential policies in order to mitigate the risk of the financial system as a whole. While the development of risk management was principally due to the advancement of internal models before the 2008 financial crisis, it is now driven by the financial regulation, which completely reshapes the finance industry.

1.2 Financial regulation

The purpose of supervision and regulatory capital has been to control the riskiness of individual banks and to increase the stability of the financial system. As explained in the previous section, it is a hard task whose bounds are not well defined. Among all the institutions that are participating to this work (see Table 1.3), four international authorities have primary responsibility of the financial regulation:

1. The Basel Committee on Banking Supervision (BCBS)
2. The International Association of Insurance Supervisors (IAIS)
3. The International Organization of Securities Commissions (IOSCO)
4. The Financial Stability Board (FSB)

The Basel Committee on Banking Supervision provides a forum for regular cooperation on banking supervisory matters. Its main objective is to improve the quality of banking supervision worldwide. The International Association of Insurance Supervisors is the equivalent of the Basel Committee for the insurance industry. Its goal is to coordinate local regulations and to promote a consistent and global supervision for insurance companies. The International Organization of Securities Commissions is the international body that develops and implements standards and rules for securities and market regulation. While these three authorities are dedicated to a specific financial industry (banks, insurers and markets), the FSB is an international body that makes recommendations about the systemic risk of the global financial system. In particular, it is in charge of defining systemically important financial institutions or SIFIs. Among those different regulators, the BCBS is by far the most active and the banking regulation is certainly the most homogeneous between countries.

These four international bodies define standards at the global level and promote convergence between local supervision. The implementation of the rules is the responsibility of national supervisors or regulators[13]. In the case of the European Union, they are the European Banking Authority (EBA), the European Insurance and Occupational Pensions Authority (EIOPA), the European Securities and Markets Authority (ESMA) and the European System of Financial Supervision (ESFS). A fifth authority, the European Systemic Risk Board (ESRB), completes the European supervision system.

The equivalent authorities in the US are the Board of Governors of the Federal Reserve System, also known as the Federal Reserve Board (FRB), the Federal Insurance Office (FIO) and the Securities and Exchange Commission (SEC). In fact, the financial supervision is more complicated in the US as shown by Jickling and Murphy (2010). The supervisor of banks is traditionally the Federal Deposit Insurance Corporation (FDIC) for federal banks and the Office of the Comptroller of the Currency (OCC) for national banks. However, the Dodd-Frank Act created the Financial Stability Oversight Council (FSOC) to monitor systemic risk. For banks and other financial institutions designated by the FSOC as SIFIs, the supervision is directly done by the FRB. The supervision of markets is shared between the SEC and the Commodity Futures Trading Commission (CFTC), which supervises derivatives trading including futures contracts and options[14].

[13] The regulator is responsible of setting rules and policy guidelines. The supervisor evaluates the safety and soundness of individual banks and verifies that the regulation rules are applied. In Europe, the regulator is EBA while the supervisor is ECB.

[14] A complete list of supervisory authorities by countries are provided on page 28.

TABLE 1.3: The supervision institutions in finance

	Banks	Insurers	Markets	All sectors
Global	BCBS	IAIS	IOSCO	FSB
EU	EBA/ECB	EIOPA	ESMA	ESFS
US	FDIC/FRB	FIO	SEC	FSOC

1.2.1 Banking regulation

The evolution of the banking supervision has highly evolved since the end of the eighties. Here are the principal dates:

1988 Publication of "*International Convergence of Capital Measurement and Capital Standards*", which is better known as "*The Basel Capital Accord*". This text sets the rules of the Cooke ratio.

1993 Development of the Capital Adequacy Directive (CAD) by the European Commission.

1996 Publication of "*Amendment to the Capital Accord to incorporate Market Risks*". This text includes the market risk to compute the Cooke ratio.

2001 Publication of the second consultative document "*The New Basel Capital Accord*" of the Basel II framework.

2004 Publication of "*International Convergence of Capital Measurement and Capital Standards – A Revisited Framework*". This text establishes the Basel II framework.

2006 Implementation of the Basel II framework.

2010 Publication of the Basel III framework.

2013 Beginning of the implementation of the Basel III framework. Its finalization is expected for January 2027.

2017 Finalization of Basel III reforms.

2019 Publication of "*Minimum Capital Requirements for Market Risk*". This is the final version of the Basel III framework for computing the market risk.

This list places the three Basel Accords within a timeframe. However, it gives a misleading image of the banking supervision dynamics. In order to have a better view, we have reported the cumulative number of standards[15] that have been published by the Basel Committee on Banking Supervision in Figure 1.5.

In 1988, the Basel Committee introduced the Cooke ratio[16], which is the minimum amount of capital a bank should maintain in case of unexpected losses. Its goal is to:

- provide an adequation between the capital held by the bank and the risk taken by the bank;
- enhance the soundness and stability of the banking system;
- and reduce the competitive inequalities between banks[17].

[15] They can be found by using the website of the BCBS: `https://www.bis.org/bcbs/publications.htm` and selecting the publication type '*Standards*'.

[16] This ratio took the name of Peter Cooke, who was the Chairman of the BCBS between 1977 and 1988.

[17] This was particularly true between Japanese banks, which were weakly capitalized, and banks in the US and Europe.

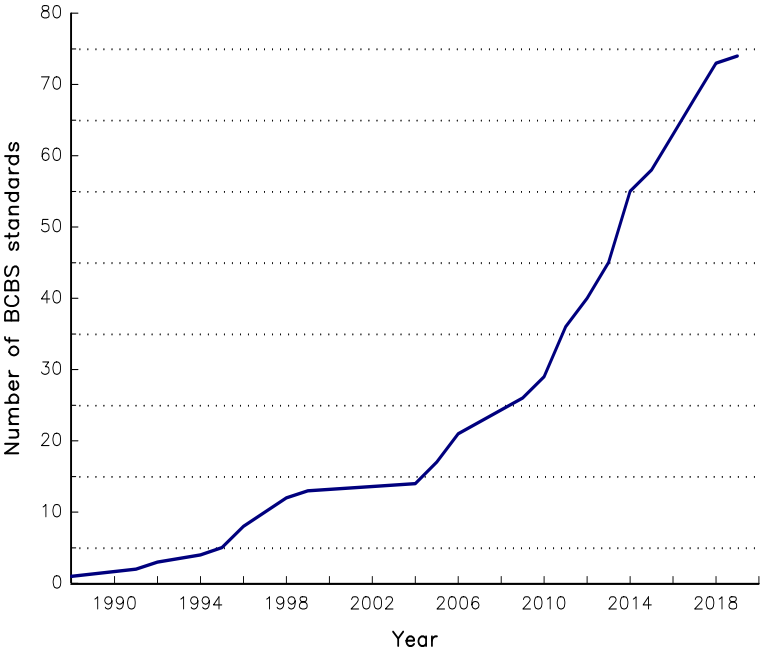


FIGURE 1.5: The huge increase of the number of banking supervision standards

Source: Basel Committee on Banking Supervision and author's calculations.

It is measured as follows:

$$\text{Cooke Ratio} = \frac{C}{\text{RWA}}$$

where C and RWA are the capital and the risk-weighted assets of the bank. A risk-weighted asset is simply defined as a bank's asset weighted by its risk score or risk weight (RW). Because bank's assets are mainly credits, the notional is generally measure by the exposure at default (EAD). To compute risk-weighted assets, we then use the following formula:

$$\text{RWA} = \text{EAD} \cdot \text{RW}$$

The original Basel Accord only considers credit risk and classifies bank's exposures into four categories depending on the value of the risk weights[18] (0%, 20%, 50% and 100%). Concerning off-balance sheet exposures, engagements are converted to credit risk equivalents by multiplying the nominal amount by a credit conversion factor (CCF) and the resulting amounts are risk-weighted according to the nature of the counterparty. Concerning the numerator of the ratio, the Basel Committee distinguishes tier 1 capital and tier 2 capital. Tier 1 capital[19] (or core capital) is composed of (1) common stock (or paid-up share

[18] These categories are defined as follows: (1) cash, gold, claims on OECD governments and central banks, claims on governments and central banks outside OECD and denominated in the national currency are risk-weighted at 0%; (2) claims on all banks with a residual maturity lower than one year, longer-term claims on OECD incorporated banks, claims on public-sector entities within the OECD are weighted at 20%; (3) loans secured on residential property are risk-weighted at 50%; (4) longer-term claims on banks incorporated outside the OECD, claims on commercial companies owned by the public sector, claims on private-sector commercial enterprises are weighted at 100%.

[19] At least 50% of the tier 1 capital should come from the common equity.

capital) and (2) disclosed reserves (or retained earnings), whereas tier 2 capital represents supplementary capital such as[20] (1) undisclosed reserves, (2) asset revaluation reserves, (3) general loan-loss reserves (or general provisions), (4) hybrid debt capital instruments and (5) subordinated debt. The Cooke ratio required a minimum capital ratio of 8% when considering both tier 1 and tier 2 capital, whereas tier 1 capital ratio should be at least half of the total capital or 4%.

Example 1 *The assets of a bank are composed of $100 mn of US treasury bonds, $100 mn of Brazilian government bonds, $50 mn of residential mortgage, $300 mn of corporate loans and $20 mn of revolving credit loans. The bank liability structure includes $25 mn of common stock and $13 mn of subordinated debt.*

For each asset, we compute the RWA by choosing the right risk weight factor. We obtain the following results:

Asset	EAD	RW	RWA
US treasury bonds	100	0%	0
Brazilian Gov. bonds	100	100%	100
Residential mortgage	50	50%	25
Corporate loans	300	100%	300
Revolving credit	20	100%	20
Total			445

The risk-weighted assets of the bank are then equal to $445 mn. We deduce that the capital adequacy ratio is:

$$\text{Cooke Ratio} = \frac{38}{445} = 8.54\%$$

This bank meets the regulatory requirements, because the Cooke ratio is higher than 8% and the tier 1 capital ratio[21] is also higher than 4%. Suppose now that the capital of the bank consists of $13 mn of common stock and $25 mn of subordinated debt. In this case, the bank does not satisfy the regulatory requirements, because the tier 2 capital cannot exceed the tier 1 capital, meaning that the Cooke ratio is equal to 8.54% and the capital tier 1 ratio is equal to 2.92%.

The Basel Accord, which has been adopted by more than 100 countries, has been implemented in the US by the end of 1992 and in Europe in 1993. In 1996, the Basel Committee published a revision of the original Accord by incorporating market risk. This means that banks have to calculate capital charges for market risk in addition to the credit risk. The major difference with the previous approach to measure credit risk is that banks have the choice between two methods for applying capital charges for the market risk:

- the standardized measurement method (SMM);
- the internal model-based approach[22] (IMA).

Within the SMM, the bank apply a fixed capital charge for each asset. The market risk requirement is therefore the sum of the capital charges for all the assets that compose the bank's portfolio. With IMA, the bank estimates the market risk capital charge by computing the 99% value-at-risk of the portfolio loss for a holding period of 10 trading days. From a

[20] The comprehensive definitions and restrictions to define all the elements of capital are defined in Appendix 1 in BCBS (1988).

[21] The tier 1 capital ratio is equal to $25/445 = 5.26\%$.

[22] The use of the internal model-based approach is subject to the approval of the national supervisor.

statistical point of view, the value-at-risk[23] with a confidence level α is defined as the quantile α associated to the probability distribution of the portfolio loss (see Figure 1.6).

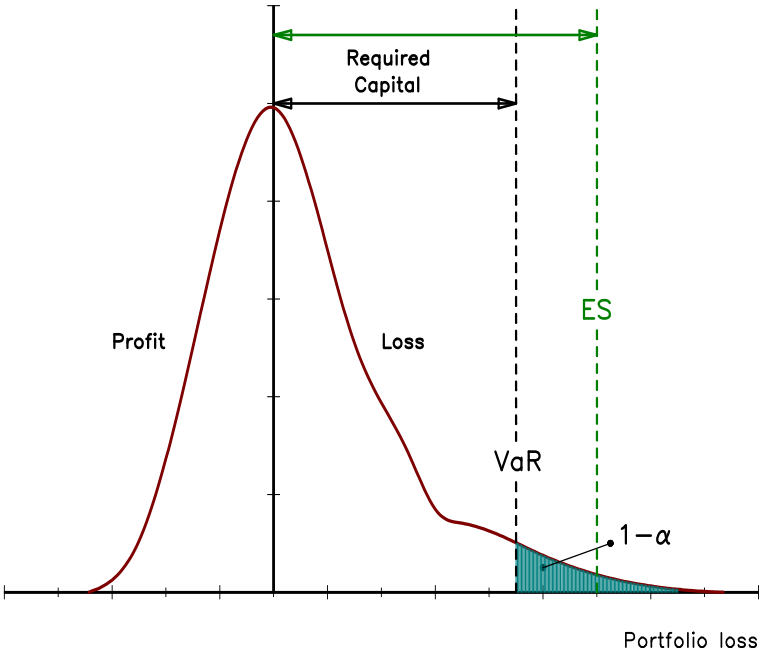


FIGURE 1.6: Probability distribution of the portfolio loss

Another difference with credit risk is that the bank directly computes the market risk capital requirement $\mathcal{K}_{\text{MR}}$ with these two approaches[24]. Therefore, the Cooke ratio becomes[25]:

$$\frac{C_{\text{Bank}}}{\text{RWA} + 12.5 \times \mathcal{K}_{\text{MR}}} \geq 8\%$$

We deduce that:

$$C_{\text{Bank}} \geq \underbrace{8\% \times \text{RWA}}_{\mathcal{K}_{\text{CR}}} + \mathcal{K}_{\text{MR}}$$

meaning that $8\% \times \text{RWA}$ can be interpreted as the credit risk capital requirement $\mathcal{K}_{\text{CR}}$, which can be compared to the market risk capital charge $\mathcal{K}_{\text{MR}}$.

Example 2 *We consider Example 1 and assume that the bank has a market risk on an equity portfolio of* \$25 *mn. The corresponding risk capital charge for a long exposure on a diversified portfolio of stocks is equal to* 12%. *Using its internal model, the bank estimates that the* 99% *quantile of the portfolio loss is equal to* \$1.71 *mn for a holding period of* 10 *days.*

[23] In the Basel III framework, the expected shortfall, which is defined as the average loss beyond the value-at-risk, replaces the value-at-risk for computing the market risk.

[24] We use the symbols C and $\mathcal{K}$ in order to make the distinction between the capital of the bank and the regulatory capital requirement.

[25] When considering market risk, the total capital may include tier 3 capital, consisting of short-term subordinated debt with an original maturity of at least 2 years.

In the case of the standardized measurement method, the market risk capital requirement is equal to \$3 mn[26]. The capital ratio becomes:

$$\text{Cooke Ratio} = \frac{25}{445 + 12.5 \times 3} = 7.88\%$$

In this case, the bank does not meet the minimum capital requirement of 8%. If the bank uses its internal model, the Cooke ratio is satisfied:

$$\text{Cooke Ratio} = \frac{25}{445 + 12.5 \times 1.71} = 8.15\%$$

The Basel Accord has been highly criticized, because the capital charge for credit risk is too simplistic and too little risk sensitive: limited differentiation of credit risk, no maturity, granularity of risk weights, etc. These resulted in regulatory arbitrage through the use of securitization between assets with same regulatory risk but different economic risk. In June 1999, the Basel Committee produced an initial consultative document with the objective to replace the 1988 Accord by a new capital adequacy framework. This paper introduces some features about Basel II, but this is really the publication of the second consultative paper in January 2001 that marks a milestone for the banking regulation. Indeed, the 2001 publication is highly detailed and comprehensive, and the implementation of this new framework seemed very complex at that time. The reaction of the banking industry was negative and somehow hostile at the beginning, in particular because the Basel Committee introduced a third capital charge for operational risk besides credit and market risks and the implementation costs were very high. It has taken a long time until the Basel Committee and the banking industry converge to an accord. Lastly, the finalized Basel II framework is published in June 2004.

TABLE 1.4: The three pillars of the Basel II framework

Pillar 1 Minimum Capital Requirements	Pillar 2 Supervisory Review Process	Pillar 3 Market Discipline
Credit risk Market risk Operational risk	Review & reporting Capital above Pillar 1 Supervisory monitoring	Capital structure Capital adequacy Models & parameters Risk management

As illustrated in Table 1.4, the new Accord consists of three pillars:

1. the first pillar corresponds to *minimum capital requirements*, that is, how to compute the capital charge for credit risk, market risk and operational risk;
2. the second pillar describes the *supervisory review process*; it explains the role of the supervisor and gives the guidelines to compute additional capital charges for specific risks, which are not covered by the first pillar;

[26] We have:

$$\mathcal{K}_{\text{MR}} = 12\% \times 25 = 3$$

3. the *market discipline* establishes the third pillar and details the disclosure of required information regarding the capital structure and the risk exposures of the bank.

Regarding the first pillar, the Cooke ratio becomes:

$$\frac{C_{\text{Bank}}}{\text{RWA} + 12.5 \times \boldsymbol{\mathcal{K}}_{\text{MR}} + 12.5 \times \boldsymbol{\mathcal{K}}_{\text{OR}}} \geq 8\%$$

where $\boldsymbol{\mathcal{K}}_{\text{OR}}$ is the capital charge for operational risk. This implies that the required capital is directly computed for market risk and operational risk whereas credit risk is indirectly measured by risk-weighted assets[27].

Example 3 *We assume that the risk-weighted assets for the credit risk are equal to \$500 mn, the capital charge for the market risk is equal to \$10 mn and the capital charge for the operational risk is equal to \$3 mn.*

We deduce that the required capital for the bank is:

$$\begin{aligned}
\boldsymbol{\mathcal{K}} &= 8\% \times (\text{RWA} + 12.5 \times \boldsymbol{\mathcal{K}}_{\text{MR}} + 12.5 \times \boldsymbol{\mathcal{K}}_{\text{OR}}) \\
&= 8\% \times \text{RWA} + \boldsymbol{\mathcal{K}}_{\text{MR}} + \boldsymbol{\mathcal{K}}_{\text{OR}} \\
&= 8\% \times 500 + 10 + 3 \\
&= \$53 \text{ mn}
\end{aligned}$$

This implies that credit risk represents 75.5% of the total risk.

With respect to the original Accord, the Basel Committee did not change the market risk approach whereas it profoundly changed the methods to compute the capital charge for the credit risk. Two approaches are proposed:

- The standardized approach (SA)
  This approach, which is more sensitive than Basel I, is based on external ratings provided by credit rating agencies. The capital charge is computed by considering a mapping function between risk weights and credit ratings.

- The internal ratings-based approach (IRB)
  This approach can be viewed as an external risk model with internal and external risk parameters. The key parameter is the default probability of the asset, which is deduced from the internal credit rating model of the bank. The Basel Committee makes the distinction between two methods. In the foundation IRB (FIRB), the bank only estimates the probability of default and uses standard values for the other risk parameters of the model. In the advanced IRB (AIRB), the bank may estimate all the risk parameters.

Regarding operational risk, the Basel Committee propose three approaches to compute the required capital:

- The Basic Indicator Approach (BIA)
  In this case, the capital charge is a fixed percentage of the gross income.

- The Standardized Approach (TSA)
  This method consists of dividing bank's activities into eight business lines. For each business line, the capital charge is a fixed percentage β of its gross income. The parameter β depends on the riskiness of the business line. The total capital is the sum of the eight regulatory capital charges.

[27] In fact, we can define risk-weighted assets for each category of risk. We have the following relationships $\text{RWA}_{\mathcal{R}} = 12.5 \times \boldsymbol{\mathcal{K}}_{\mathcal{R}}$ and $\boldsymbol{\mathcal{K}}_{\mathcal{R}} = 8\% \times \text{RWA}_{\mathcal{R}}$ where $\boldsymbol{\mathcal{K}}_{\mathcal{R}}$ is the required capital for the risk $\mathcal{R}$. The choice of defining either $\text{RWA}_{\mathcal{R}}$ or $\boldsymbol{\mathcal{K}}_{\mathcal{R}}$ is a mere convention.

- Advanced Measurement Approaches (AMA)
  In this approach, the bank uses a statistical model with internal data for estimating the total capital.

A summary of the different options is reported in Figure 1.7.

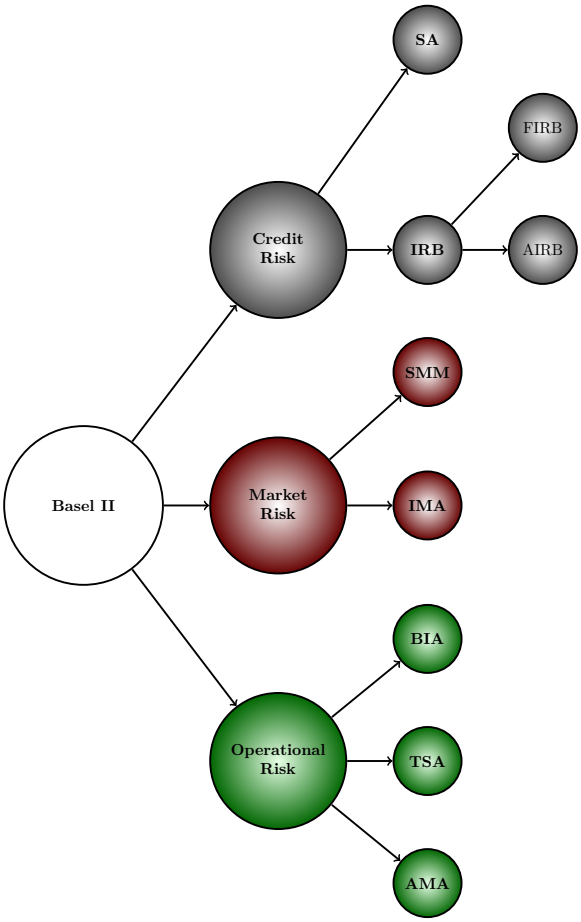


FIGURE 1.7: Minimum capital requirements in the Basel II framework

The European Union has adopted the Basel II framework in June 2006 with the capital requirements directive[28] (CRD). In the United States, Basel II is partially applied since 2006 and only concerns the largest banking institutions (Getter, 2014). Since the 2004 publication, more than 40 countries have fully implemented Basel II (Hong Kong in January 2007, Japan in March 2007, Canada in November 2007, South Korea in December 2007, Australia in January 2008, South Africa in January 2008, etc.). However, the subprime crisis in 2007 and the collapse of Lehman Brothers in September 2008 illustrated the limits of the New Accord concerning the issues of leverage and liquidity. In response to the financial market crisis, the Basel Committee enhances then the New Accord by issuing a set of documents between 2009 and 2010. In July 2009, the Basel Committee approved a package of measures to strengthen the rules governing trading book capital, particularly the market risk associated

[28] It replaces CAD II (or the 98/31/EEC directive), which is the revision of the original CAD and incorporates market risk.

to securitization and credit-related products. Known as the Basel 2.5 framework, these new rules can be summarized into four main elements, which are:

1. the incremental risk charge (IRC), which is an additional capital charge to capture default risk and migration risk for unsecuritized credit products;
2. the stressed value-at-risk requirement (SVaR), which is intended to capture stressed market conditions;
3. the comprehensive risk measure (CRM), which is an estimate of risk in the credit correlation trading portfolio (CDS baskets, CDO products, etc.);
4. new standardized charges on securitization exposures, which are not covered by CRM.

In addition to these elements affecting the first pillar, the Basel Committee also expands the second pillar (largest exposures and risk concentrations, remuneration policies, governance and risk management) and enhances the third pillar (securitization and re-securitization exposures). The coming into force of Basel 2.5 was December 2011 in the European Union[29] and January 2013 in the United States (BCBS, 2015b).

In December 2010, the Basel Committee published a new regulatory framework in order to enhance risk management, increase the stability of the financial markets and improve the banking industry's ability to absorb macro-economic shocks. The Basel III framework consists of micro-prudential and macro-prudential regulation measures concerning;

- a new definition of the risk-based capital;
- the introduction of a leverage ratio;
- the management of the liquidity risk.

The capital is redefined as follows. Tier 1 capital is composed of common equity tier 1 capital (common equity and retained earnings or CET1) and additional tier 1 capital (AT1). The new capital ratios are 4.5% for CET1, 6% for tier 1 and 8% for total capital (T1 + T2). Therefore, Basel III gives preference to tier 1 capital rather than tier 2 capital whereas the tier 3 risk capital is eliminated. BCBS (2010) introduced also a surplus of CET1, which is "*designed to ensure that banks build up capital buffers outside periods of stress which can be drawn down as losses are incurred*". This capital conservation buffer (CB), which is equal to 2.5% of RWA, applies at all the times outside periods of stress. The aim is to reduce the distribution of earnings and to support the business of bank through periods of stress. A macro-prudential approach completes capital requirements by adding a second capital buffer called the countercyclical capital buffer (CCB). During periods of excessive credit growth, national authorities may require an additional capital charge between 0% and 2.5%, which increases the CET1 ratio until 9.5% (including the conservation buffer). The underlying idea is to smooth the credit cycle, to reduce the procyclicality and to help banks to provide credit during bad periods of economic growth. The implementation of this new framework is progressive from April 2013 until March 2019. A summary of capital requirements[30] and transitional periods is given in Table 1.5.

This new definition of the capital is accompanied by a change of the required capital for counterparty credit risk (CCR). In particular, BCBS (2010) adds a credit valuation

[29] The Basel 2.5 framework was adopted in two stages: CRD II (or the 2009/111/EC directive) in November 2009 and CRD III (or the 2010/76/EU directive) in December 2010.

[30] Basel III defines a third capital buffer for systemic banks, which can vary between 1% and 3.5%. This topic will be presented later on the paragraph dedicated to systemically important financial institutions on page 26.

TABLE 1.5: Basel III capital requirements

Capital ratio	2013	2014	2015	2016	2017	2018	2019
CET1	3.5%	4.0%	4.5%				**4.5**%
CB				0.625%	1.25%	1.875%	**2.5**%
CET1 + CB	3.5%	4.0%	4.5%	5.125%	5.75%	6.375%	**7.0**%
Tier 1	4.5%	5.5%	6.0%				**6.0**%
Total	8.0%						**8.0**%
Total + CB	8.0%			8.625%	9.25%	9.875%	**10.5**%
CCB	**0**% − **2.5**%						

Source: Basel Committee on Banking Supervision, www.bis.org/bcbs/basel3.htm.

adjustment charge (CVA) for OTC derivative trades. CVA is defined as the market risk of losses caused by changes in the credit spread of a counterparty due to changes in its credit quality. It also corresponds to the market value of counterparty credit risk.

Basel III also includes a leverage ratio to prevent the build-up of excessive on- and off-balance sheet leverage in the banking sector. BCBS (2014a) defines this ratio as follows:

$$\text{Leverage ratio} = \frac{\text{Tier 1 capital}}{\text{Total exposures}} \geq 3\%$$

where the total exposures is the sum of on-balance sheet exposures, derivative exposures and some adjustments concerning off-balance sheet items. The leverage ratio can be viewed as the second macro-prudential measure of Basel III. Indeed, during credit boom, we generally observe compression of risk weight assets and a growth of the leverage, because the number of profitable projects increases during economic good times. For instance, Brei and Gambacorta (2014) show that the Basel III leverage ratio is negatively correlated with GDP or credit growth. By introducing a floor value, the Basel Committee expects that the leverage ratio will help to reduce the procyclicality like the countercyclical capital buffer.

The management of the liquidity is another important issue of Basel III. The bankruptcy of Lehman Brothers was followed by a lack of liquidity, which is one of the main sources of systemic risk. For instance, Brunnermeier and Pedersen (2009) demonstrated that a liquidity dry-up event arising from a fight-to-quality environment can result in runs, fire sales, and asset liquidations in general transforming the market into a contagion mechanism. In order to prevent such events, the Basel Committee proposed several liquidity rules and introduced in particular two liquidity ratios: the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR). The objective of the LCR is to promote short-term resilience of the bank's liquidity risk profile. It is expressed as:

$$\text{LCR} = \frac{\text{HQLA}}{\text{Total net cash outflows}} \geq 100\%$$

where HQLA is the stock of high quality liquid assets and the denominator is the total net cash outflows over the next 30 calendar days. Therefore, the LCR is designed to ensure that the bank has the necessary assets to face a one-month stressed period of outflows. On the contrary, NSFR is designed in order to promote long-term resilience of the bank's liquidity profile. It is defined as the amount of available stable funding (ASF) relative to the amount of required stable funding (RSF):

$$\text{NSFR} = \frac{\text{Available amount of stable funding}}{\text{Required amount of stable funding}} \geq 100\%$$

The amount of available stable funding is equal to the regulatory capital[31] plus the other liabilities to which we apply a scaling factor between 0% and 100%. The amount of required stable funding is the sum of two components: risk-weighted assets and off-balance sheet exposures.

The implementation of Basel III was due to January 2013, but some countries have delayed the adoption of the full package. According to BCBS (2015b), the rules for risk-based capital are more adopted than those concerning the liquidity ratio or the leverage ratio. In the US, the rules for risk-based capital and the leverage ratio are effective since January 2014, while the LCR rule came into effect in January 2015. In the European Union, the Basel III agreement is transposed on July 2013 into two texts: the CRD IV (or the 2013/36/EU directive) and the capital requirements regulation (CRR) (or the 575/2013 EU regulation). Therefore, Basel III is effective since January 2014 for the rules of risk-based capital and leverage ratio and October 2015 for the LCR rule.

Even before Basel III is fully implemented, the Basel Committee has published a set of consultative documents, which has been viewed as the basis of a future Basel IV Accord. The guiding principle of these works is to simplify the different approaches to compute the regulatory capital and to reduce the risk of arbitrage between standardized and advanced methods. These new proposals concern review of the market risk measurement (BCBS, 2013b, 2014h, 2016a), revision to the standardized approach for credit (BCBS, 2015d) and operational risks (BCBS, 2014f, 2016b), minimum capital requirements for interest rate risk in the banking book (BCBS, 2016d) and a modified framework for the CVA risk (BCBS, 2015c). Finally, the Basel Committee created in 2017 a surprise by announcing that all these reforms correspond to the finalization of the Basel III Accord. The changes are very significant. For instance, it replaces the VaR measure by the expected shortfall measure. The risk weight of residential real estate exposures will depend on the loan-to-value (LTV) ratio. It also imposes some constraints on the use of internal credit risk models, in particular the remove of the IRB approach for bank, large corporate and equity exposures. CVA requirements will be based on two approaches: SA-CVA and BA-CVA. For counterparty credit risk, the IMM-CCR method will be constrained by a floor with respect to the SA-CCR method. In the case of operational risk, the three approaches (BIA, TSA and AMA) are replaced by a unique approach called the Standardized Measurement Approach (SMA). For market risk, the boundary between trading book and banking book is changed, and the standard approach is fully revisited and is based on risk sensitivities. Finally, the interest rate risk of the banking book continues to be monitored in Pillar 2, but its measure is highly reinforced.

1.2.2 Insurance regulation

Contrary to the banking industry, the regulation in insurance is national. The International Association of Insurance Supervisors (IAIS) is an association to promote globally consistent supervision. For that, the IAIS is responsible for developing principles and standards, which form the Insurance Core Principles (ICP). For instance, the last release of ICP was in November 2018 and contained 26 ICPs[32]. However, its scope of intervention is more limited than this of the BCBS. In particular, the IAIS does not produce any methodologies of risk management or formula to compute risk-based capital. In Europe, the regulatory framework is the Solvency II directive (or the 2009/138/EC directive), which harmonizes the insurance regulation and capital requirements in the European Union. In the US, the

[31] Excluding tier 2 instruments with residual maturity of less than one year.

[32] ICP 1 concerns the objectives, powers and responsibilities of the supervisor, ICP 17 is dedicated to capital adequacy, ICP 24 presents the macro-prudential surveillance and insurance supervision, etc.

supervisor is the National Association of Insurance Commissioners (NAIC). In 2008, it has created a Solvency Modernization Initiative (SMI) in order to reform the current framework in the spirit of Solvency II. However, the convergence across the different jurisdictions is far to being reached.

Solvency I (or the 2002/13/EC directive) is a set of rules to define the insurance solvency regime and was put in place on January 2004 in the European Union. It defined how an insurance company should calculate its liabilities and the required capital. In this framework, the capital is the difference between the book value of assets and the technical provisions (or insurance liabilities). This capital is decomposed in the solvency capital requirement (or SCR) and the surplus (see Figure 1.8). One of the main drawbacks of Solvency I is that assets and liabilities are evaluated using an accounting approach (historical or amortized cost).

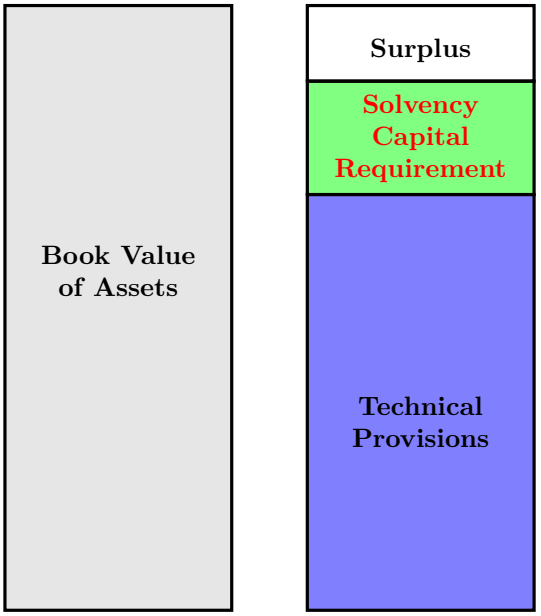


FIGURE 1.8: Solvency I capital requirement

In an address to the European Insurance Forum 2013, Matthew Elderfield, Deputy Governor of the Central Bank of Ireland, justifies the reform of the insurance regulation in Europe as follows:

> "*[...] it is unacceptable that the common regulatory framework for insurance in Europe in the 21st-century is not risk-based and only takes account, very crudely, of one side of the balance sheet. The European Union urgently needs a new regulatory standard which differentiates solvency charges based on the inherent risk of different lines of business and which provides incentives for enhanced risk management. It urgently needs a framework that takes account of asset risks in an insurance company. It urgently needs a framework that encourages better governance and management of risk. And it urgently needs a framework that provides better disclosure to market participants*" (Elderfield, 2013, page 1).

With Solvency II, capital requirements are then based on an economic valuation of the insurer balance sheet, meaning that:

- assets are valued at their market value;
- liabilities are valued on a best estimate basis.

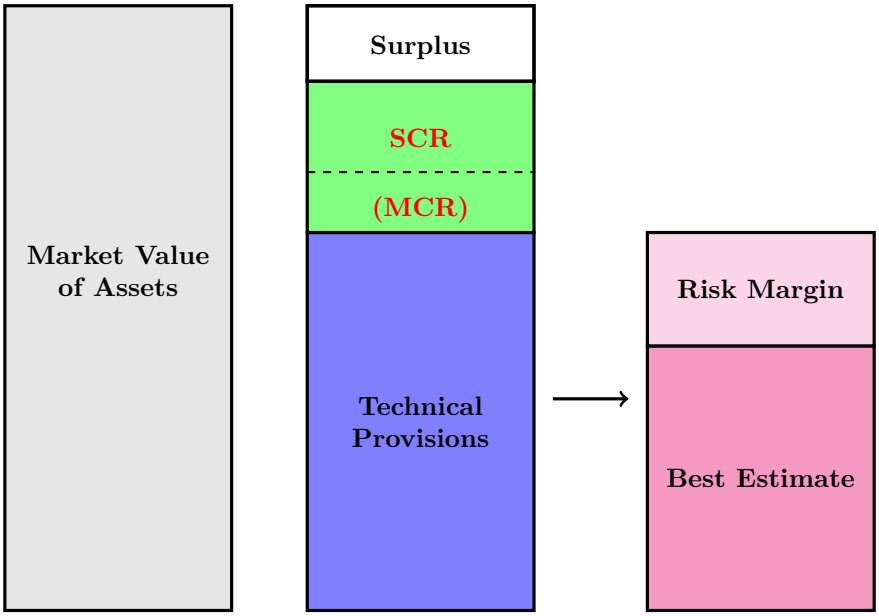


FIGURE 1.9: Solvency II capital requirement

In this framework, the economic value of liabilities corresponds to the expected present value of the future cash flows. Technical provisions are then the sum of the liabilities best estimate and a risk margin (or prudence margin) in order to take into account non-hedgeable risk components. Solvency II defines two levels of capital requirements. The minimum capital requirement (MCR) is the required capital under which risks are considered as being unacceptable. The solvency capital requirement (SCR) is the targeted required capital (SCR $\geq$ MCR). The underlying idea is to cover the different source of risk at a 99.5% confidence level[33] for a holding period of one year. The insurance company may opt for the standard formula or its own internal model for computing the required capital. In the case of the standard formula method, the SCR of the insurer is equal to:

$$\text{SCR} = \sqrt{\sum_{i,j}^{m} \rho_{i,j} \cdot \text{SCR}_i \cdot \text{SCR}_j} + \text{SCR}_{\text{OR}}$$

where SCR_i is the SCR of the risk module i, SCR_{OR} is the SCR associated to the operational risk and $\rho_{i,j}$ is the correlation factor between risk modules i and j. Solvency II considers several risk components: underwriting risk (non-life, life, health, etc.), market risk, default and counterpart credit risk[34]. For each risk component, a formula is provided to compute the SCR of the risk factors. Regarding the capital C, own funds are classified into basic own funds and ancillary own funds. The basic own funds consist of the excess of assets over

[33] It is set to 85% for the MCR.

[34] Solvency II is an ambitious and complex framework because it mixes both assets and liabilities, risk management and ALM.

liabilities, and subordinated liabilities. The ancillary own funds correspond to other items which can be called up to absorb losses. Examples of ancillary own funds are unpaid share capital or letters of credit and guarantees. Own funds are then divided into tiers depending on their permanent availability and subordination. For instance, tier 1 corresponds to basic own funds which are immediately available and fully subordinated. The solvency ratio is then defined as:

$$\text{Solvency Ratio} = \frac{C}{\text{SCR}}$$

This solvency ratio must be larger than 33% for tier 1 and 100% for the total own funds.

The quantitative approach to compute MCR, SCR and the technical provisions define Pillar 1 (Figure 1.9). As in Basel II framework, it is completed by two other pillars. Pillar 2 corresponds to the governance of the solvency system and concerns qualitative requirements, rules for supervisors and own risk and solvency assessment (ORSA). Pillar 3 includes market disclosures and also supervisory reporting.

1.2.3 Market regulation

Banks and insurers are not the only financial institutions that are regulated and the financial regulatory framework does not reduce to Basel III and Solvency II. In fact, a whole variety of legislation measures helps to regulate the financial market and the participants.

In Europe, the markets in financial instruments directive or MiFID[35] came in force since November 2007. Its goal was to establish a regulatory framework for the provision of investment services in financial instruments (such as brokerage, advice, dealing, portfolio management, underwriting, etc.) and for the operation of regulated markets by market operators. The scope of application concerns various aspects such as passporting, client categorization (retail/professional investor), pre-trade and post-trade transparency or best execution procedures. In August 2012, MiFID is completed by the European market infrastructure regulation (EMIR), which is specifically designed to increase the stability of OTC derivative markets by promoting central counterparty clearing and trade repositories. In June 2014, MiFID is revised (MiFID 2) and the regulation on markets in financial instruments (MiFIR) replaces EMIR. According to ESMA[36], this supervisory framework concerns 104 European regulated markets at the date of May 2015. On April 2014, the European parliament completes the framework by publishing new rules to protect retail investors (packaged retail and insurance-based investment products or PRIIPS). These rules complete the various UCITS directives, which organize the distribution of mutual funds in Europe.

In the US, the regulation of the market dates back to the 1930s:

- The Securities Act of 1933 concerns the distribution of new securities.
- The Securities Exchange Act of 1934 regulates trading securities, brokers, and exchanges, whereas the Commodity Exchange Act regulates the trading of commodity futures.
- The Trust Indenture Act of 1939 defines the regulating rules for debt securities.
- The Investment Company Act of 1940 is the initial regulation framework of mutual funds.
- The Investment Advisers Act of 1940 is dedicated to investment advisers.

[35] It corresponds to the 2004/39/EC directive.
[36] See the website www.esma.europa.eu/databases-library/registers-and-data.

At the same time, the Securities and Exchange Commission (SEC) was created to monitor financial markets (stocks and bonds). Now, the area of SEC supervision is enlarged and concerns stock exchanges, brokers, mutual funds, investment advisors, some hedge funds, etc. In 1974, the Commodities Futures Trading Commission Act established the Commodity Futures Trading Commission (CFTC) as the supervisory agency responsible for regulating the trading of futures contracts. The market regulation in the US has not changed significantly until the 2008 Global Financial Crisis (GFC). In 2010, President Barack Obama signed an ambitious federal law, the Dodd-Frank Wall Street Reform and Consumer Protection Act also named more simply Dodd-Frank, which is viewed as a response to the crisis. This text has an important impact on various areas of regulation (banking, market, investors, asset managers, etc.). It also introduces a new dimension in regulation. It concerns the coordination among regulators with the creation of the Financial Stability Oversight Council (FSOC), whose goal is to monitor the systemic risk.

1.2.4 Systemic risk

The 2008 financial crisis has an unprecedent impact on the financial regulation. It was responsible for Basel III, Dodd-Frank, Volcker rule, etc., but it has also inspired new considerations on the systemic risk. Indeed, the creation of the Financial Stability Board (FSB) in April 2009 was motivated to establish an international body that monitors and makes recommendations about the global financial system, and especially the associated systemic risk. Its area of intervention covers not only banking and insurance, but also all the other financial institutions including asset managers, finance companies, market intermediaries, investors, etc.

The main task of the FSB is to develop assessment methodologies for defining systemically important financial institutions (SIFIs) and to make policy recommendations for mitigating the systemic risk of the financial system. According to FSB (2010), SIFIs are institutions whose "*distress or disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity*". By monitoring SIFIs in a different way than other financial institutions, the objective of the supervisory authorities is obviously to address the '*too big too fail*' problem. A SIFI can be global (G-SIFI) or domestic (D-SIFI). The FSB also distinguishes between three types of G-SIFIs:

1. G-SIBs correspond to global systemically important banks.
2. G-SIIs designate global systemically important insurers.
3. The third category is defined with respect to the two previous ones. It incorporates other SIFIs than banks and insurers (non-bank non-insurer global systemically important financial institutions or NBNI G-SIFIs).

The FSB/BCBS framework for identifying G-SIBs is a scoring system based on five categories: size, interconnectedness, substitutability/financial institution infrastructure, complexity and cross-jurisdictional activity (BCBS, 2014g). In November 2018, there were 29 G-SIBs (FSB, 2015b). Depending on the score value, the bank is then assigned to a specific bucket, which is used to calculate the higher loss absorbency (HLA) requirement. This additional capital requirement is part of the Basel III framework and ranges from 1% to 3.5% common equity tier 1. According to FSB (2018b), the most systemically important bank is JPMorgan Chase, which is assigned to an additional capital buffer of 2.5% CET1. This means that the total capital for this banks can go up to 15.5% with the following

decomposition: tier 1 = 6.0%, tier 2 = 2.0%, conservation buffer = 2.5%, countercyclical buffer = 2.5% and systemic risk capital = 2.5%.

For insurers, the assessment methodology is close to the methodology for G-SIBs and is based on five categories: size, global activity, interconnectedness, non-traditional insurance and non-insurance activities and substitutability (IAIS, 2013a). However, this quantitative approach is completed by a qualitative analysis and the final list of G-SIIs is the result of the IAIS supervisory judgment. In November 2015, there were 9 G-SIIs (FSB, 2015c). The associated policy measures are documented in IAIS (2013b) and consist of three main axes: recovery and resolution planning requirements, enhanced supervision and higher loss absorbency requirements.

Concerning NBNI SIFIs, FSB and IOSCO are still in a consultation process in order to finalize the assessment methodologies (FSB, 2015a). Indeed, the second consultation paper considers three categories of participants in the financial sectors that it identifies as potential NBNI SIFIs:

1. finance companies;
2. market intermediaries, especially securities broker-dealers;
3. investment funds, asset managers and hedge funds.

The final assessment methodology was planned for the end of 2015, but it has never been published until now. However, the fact that the FSB already considers that there are other SIFIs than banks and insurers suggests that financial regulation will be strengthened for many financial institutions including the three previous categories but also other financial institutions such as pension funds, sovereign wealth funds, etc.

The identification of SIFIs is not the only task of the FSB. The other important objective is to monitor the shadow banking system and to understand how it can pose systemic risk. The shadow banking system can be described as "*credit intermediation involving entities and activities outside the regular banking system*" (FSB, 2011). It is also called non-bank credit intermediation. The shadow banking system may expose the traditional banking system to systemic risk, because they may be spill-over effects between the two systems. Moreover, shadow banking entities (SBEs) are not subject to tight regulation like banks. However, it runs bank-like activities such as maturity transformation, liquidity transformation, leverage and credit risk transfer. Examples of shadow banking are for instance money market funds, securitization, securities lending, repos, etc. The task force formed by the FSB follows a three-step process:

- the first step is to scan and map the overall shadow banking system and to understand its risks;
- the second step is to identify the aspects of the shadow banking system posing systemic risk or regulatory arbitrage concerns;
- the last step is to assess the potential impact of systemic risk induced by the shadow banking system.

Even if this process is ongoing, shadow banking regulation can be found in Dodd-Frank or 2015 consultation paper of the EBA. However, until now regulation is principally focused on money market funds.

1.3 Financial regulation overview

1.3.1 List of supervisory authorities

We use the following correspondence: **B** for banking supervision, **I** for insurance supervision, **M** for market supervision and **S** for systemic risk supervision.

International authorities

BCBS Basel Committee on Banking Supervision; www.bis.org/bcbs; **B**
FSB Financial Stability Board; www.fsb.org; **S**
IAIS International Association of Insurance Supervisors; www.iaisweb.org; **I**
IOSCO International Organization of Securities Commissions; www.iosco.org; **M**

European authorities

EBA European Banking Authority; eba.europa.eu; **B**
ECB/SSM European Central Bank/Single Supervisory Mechanism; www.bankingsupervision.europa.eu; **B**
EIOPA European Insurance and Occupational Pensions Authority; eiopa.europa.eu; **I**
ESMA European Securities and Markets Authority; www.esma.europa.eu; **M**
ESRB European Systemic Risk Board; www.esrb.europa.eu; **S**

US authorities

CFTC Commodity Futures Trading Commission; www.cftc.gov; **M**
FRB Federal Reserve Board; www.federalreserve.gov/supervisionreg.htm; **B/S**
FDIC Federal Deposit Insurance Corporation; www.fdic.gov; **B**
FIO Federal Insurance Office; home.treasury.gov/policy-issues/financial-markets-financial-institutions-and-fiscal-service/federal-insurance-office; **I**
FSOC Financial Stability Oversight Council; home.treasury.gov/policy-issues/financial-markets-financial-institutions-and-fiscal-service/fsoc; **S**
OCC Office of the Comptroller of the Currency; www.occ.gov; **B**
SEC Securities and Exchange Commission; www.sec.gov; **M**

Some national authorities

Canada

CSA Canadian Securities Administrators; www.securities-administrators.ca; **M**
OSFI Office of the Superintendent of Financial Institutions; www.osfi-bsif.gc.ca; **B/I**
IIROC Investment Industry Regulatory Organization of Canada; www.iiroc.ca; **M**

China

CBRC China Banking Regulatory Commission; www.cbrc.gov.cn; **B**
CIRC China Insurance Regulatory Commission; www.circ.gov.cn; **I**
CSRC China Securities Regulatory Commission; www.csrc.gov.cn; **M**

France

AMF	Autorité des Marchés Financiers; www.amf-france.org; **M**
ACPR	Autorité de Contrôle Prudentiel et de Résolution; acpr.banque-france.fr; **B/I**

Germany

BAFIN	Bundesanstalt für Finanzdienstleistungsaufsicht; www.bafin.de; **B/I/M**

Italy

BdI	Banca d'Italia; www.bancaditalia.it; **B**
CONSOB	Commissione Nazionale per le Società e la Borsa; www.consob.it; **M**
IVASS	Istituto per la Vigilanza sulle Assicurazioni; www.ivass.it; **I**

Japan

FSA	Financial Services Agency; www.fsa.go.jp; **B/I/M**

Luxembourg

CAA	Commissariat aux Assurances; www.caa.lu; **I**
CSSF	Commission de Surveillance du Secteur Financier; www.cssf.lu; **B/M**

Spain

BdE	Banco de España; www.bde.es; **B**
CNMV	Comisión Nacional del Mercado de Valores; www.cnmv.es; **M**
DGS	Dirección General de Seguros y Pensiones; www.dgsfp.mineco.es; **I**

Switzerland

FINMA	Swiss Financial Market Supervisory Authority; www.finma.ch; **B/I/M**

United Kingdom

FCA	Financial Conduct Authority; www.fca.org.uk; **M**
PRA	Prudential Regulation Authority; www.bankofengland.co.uk/prudential-regulation; **B/I**

1.3.2 Timeline of financial regulation

In this section, we give the major dates which marketed the important stages of the financial regulation. We can consider four periods: before 1980, the years 1980 – 2000, the period until the 2008 Global Financial Crisis and the last 10 years.

Before 1980

Before 1980, the financial regulation is mainly developed in the US with several acts, which are voted in after the Great Depression in the 1930s. These acts concerns a wide range of financial activities, in particular banking, markets and investment sectors. The Basel Committee on Banking Supervision was established in 1974. In Europe, two directives established a regulatory framework for insurance companies.

Banking Regulation	1913	Federal Reserve Act (establishment of the Federal Reserve System as the central banking system of the US)
	1933	Glass-Steagall Act (separation of commercial and investment banking in the US)
	1933	US Banking Act (creation of FDIC and insurance deposit)
BCBS	1974	Creation of the Basel Committee on Banking Supervision
Solvency I	1973-07-24	Publication of the non-life insurance directive (73/239/EEC) dedicated to solvency margin requirements
	1979-03-05	Publication of the life insurance directive (79/267/EEC) dedicated to solvency margin requirements
Market Regulation	1933-05-27	Securities Act (registration and prospectus of securities)
	1934-06-06	Securities Exchange Act (regulation of the secondary markets and creation of the SEC)
	1936-06-15	Commodity Exchange Act (regulation of the commodity futures)
	1939-08-03	Trust Indenture Act (regulation of debt securities)
	1940-08-22	Investment Advisers Act (regulation of investment advisers)
	1940-08-22	Investment Company Act (regulation of mutual funds)
	1974-10-23	Commodity Futures Trading Commission Act (the CFTC replaces the Commodity Exchange Commission)

The years 1980 – 2000

The years 1980 – 2000 were marked by the development of the banking regulation and the publication of the Basel Accord dedicated to credit risk. Moreover, the end of the 1990s saw the implementation of the regulatory framework concerning market risks. In Europe, the UCITS directive is also an important step concerning the investment industry. In the US, the insurance regulation is reformed with the risk-based capital framework whereas Solvency I is reinforced in Europe.

Basel I	1987-12-15	Publication of the consultative paper on the Cooke ratio
	1988-07-04	Publication of the Basel Capital Accord
	1996-01-18	Publication of the amendment to incorporate market risks
CAD	1993-03-15	Publication of the Capital Adequacy Directive (93/6/EEC) known as CAD I
	1998-06-22	Revision of the CAD (98/31/EEC) known as CAD II
Solvency I	1988-06-22	Second non-life insurance directive 88/357/EEC
	1990-11-08	Second life insurance directive 90/619/EEC
	1992-06-18	Third non-life insurance directive 92/49/EEC
	1992-11-10	Third life insurance directive 92/96/EEC
RBC	1990	NAIC created the US RBC regime
	1992	Implementation of RBC in US insurance
	1993	Finalization of the RBC formula for life insurance
	1994	Finalization of the RBC formula for property and casuality insurance
	1998	Finalization of the RBC formula for health insurance
Market Regulation	1985-12-20	Publication of the first UCITS Directive (85/611/EEC)
	2000-12-14	Commodity Futures Modernization Act (regulation of OTC derivatives in the US)

The years 2000 – 2008

In the 2000s, banks and regulators have invested significant effort and resources to put in place the Basel II framework. This is during this period that modern risk management was significantly developed in the banking sector. The Solvency II reform emerged in 2004 and intensive work was underway to calibrate this new proposition on insurance regulation.

Basel II	1999-06-02	Publication of the first CP on Basel II
	2001-01-29	Publication of the second CP on Basel II
	2001-11-05	Results of the QIS 2
	2002-06-25	Results of the QIS 2.5
	2003-04-29	Publication of the third CP on Basel II
	2003-05-05	Results of the QIS 3
	2004-06-10	Publication of the Basel II Accord
	2004–2005	Conduct of QIS 4 (national impact study and tests)
	2005-07-30	Publication of "*The Application of Basel II to Trading Activities and the Treatment of Double Default Effects*"
	2006-06-16	Results of the QIS 5
	2006-06-30	Publication of the Basel II Comprehensive Version (including Basel I, Basel II and 2005 revisions)
CRD	2006-05-14	Publication of the directive 2006/48/EC
	2006-05-14	Publication of the directive 2006/49/EC (CRD)
Solvency I	2002-03-05	Non-life insurance directive 2002/13/EC (revision of solvency margin requirements)
	2002-11-05	Life insurance recast directive 2002/83/EC
Solvency II	2004	Initial works on Solvency II
	2006-03-17	Report on the first QIS
	2007	Report on the second QIS
	2007-11-01	Report on the third QIS
Market Regulation	2002-01-22	Publication of the directives 2001/107/EC and 2001/108/EC (UCITS III)
	2004-04-21	Publication of the directive 2004/39/EC (MiFID 1)

The years 2008 – 2019

The 2008 Global Financial Crisis completely changed the landscape of financial regulation. Under political pressures, we assist to a frenetic race of regulatory reforms. For instance, the Basel Committee had published 21 regulatory standards before 2007. From January 2008 to December 2014, this number has dramatically increased with 34 new regulatory standards. With Basel 2.5, new capital requirements are put in place for market risk. The Basel III framework is published at the end of 2010 and introduces new standards for managing the liquidity risk. However, the finalized version of Basel III reforms will be only published in 2017. In Europe, market regulation is the new hot topic for regulators. However, the major event of the beginning of this decade concerns systemic risk. New regulations have emerged and new financial activities are under scrutiny (shadow banking system, market infrastructures, investment management).

Basel 2.5	2007-10-12	Publication of the first CP on the incremental risk charge
	2008-07-22	Proposed revisions to the Basel II market risk framework
	2009-07-13	Publication of the final version of Basel 2.5
Basel III	2010-12-16	Publication of the original version of Basel III
	2011-06-01	Revised version of the Basel III capital rules reflecting the CVA modification
	2013-01-07	Publication of the rules concerning the liquidity coverage ratio
	2013-10-31	Fundamental review of the trading book (FRTB)
	2013-12-13	Capital requirements for banks' equity investments in funds
	2014-01-12	Publication of the leverage ratio
	2014-03-31	Publication of SA-CCR
	2014-04-10	Capital requirements for bank exposures to central counterparties
	2014-04-15	Supervisory framework for measuring and controlling large exposures
	2014-10-31	Publication of the rules concerning the net stable funding ratio
	2016-04-21	Interest rate risk in the banking book (IRRBB)
	2016-07-11	Revisions to the securitization framework
	2017-12-07	Final version of Basel III reforms
	2019-01-14	Publication of the Basel III comprehensive version for market risk
CRD/CRR	2009-09-16	Directive 2009/111/EC (CRD II)
	2010-09-24	Directive 2010/76/EU (CRD III)
	2013-06-26	Directive 2013/36/EU (CRD IV)
	2013-06-26	Publication of the capital requirements regulation 575/2013 (CRR)
	2013-10-15	Council regulation 1024/2013 concerning the European Central Bank and the prudential supervision
	2014-10-10	Commission delegated regulation 2015/62 of on the leverage ratio
	2017-12-12	Regulation 2017/2401 on securitizations
	2019	Publication of CRD V & CRR 2
Solvency II	2008-11-19	Report on the fourth QIS
	2009-11-25	Solvency II directive 2009/138/EC
	2011-03-14	Report on the fifth QIS
	2014-04-16	Publication of the Omnibus II directive 2014/51/UE
	2015-10-10	Publication of the commission delegated regulation 2015/35
	2015-12-02	Commission implementing regulation 2015/2450
Market Regulation	2009-07-13	Directive 2009/65/EC (UCITS IV)
	2010-06-08	AIFM directive (2011/61/EU)
	2012-07-04	EU regulation 648/2012 (EMIR)
	2014-05-15	Directive 2014/65/EU (MiFID II)
	2012-05-15	EU regulation 600/2014 (MiFIR)
	2014-07-23	Directive 2014/91/EU (UCITS V)
	2014-11-26	EU regulation 1286/2014 (PRIIPS)
	2015-11-25	EU regulation 2015/2365 on securities financing transactions
	2016-06-08	EU regulation 2016/1011 on indices and benchmarks
	2017-06-14	EU regulation 2017/1131 on money market funds

Continued on next page

Continued from previous page

	2009-04	Creation of the Financial Stability Board (FSB)
	2010-07-21	Dodd-Frank Wall Street Reform and Consumer Protection Act
Systemic	2010-07-21	Volcker Rule (§619 of the Dodd-Frank Act)
Risk	2011-11-04	Publication of the G-SIB assessment methodology (BCBS)
	2013-07-03	Update of the G-SIB assessment methodology (BCBS)
	2015-03-04	Second CP on assessment methodologies for identifying NBNI-SIFIs (FSB-IOSCO)

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part I

Risk Management in the Financial Sector

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 2

Market Risk

This chapter begins with the presentation of the regulatory framework. It will help us to understand how the supervision on market risk is organized and how the capital charge is computed. Then we will study the different statistical approaches to measure the value-at-risk and the expected shortfall. Specifically, a section is dedicated to the risk management of derivatives and exotic products. We will see the main concepts, but we will present the more technical details later in Chapter 9 dedicated to model risk. Advanced topics like Monte Carlo methods and stress testing models will also be addressed in Part II. Finally, the last part of the chapter is dedicated to risk allocation.

2.1 Regulatory framework

We recall that the original Basel Accord only concerned credit risk in 1988. However, the occurrences of market shocks were more important and the rapid development of derivatives created some stress events at the end of the eighties and the beginning of the nineties. On 19 October 1987, stock markets crashed and the Dow Jones Industrial Average index dropped by more than 20% in the day. In 1990, the collapse of the Japanese asset price bubble (both in stock and real estate markets) caused a lot of damage in the Japanese banking system and economy. The unexpected rise of US interest rates in 1994 resulted in a bond market massacre and difficulties for banks, hedge funds and money managers. In 1994-1995, several financial disasters occurred, in particular the bankruptcy of Barings and the Orange County affair (Jorion, 2007).

In April 1993, the Basel Committee published a first consultative paper to incorporate market risk in the Cooke ratio. Two years later, in April 1995, it accepted the idea to compute the capital charge for market risks with an internal model. This decision is mainly due to the publication of *RiskMetrics* by J.P. Morgan in October 1994. Finally, the Basel Committee published the amendment to the capital accord to incorporate market risks in January 1996. This proposal has remained the supervisory framework for market risk during many years. However, the 2008 Global Financial Crisis had a big impact in terms of market risk. Just after the crisis, a new approach called Basel 2.5 has been accepted. In 2012, the Basel Committee launched a major project: the fundamental review of the trading book (FRTB). These works resulted in the publication of a new comprehensive framework in January 2019 (BCBS, 2019). This is the Basel III framework for computing the minimum capital requirements for market risk as of January 2022.

According to BCBS (2019), market risk is defined as "*the risk of losses (in on- and off-balance sheet positions) arising from movements in market prices. The risks subject to market risk capital requirements include but are not limited to:*

- *default risk, interest rate risk, credit spread risk, equity risk, foreign exchange (FX) risk and commodities risk for trading book instruments*;
- *FX risk and commodities risk for banking book instruments*."

The following table summarizes the perimeter of markets risks that require regulatory capital:

Portfolio	Fixed Income	Equity	Currency	Commodity	Credit
Trading	✓	✓	✓	✓	✓
Banking			✓	✓	

The Basel Committee makes the distinction between the trading book and the banking book. Instruments to be included in the trading book are subject to market risk capital requirements, while instruments to be included in the banking book are subject to credit risk capital requirements (with the exception of foreign exchange and commodity instruments). The trading book refers to positions in assets held with trading intent or for hedging other elements of the trading book. These assets are systematically valuated on a fair value (mark-to-market or mark-to-model) basis, are actively managed and their holding is intentionally for short-term resale. Examples are proprietary trading, market-making activities, hedging portfolios of derivatives products, listed equities, repo transactions, etc. The banking book refers to positions in assets that are expected to be held until the maturity. These assets may be valuated at their historic cost or with a fair value approach. Examples are unlisted equities, real estate holdings, hedge funds, etc.

The first task of the bank is therefore to define trading book assets and banking book assets. For instance, if the bank sells an option on the Libor rate to a client, a capital charge for the market risk is required. If the bank provides a personal loan to a client with a fixed interest rate, there is a market risk if the interest rate risk is not hedged. However, a capital charge is not required in this case, because the exposure concerns the banking book. Exposures on stocks may be included in the banking book if the objective is a long-term investment.

2.1.1 The Basel I/II framework

To compute the capital charge, banks have the choice between two approaches:

1. the standardized measurement method (SMM);
2. the internal model-based approach (IMA).

The standardized measurement method has been implemented by banks at the end of the nineties. However, banks quickly realized that they can sharply reduce their capital requirements by adopting internal models. This explained that SMM was only used by a few number of small banks in the 2000s.

2.1.1.1 Standardized measurement method

Five main risk categories are identified: interest rate risk, equity risk, currency risk, commodity risk and price risk on options and derivatives. For each category, a capital charge is computed to cover the general market risk, but also the specific risk. According to the Basel Committee, specific risk includes the risk "*that an individual debt or equity security moves by more or less than the general market in day-to-day trading and event risk (e.g. takeover risk or default risk)*". The use of internal models is subject to the approval of the supervisor and the bank can mix the two approaches under some conditions. For instance, the bank may use SMM for the specific risk and IMA for the general market risk.

In this approach, the capital charge $\mathcal{K}$ is equal to the risk exposure E times the capital charge weight K:

$$\mathcal{K} = E \cdot K$$

For the specific risk, the risk exposure corresponds to the notional of the instrument, whether it is a long or a short position. For the general market risk, long and short positions on different instruments can be offset. In what follows, we give the main guidelines and we invite the reader to consult BCBS (1996a, 2006) to obtain the computational details.

Interest rate risk Let us first consider the specific risk. The Basel Committee makes the distinction between sovereign and other fixed income instruments. In the case of government instruments, the capital charge weights are:

Rating	AAA to AA−	A+ to BBB−			BB+ to B−	Below B−	NR
Maturity		0−6M	6M−2Y	2Y+			
K	0%	0.25%	1.00%	1.60%	8%	12%	8%

This capital charge depends on the rating and also the residual maturity for A+ to BBB− issuers[1]. The category NR stands for non-rated issuers. In the case of other instruments issued by public sector entities, banks and corporate companies, the capital charge weights are:

Rating	AAA to BBB−			BB+ to BB−	Below BB−	NR
Maturity	0−6M	6M−2Y	2Y+			
K	0.25%	1.00%	1.60%	8%	12%	8%

Example 4 *We consider a trading portfolio with the following exposures: a long position of* \$50 *mn on Euro-Bund futures, a short position of* \$100 *mn on three-month T-Bills and a long position of* \$10 *mn on an investment grade (IG) corporate bond with a three-year residual maturity.*

The underlying asset of Euro-Bund futures is a German bond with a long maturity (higher than 6 years). We deduce that the capital charge for specific risk for the two sovereign exposures is equal to zero, because both Germany and US are rated above A+. Concerning the corporate bond, we obtain:

$$\mathcal{K} = 10 \times 1.60\% = \$160\,000$$

For the general market risk, the bank has the choice between two methods: the maturity approach and the duration approach. In the maturity approach, long and short positions are slotted into a maturity-based ladder comprising fifteen time-bands (less than one month, between one and three months, ... between 12 and 20 years, greater than 20 years). The risk weights depend on the time band and the value of the coupon[2], and apply to the net exposure on each time band. For example, a capital charge of 8% is used for the net

[1] Three maturity periods are defined: 6 months or less, greater than 6 months and up to 24 months, more than 24 months.

[2] We distinguish coupons less than 3% (small coupons or SC) and coupons 3% or more (big coupons or BC).

exposure of instruments (with small coupons), whose maturity is between 12 and 20 years. For reflecting basis and gap risks, the bank must also include a 10% capital charge to the smallest exposure of the matched positions. This adjustment is called the '*vertical disallowance*'. The Basel Committee considers a second adjustment for horizontal offsetting (the '*horizontal disallowance*'). For that, it defines 3 zones (less than 1 year, one year to four years and more than four years). The offsetting can be done within and between the zones. The adjustment coefficients are 30% within the zones 2 and 3, 40% within the zone 1, between the zones 1 and 2, and between the zones 2 and 3, and 100% between the zones 1 and 3. Therefore, the regulatory capital for the general market risk is the sum of the three components:

$$\mathcal{K} = \mathcal{K}^{\text{OP}} + \mathcal{K}^{\text{VD}} + \mathcal{K}^{\text{HD}}$$

where $\mathcal{K}^{\text{OP}}$, $\mathcal{K}^{\text{VD}}$ and $\mathcal{K}^{\text{HD}}$ are the required capital for the overall net open position, the vertical disallowance and the horizontal disallowance.

With the duration approach, the bank computes the price sensitivity of each position with respect to a change in yield Δy, slots the sensitivities into a duration-based ladder and applies adjustments for vertical and horizontal disallowances. The computation of the required capital is exactly the same as previously, but with a different definition of time bands and zones.

Equity risk For equity exposures, the capital charge for specific risk is 4% if the portfolio is liquid and well-diversified and 8% otherwise. For the general market risk, the risk weight is equal to 8% and applies to the net exposure.

Example 5 *We consider a* \$100 *mn short exposure on the S&P 500 index futures contract and a* \$60 *mn long exposure on the Apple stock.*

The capital charge for specific risk is[3]:

$$\begin{aligned}\mathcal{K}^{\text{Specific}} &= 100 \times 4\% + 60 \times 8\% \\ &= 4 + 4.8 \\ &= 8.8\end{aligned}$$

The net exposure is −\$40 mn. We deduce that the capital charge for the general market risk is:

$$\begin{aligned}\mathcal{K}^{\text{General}} &= |-40| \times 8\% \\ &= 3.2\end{aligned}$$

It follows that the total capital charge for this equity portfolio is \$12 mn.

Remark 1 *Under Basel 2.5, the capital charge for specific risk is set to 8% whatever the liquidity of the portfolio.*

Foreign exchange risk The Basel Committee includes gold in this category and not in the commodity category because of its specificity in terms of volatility and its status of safe-heaven currency. The bank has first to calculate the net position (long or short) of each currency. The capital charge is then 8% of the global net position defined as the sum of:

[3] We assume that the S&P 500 index is liquid and well-diversified, whereas the exposure on the Apple stock is not diversified.

- the maximum between the aggregated value $\mathcal{L}_{\text{FX}}$ of long positions and the aggregated value $\mathcal{S}_{\text{FX}}$ of short positions and,
- the absolute value of the net position $\mathcal{N}_{\text{Gold}}$ in gold.

We have:

$$\mathcal{K} = 8\% \times \left(\max\left(\mathcal{L}_{\text{FX}}, \mathcal{S}_{\text{FX}}\right) + \left|\mathcal{N}_{\text{Gold}}\right|\right)$$

Example 6 *We consider a bank which has the following long and short positions expressed in \$ mn*[4]*:*

Currency	EUR	JPY	GBP	CHF	CAD	AUD	ZAR	Gold
$\mathcal{L}_i$	170	0	25	37	11	3	8	33
$\mathcal{S}_i$	80	50	12	9	28	0	8	6

We first compute the net exposure $\mathcal{N}_i$ for each currency:

$$\mathcal{N}_i = \mathcal{L}_i - \mathcal{S}_i$$

We obtain the following figures:

Currency	EUR	JPY	GBP	CHF	CAD	AUD	ZAR	Gold
$\mathcal{N}_i$	90	−50	13	28	−17	3	0	27

We then calculate the aggregated long and short positions:

$$\begin{aligned}
\mathcal{L}_{\text{FX}} &= 90 + 13 + 28 + 3 + 0 = 134 \\
\mathcal{S}_{\text{FX}} &= 50 + 17 = 67 \\
\mathcal{N}_{\text{Gold}} &= 27
\end{aligned}$$

We finally deduce that the capital charge is equal to \$12.88 mn:

$$\begin{aligned}
\mathcal{K} &= 8\% \times \left(\max\left(134, 67\right) + \left|27\right|\right) \\
&= 8\% \times 161 \\
&= 12.88
\end{aligned}$$

Commodity risk Commodity risk concerns both physical and derivative positions (forward, futures[5] and options). This includes energy products (oil, gas, ethanol, etc.), agricultural products (grains, oilseeds, fiber, livestock, etc.) and metals (industrial and precious), but excludes gold which is covered under foreign exchange risk. The Basel Committee makes the distinction between the risk of spot or physical trading, which is mainly affected by the directional risk and the risk of derivative trading, which includes the directional risk, the basis risk, the cost-of-carry and the forward gap (or time spread) risk. The SMM for commodity risk includes two options: the simplified approach and the maturity ladder approach.

Under the simplified approach, the capital charge for directional risk is 15% of the absolute value of the net position in each commodity. For the other three risks, the capital charge is equal to 3% of the global gross position. We have:

$$\mathcal{K} = 15\% \times \sum_{i=1}^{m} \left|\mathcal{L}_i - \mathcal{S}_i\right| + 3\% \times \sum_{i=1}^{m} \left(\mathcal{L}_i + \mathcal{S}_i\right)$$

[4] We implicity assume that the reporting currency of the bank is the US dollar.

[5] The most traded futures contracts are crude oil, brent, heating oil, gas oil, natural oil, rbob gasoline silver, platinum, palladium, zinc, lead, aluminium, cocoa, soybeans, corn, cotton, wheat, sugar, live cattle, coffee and soybean oil.

where m is the number of commodities, $\mathcal{L}_i$ is the long position on commodity i and $\mathcal{S}_i$ is the short position on commodity i.

Example 7 *We consider a portfolio of five commodities. The mark-to-market exposures expressed in* \$ *mn are the following:*

Commodity	Crude Oil	Coffee	Natural Gas	Cotton	Sugar
$\mathcal{L}_i$	23	5	3	8	11
$\mathcal{S}_i$	0	0	19	2	6

The aggregated net exposure $\sum_{i=1}^{5} |\mathcal{L}_i - \mathcal{S}_i|$ is equal to \$55 mn whereas the gross exposure $\sum_{i=1}^{5} (\mathcal{L}_i + \mathcal{S}_i)$ is equal to \$77 mn. We deduce that the required capital is $15\% \times 55 + 3\% \times 77$ or \$10.56 mn.

Under the maturity ladder approach, the bank should spread long and short exposures of each commodity to seven time bands: 0-1M, 1M-3M, 3M-6M, 6M-1Y, 1Y-2Y, 2Y-3Y, 3Y+. For each time band, the capital charge for the basis risk is equal to 1.5% of the matched positions (long and short). Nevertheless, the residual net position of previous time bands may be carried forward to offset exposures in next time bands. In this case, a surcharge of 0.6% of the residual net position is added at each time band to cover the time spread risk. Finally, a capital charge of 15% is applied to the global net exposure (or the residual unmatched position) for directional risk.

Option's market risk There are three approaches for the treatment of options and derivatives. The first method, called the simplified approach, consists of calculating separately the capital charge of the position for the option and the associated underlying. In the case of an hedged exposure (long cash and long put, short cash and long call), the required capital is the standard capital charge of the cash exposure less the amount of the in-the-money option. In the case of a non-hedged exposure, the required capital is the minimum value between the mark-to-market of the option and the standard capital charge for the underlying.

Example 8 *We consider a variant of Example 5. We have a* \$100 *mn short exposure on the S&P 500 index futures contract and a* \$60 *mn long exposure on the Apple stock. We assume that the current stock price of Apple is* \$120*. Six months ago, we have bought* 400 000 *put options on Apple with a strike of* \$130 *and a one-year maturity. We also decide to buy* 10 000 *ATM call options on Google. The current stock price of Google is* \$540 *and the market value of the option is* \$45.5*.*

We deduce that we have 500 000 shares of the Apple stock. This implies that \$48 mn of the long exposure on Apple is hedged by the put options. Concerning the derivative exposure on Google, the market value is equal to \$0.455 mn. We can therefore decompose this portfolio into three main exposures:

- a directional exposure composed by the \$100 mn short exposure on the S&P 500 index and the \$12 mn remaining long exposure on the Apple stock;
- a \$48 mn hedged exposure on the Apple stock;
- a \$0.455 mn derivative exposure on the Google stock.

For the directional exposure, we compute the capital charge for specific and general market risks[6]:

$$\begin{aligned}\mathcal{K} &= (100 \times 4\% + 12 \times 8\%) + 88 \times 8\% \\ &= 4.96 + 7.04 \\ &= 12\end{aligned}$$

For the hedged exposure, we proceed as previously but we deduce the in-the-money value[7]:

$$\begin{aligned}\mathcal{K} &= 48 \times (8\% + 8\%) - 4 \\ &= 3.68\end{aligned}$$

The market value of the Google options is \$0.455 mn. We compare this value to the standard capital charge[8] to determine the capital charge:

$$\begin{aligned}\mathcal{K} &= \min(5.4 \times 16\%, 0.455) \\ &= 0.455\end{aligned}$$

We finally deduce that the required capital is \$16.135 mn.

The second approach is the delta-plus method. In this case, the directional exposure of the option is calculated by its delta. Banks will also required to compute an additional capital charge for gamma and vega risks. We consider different options and we note $j \in \mathcal{A}_i$ when the option j is written on the underlying asset i. We first compute the (signed) capital charge for the 4 risks at the asset level:

$$\begin{aligned}\mathcal{K}_i^{\text{Specific}} &= \left(\sum_{j \in \mathcal{A}_i} N_j \cdot \boldsymbol{\Delta}_j\right) \cdot S_i \cdot K_i^{\text{Specific}} \\ \mathcal{K}_i^{\text{General}} &= \left(\sum_{j \in \mathcal{A}_i} N_j \cdot \boldsymbol{\Delta}_j\right) \cdot S_i \cdot K_i^{\text{General}} \\ \mathcal{K}_i^{\text{Gamma}} &= \frac{1}{2}\left(\sum_{j \in \mathcal{A}_i} N_j \cdot \boldsymbol{\Gamma}_j\right) \cdot \left(S_i \cdot K_i^{\text{Gamma}}\right)^2 \\ \mathcal{K}_i^{\text{Vega}} &= \sum_{j \in \mathcal{A}_i} N_j \cdot \boldsymbol{\upsilon}_j \cdot (25\% \cdot \Sigma_j)\end{aligned}$$

where S_i is the current market value of the asset i, K_i^{Specific} and K_i^{General} are the corresponding standard capital charge for specific and general market risk and K_i^{Gamma} is the capital charge for gamma impact[9]. Here, N_j, $\boldsymbol{\Delta}_j$, $\boldsymbol{\Gamma}_j$ and $\boldsymbol{\upsilon}_j$ are the exposure, delta, gamma and vega of the option j. For the vega risk, the shift corresponds to $\pm 25\%$ of the implied volatility Σ_j. For a portfolio of assets, the traditional netting rules apply to specific and general market risks. The total capital charge for gamma risk corresponds to the opposite of the sum of the negative individual capital charges for gamma risk whereas the total capital charge for vega risk corresponds to the sum of the absolute value of individual capital charges for vega risk.

[6] The net short exposure is equal to \$88 mn.

[7] It is equal to $400\,000 \times \max(130 - 120, 0)$.

[8] It is equal to $10\,000 \times 540 \times (8\% + 8\%)$.

[9] It is equal to 8% for equities, 8% for currencies and 15% for commodities. In the case of interest rate risk, it corresponds to the standard value $K(t)$ for the time band t (see the table on page 8 in BCBS (1996a)).

Example 9 *We consider a portfolio of 4 options written on stocks with the following characteristics:*

Option	Stock	Exposure	Type	Price	Strike	Maturity	Volatility
1	A	−5	call	100	110	1.00	20%
2	A	−10	call	100	100	2.00	20%
3	B	10	call	200	210	1.00	30%
4	B	8	put	200	190	1.25	35%

This means that we have 2 assets. For stock A, we have a short exposure on 5 call options with a one-year maturity and a short exposure on 10 call options with a two-year maturity. For stock B, we have a long exposure on 10 call options with a one-year maturity and a long exposure on 8 put options with a maturity of one year and three months.

Using the Black-Scholes model, we first compute the Greek coefficients for each option j. Because the options are written on single stocks, the capital charges K_i^{Specific}, K_i^{General} and K_i^{Gamma} are all equal to 8%. Using the previous formulas, we then deduce the individual capital charges for each option[10]:

j	1	2	3	4
$\boldsymbol{\Delta}_j$	0.45	0.69	0.56	−0.31
$\boldsymbol{\Gamma}_j$	0.02	0.01	0.01	0.00
$\boldsymbol{\upsilon}_j$	39.58	49.91	78.85	79.25
$\boldsymbol{\mathcal{K}}_j^{\text{Specific}}$	−17.99	−55.18	89.79	−40.11
$\boldsymbol{\mathcal{K}}_j^{\text{General}}$	−17.99	−55.18	89.79	−40.11
$\boldsymbol{\mathcal{K}}_j^{\text{Gamma}}$	−3.17	−3.99	8.41	4.64
$\boldsymbol{\mathcal{K}}_j^{\text{Vega}}$	−9.89	−24.96	59.14	55.48

We can now aggregate the previous individual capital charges for each stock. We obtain:

Stock	$\boldsymbol{\mathcal{K}}_i^{\text{Specific}}$	$\boldsymbol{\mathcal{K}}_i^{\text{General}}$	$\boldsymbol{\mathcal{K}}_i^{\text{Gamma}}$	$\boldsymbol{\mathcal{K}}_i^{\text{Vega}}$
A	−73.16	−73.16	−7.16	−34.85
B	49.69	49.69	13.05	114.61
Total	122.85	23.47	7.16	149.46

To compute the total capital charge, we apply the netting rule for the general market risk, but not for the specific risk. This means that $\boldsymbol{\mathcal{K}}^{\text{Specific}} = |-73.16| + |49.69| = 122.85$ and $\boldsymbol{\mathcal{K}}^{\text{General}} = |-73.16 + 49.69| = 23.47$. For gamma risk, we only consider negative impacts and we have $\boldsymbol{\mathcal{K}}^{\text{General}} = |-7.16| = 7.16$. For vega risk, there is no netting rule: $\boldsymbol{\mathcal{K}}^{\text{Vega}} = |-34.85| + |114.61| = 149.46$. We finally deduce that the overall capital is 302.94.

The third method is the scenario approach. In this case, we evaluate the profit and loss (P&L) for simultaneous changes in the underlying price and in the implied volatility of the option. For defining these scenarios, the ranges are the standard shifts used previously. For instance, we use the following ranges for equities:

		S_i	
		−8%	+8%
Σ_j	−25%		
	+25%		

[10] For instance, the individual capital charge of the second option for the gamma risk is

$$\boldsymbol{\mathcal{K}}_j^{\text{Gamma}} = \frac{1}{2} \times (-10) \times 0.0125 \times (100 \times 8\%)^2 = -3.99$$

The scenario matrix corresponds to intermediate points on the 2×2 grid. For each cell of the scenario matrix, we calculate the P&L of the option exposure[11]. The capital charge is then the largest loss.

Securitization instruments The treatment of specific risk of securitization positions is revised in Basel 2.5 and is based on external ratings. For instance, the capital charge for securitization exposures is 1.6% if the instrument is rated from AAA to AA−. For resecuritization exposures, it is equal to 3.2%. If the rating of the instrument is from BB+ to BB−, the risk capital charges becomes respectively[12] 28% and 52%.

2.1.1.2 Internal model-based approach

The use of an internal model is conditional upon the approval of the supervisory authority. In particular, the bank must meet certain criteria concerning different topics. These criteria concerns the risk management system, the specification of market risk factors, the properties of the internal model, the stress testing framework, the treatment of the specific risk and the backtesting procedure. In particular, the Basel Committee considers that the bank must have "*sufficient numbers of staff skilled in the use of sophisticated models not only in the trading area but also in the risk control, audit, and if necessary, back office areas*". We notice that the Basel Committee first insists on the quality of the trading department, meaning that the trader is the first level of risk management. The validation of an internal model does not therefore only concern the risk management department, but the bank as a whole.

Qualitative criteria BCBS (1996a) defines the following qualitative criteria:

- "*The bank should have an independent risk control unit that is responsible for the design and implementation of the bank's risk management system. [...] This unit must be independent from business trading units and should report directly to senior management of the bank*".
- The risk management department produces and analyzes daily reports, is responsible for the backtesting procedure and conducts stress testing analysis.
- The internal model must be used to manage the risk of the bank in the daily basis. It must be completed by trading limits expressed in risk exposure.
- The bank must document internal policies, controls and procedures concerning the risk measurement system (including the internal model).

It is today obvious that the risk management department should not report to the trading and sales department. Twenty-five years ago, it was not the case. Most of risk management units were incorporated to business units. It has completely changed because of the regulation and risk management is now independent from the front office. The risk management function has really emerged with the amendment to incorporate market risks and even more with the Basel II reform, whereas the finance function has long been developed in banks. For instance, it's very recent that the head of risk management[13] is also a member of the executive committee of the bank whereas the head of the finance department[14] has always been part of the top management.

[11] It may include the cash exposure if the option is used for hedging purposes.
[12] See pages 4-7 of BCBS (2009b) for the other risk capital charges.
[13] He is called the chief risk officer or CRO.
[14] He is called the chief financial officer or CFO.

From the supervisory point of view, an internal model does not reduce to measure the risk. It must be integrated in the management of the risk. This is why the Basel Committee points out the importance between the outputs of the model (or the risk measure), the organization of the risk management and the impact on the business.

Quantitative criteria The choice of the internal model is left to the bank, but it must respect the following quantitative criteria:

- The value-at-risk (VaR) is computed on a daily basis with a 99% confidence level. The minimum holding period of the VaR is 10 trading days. If the bank computes a VaR with a shorter holding period, it can use the square-root-of-time rule.
- The risk measure can take into account diversification, that is the correlations between the risk categories.
- The model must capture the relevant risk factors and the bank must pay attention to the specification of the appropriate set of market risk factors.
- The sample period for calculating the value-at-risk is at least one year and the bank must update the data set frequently (every month at least).
- In the case of options, the model must capture the non-linear effects with respect to the risk factors and the vega risk.
- "*Each bank must meet, on a daily basis, a capital requirement expressed as the higher of (i) its previous day's value-at-risk number [...] and (ii) an average of the daily value-at-risk measures on each of the preceding sixty business days, multiplied by a multiplication factor*".
- The value of the multiplication factor depends on the quality of the internal model with a range between 3 and 4. The quality of the internal model is related to its ex-post performance measured by the backtesting procedure.

The holding period to define the capital is 10 trading days. However, it is difficult to compute the value-at-risk for such holding period. In practice, the bank computes the one-day value-at-risk and converts this number into a ten-day value-at-risk using the square-root-of-time rule:

$$\mathrm{VaR}_\alpha\left(w; \text{ten days}\right) = \sqrt{10} \times \mathrm{VaR}_\alpha\left(w; \text{one day}\right)$$

This rule comes from the scaling property of the volatility associated to a geometric Brownian motion. It has the advantage to be simple and objective, but it generally underestimates the risk when the loss distribution exhibits fat tails[15].

The required capital at time t is equal to:

$$\boldsymbol{\mathcal{K}}_t = \max\left(\mathrm{VaR}_{t-1}, (3+\xi)\cdot\frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}_{t-i}\right) \tag{2.1}$$

where VaR_t is the value-at-risk calculated at time t and ξ is the penalty coefficient ($0 \leq \xi \leq 1$). In normal periods where $\mathrm{VaR}_{t-1} \simeq \mathrm{VaR}_{t-i}$, the required capital is the average of the last 60 value-at-risk values times the multiplication factor[16] $m_c = 3 + \xi$. In this case, we have:

$$\boldsymbol{\mathcal{K}}_t = \boldsymbol{\mathcal{K}}_{t-1} + \frac{m_c}{60}\cdot\left(\mathrm{VaR}_{t-1} - \mathrm{VaR}_{t-61}\right)$$

[15] See for instance Diebold *et al.* (1998), Daníelsson and Zigrand (2006) or Wang *et al.* (2011).

[16] The complementary factor is explained on page 88.

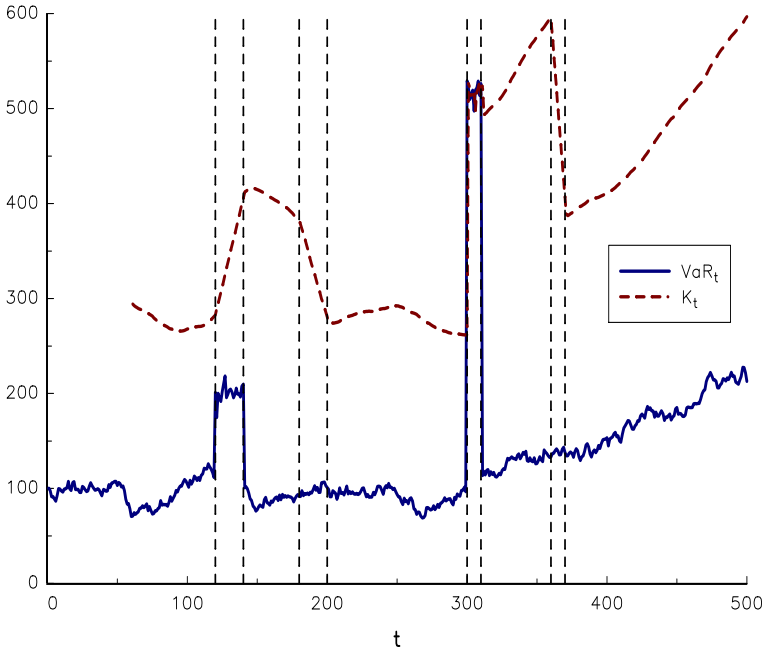


FIGURE 2.1: Calculation of the required capital with the VaR

The impact of VaR_{t-1} is limited because the factor $(3+\xi)/60$ is smaller than 6.7%. The required capital can only be equal to the previous day's value-at-risk if the bank faces a stress $\text{VaR}_{t-1} \gg \text{VaR}_{t-i}$. We also notice that a shock on the VaR vanishes after 60 trading days. To understand the calculation of the capital, we report an illustration in Figure 2.1. The solid line corresponds to the value-at-risk VaR_t whereas the dashed line corresponds to the capital $\boldsymbol{\mathcal{K}}_t$. We assume that $\xi = 0$ meaning that the multiplication factor is equal to 3. When $t < 120$, the value-at-risk varies around a constant. The capital is then relatively smooth and is three times the average VaR. At time $t = 120$, we observe a shock on the value-at-risk, which lasts 20 days. Immediately, the capital increases until $t \leq 140$. Indeed, at this time, the capital takes into account the full period of the shocked VaR (between $t = 120$ and $t = 139$). The full effect of this stressed period continues until $t \leq 180$, but this effect becomes partial when $t > 180$. The impact of the shock vanishes when $t = 200$. We then observe a period of 100 days where the capital is smooth because the daily value-at-risk does not change a lot. A second shock on the value-at-risk occurs at time $t = 300$, but the magnitude of the shock is larger than previously. During 10 days, the required capital is exactly equal to the previous day's value-at-risk. After 10 days, the bank succeeds to reduce the risk of its portfolio. However, the daily value-at-risk increases from $t = 310$ to $t = 500$. As previously, the impact of the second shock vanishes 60 days after the end of shock. However, the capital increases strongly at the end of the period. This is due to the effect of the multiplication factor m_c on the value-at-risk.

Stress testing Stress testing is a simulation method to identify events that could have a great impact on the soundness of the bank. The framework consists of applying stress scenarios and low-probability events on the trading portfolio of the bank and to evaluate the maximum loss. Contrary to the value-at-risk[17], stress testing is not used to compute the

[17]The 99% VaR is considered as a risk measure in normal markets and therefore ignores stress events.

required capital. The underlying idea is more to identify the adverse scenarios for the bank, evaluate the corresponding losses, reduce eventually the too risky exposures and anticipate the management of such stress periods.

Stress tests should incorporate both market and liquidity risks. The Basel Committee considers two types of stress tests:

1. supervisory stress scenarios;
2. stress scenarios developed by the bank itself.

The supervisory stress scenarios are standardized and apply to the different banks. This allows the supervisors to compare the vulnerability between the different banks. The bank must complement them by its own scenarios in order to evaluate the vulnerability of its portfolio according to the characteristics of the portfolio. In particular, the bank may be exposed to some political risks, regional risks or market risks that are not taken into account by standardized scenarios. The banks must report their test results to the supervisors in a quarterly basis.

Stress scenarios may be historical or hypothetical. In the case of historical scenarios, the bank computes the worst-case loss associated to different crisis: the Black Monday (1987), the European monetary system crisis (1992), the bond market sell-off (1994), the internet bubble (2000), the subprime mortgage crisis (2007), the liquidity crisis due to Lehman Brothers collapse (2008), the Euro zone crisis (2011-2012), etc. Hypothetical scenarios are more difficult to calibrate, because they must correspond to extreme but also plausible events. Moreover, the multidimensional aspect of stress scenarios is an issue. Indeed, the stress scenario is defined by the extreme event, but the corresponding loss is evaluated with respect to the shocks on market risk factors. For instance, if we consider a severe Middle East crisis, this event will have a direct impact on the oil price, but also indirect impacts on other market risk factors (equity prices, US dollar, interest rates). Whereas historical scenarios are objective, hypothetical scenarios are by construction subjective and their calibration will differ from one financial institution to another. In the case of the Middle East crisis, one bank may consider that the oil price could fall by 30% whereas another bank may use a price reduction of 50%.

In 2009, the Basel Committee revised the market risk framework. In particular, it introduces the stressed value-at-risk measure. The stressed VaR has the same characteristics than the traditional VaR (99% confidence level and 10-day holing period), but the model inputs are "*calibrated to historical data from a continuous 12-month period of significant financial stress relevant to the bank's portfolio*". For instance, a typical period is the 2008 year which both combines the subprime mortgage crisis and the Lehman Brothers bankruptcy. This implies that the historical period to compute the SVaR is completely different than the historical period to compute the VaR (see Figure 2.2). In Basel 2.5, the capital requirement for stressed VaR is:

$$\boldsymbol{\mathcal{K}}_t^{\text{SVaR}} = \max\left(\text{SVaR}_{t-1}, m_s \cdot \frac{1}{60}\sum_{i=1}^{60}\text{SVaR}_{t-i}\right)$$

where SVaR_t is the stressed VaR measure computed at time t. Like the coefficient m_c, the multiplication factor m_s for the stressed VaR is also calibrated with respect to the backtesting outcomes, meaning that we have $m_s = m_c$ in many cases.

Specific risk and other risk charges In the case where the internal model does not take into account the specific risk, the bank must compute a specific risk charge (SRC) using

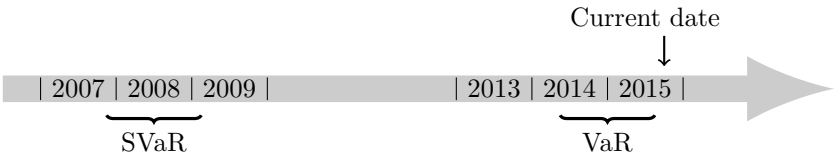


FIGURE 2.2: Two different periods to compute the VaR and the SVaR

the standardized measurement method. To be validated as a value-at-risk measure with specific risks, the model must satisfy at least the following criteria: it captures concentrations (magnitude and changes in composition), it captures name-related basis and event risks and it considers the assessment of the liquidity risk. For instance, an internal model built with a general market risk factor[18] does not capture specific risk. Indeed, the risk exposure of the portfolio is entirely determined by the beta of the portfolio with respect to the market risk factor. This implies that two portfolios with the same beta but with a different composition, concentration or liquidity have the same value-at-risk.

Basel 2.5 established a new capital requirement "*in response to the increasing amount of exposure in banks' trading books to credit-risk related and often illiquid products whose risk is not reflected in value-at-risk*" (BCBS, 2009b). The incremental risk charge (IRC) measures the impact of rating migrations and defaults, corresponds to a 99.9% value-at-risk for a one-year time horizon and concerns portfolios of credit vanilla trading (bonds and CDS). The IRC may be incorporated into the internal model or it may be treated as a surcharge from a separate calculation. Also under Basel 2.5, the Basel Committee introduced the comprehensive risk measure (CRM), which corresponds to a supplementary capital charge for credit exotic trading portfolios[19]. The CRM is also a 99.9% value-at-risk for a one-year time horizon. For IRC and CRM, the capital charge is the maximum between the most recent risk measure and the average of the risk measure over 12 weeks[20]. We finally obtain the following formula to compute the capital charge for the market risk under Basel 2.5:

$$\mathcal{K}_t = \mathcal{K}_t^{\text{VaR}} + \mathcal{K}_t^{\text{SVaR}} + \mathcal{K}_t^{\text{SRC}} + \mathcal{K}_t^{\text{IRC}} + \mathcal{K}_t^{\text{CRM}}$$

where $\mathcal{K}_t^{\text{VaR}}$ is given by Equation (2.1) and $\mathcal{K}_t^{\text{SRC}}$ is the specific risk charge. In this formula, $\mathcal{K}_t^{\text{SRC}}$ and/or $\mathcal{K}_t^{\text{IRC}}$ may be equal to zero if the modeling of these two risks is included in the value-at-risk internal model.

Backtesting and the ex-post evaluation of the internal model The backtesting procedure is described in the document *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements* published by the Basel Committee in January 1996. It consists of verifying that the internal model is consistent with a 99% confidence level. The idea is then to compare the outcomes of the risk model with realized loss values. For instance, we expect that the realized loss exceeds the VaR figure once every 100 observations on average.

The backtesting is based on the one-day holding period and compares the previous day's value-at-risk with the daily realized profit and loss. An exception occurs if the loss exceeds the value-at-risk. For a given period, we compute the number of exceptions. Depending of the frequency of exceptions, the supervisor determines the value of the penalty function between

[18] This is the case of the capital asset pricing model (CAPM) developed by Sharpe (1964).
[19] This concerns correlation trading activities on credit derivatives.
[20] Contrary to the VaR and SVaR measures, the risk measure is not scaled by a multiplication factor for IRC and CRM.

0 and 1. In the case of a sample based on 250 trading days, the Basel Committee defines three zones and proposes the values given in Table 2.1. The green zone corresponds to a number of exceptions less or equal to 4. In this case, the Basel Committee considers that there is no problem and the penalty coefficient ξ is set to 0. If the number of exceptions belongs to the yellow zone (between 5 and 9 exceptions), it may indicate that the confidence level of the internal model could be lower than 99% and implies that ξ is greater than zero. For instance, if the number of exceptions for the last 250 trading days is 6, the Basel Committee proposes that the penalty coefficient ξ is set to 0.50, meaning that the multiplication coefficient m_c is equal to 3.50. The red zone is a concern. In this case, the supervisor must investigate the reasons of such large number of exceptions. If the problem comes from the relevancy of the model, the supervisor can invalidate the internal model-based approach.

TABLE 2.1: Value of the penalty coefficient ξ for a sample of 250 observations

Zone	Number of exceptions	ξ
Green	0 – 4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	10+	1.00

The definition of the color zones comes from the statistical analysis of the exception frequency. We note w the portfolio, $L_t(w)$ the daily loss at time t and $\mathrm{VaR}_\alpha(w;h)$ the value-at-risk calculated at time $t-1$. By definition, $L_t(w)$ is the opposite of the P&L $\Pi_t(w)$:

$$\begin{aligned} L_t(w) &= -\Pi_t(w) \\ &= \mathrm{MtM}_{t-1} - \mathrm{MtM}_t \end{aligned}$$

where MtM_t is the mark-to-market of the trading portfolio at time t. By definition, we have:

$$\Pr\{L_t(w) \geq \mathrm{VaR}_\alpha(w;h)\} = 1-\alpha$$

where α is the confidence level of the value-at-risk. Let e_t be the random variable which is equal to 1 if there is an exception and 0 otherwise. e_t is a Bernoulli random variable with parameter p:

$$\begin{aligned} p &= \Pr\{e_t = 1\} \\ &= \Pr\{L_t(w) \geq \mathrm{VaR}_\alpha(w;h)\} \\ &= 1-\alpha \end{aligned}$$

In the case of the Basel framework, α is set to 99% meaning that we have a probability of 1% to observe an exception every trading day. For a given period $[t_1, t_2]$ of n trading days, the probability to observe exactly m exceptions is given by the binomial formula:

$$\Pr\{N_e(t_1;t_2) = m\} = \binom{n}{m}(1-\alpha)^m \alpha^{n-m}$$

where $N_e(t_1;t_2) = \sum_{t=t_1}^{t_2} e_t$ is the number of exceptions for the period $[t_1, t_2]$. We obtain this result under the assumption that the exceptions are independent across time. $N_e(t_1;t_2)$

is then the binomial random variable $\mathcal{B}(n; 1-\alpha)$. We deduce that the probability to have up to m exceptions is:

$$\Pr\{N_e(t_1;t_2) \leq m\} = \sum_{j=0}^{m} \binom{n}{j} (1-\alpha)^i \alpha^{n-j}$$

The three previous zones are then defined with respect to the statistical confidence level of the assumption $\mathcal{H} : \alpha = 99\%$. The green zone corresponds to the 95% confidence level: $\Pr\{N_e(t_1;t_2) \leq m\} < 95\%$. In this case, the hypothesis $\mathcal{H} : \alpha = 99\%$ is not rejected at the 95% confidence level. The yellow and red zones are respectively defined by $95\% \leq \Pr\{N_e(t_1;t_2) \leq m\} < 99.99\%$ and $\Pr\{N_e(t_1;t_2) \leq m\} \geq 99.99\%$. This implies that the hypothesis $\mathcal{H} : \alpha = 99\%$ is rejected at the 99.99% confidence level if the number of exceptions belongs to the red zone.

TABLE 2.2: Probability distribution (in %) of the number of exceptions ($n = 250$ trading days)

	$\alpha = 99\%$		$\alpha = 98\%$	
m	$\Pr\{N_e = m\}$	$\Pr\{N_e \leq m\}$	$\Pr\{N_e = m\}$	$\Pr\{N_e \leq m\}$
0	8.106	8.106	0.640	0.640
1	20.469	28.575	3.268	3.908
2	25.742	54.317	8.303	12.211
3	21.495	75.812	14.008	26.219
4	13.407	89.219	17.653	43.872
5	6.663	95.882	17.725	61.597
6	2.748	98.630	14.771	76.367
7	0.968	99.597	10.507	86.875
8	0.297	99.894	6.514	93.388
9	0.081	99.975	3.574	96.963
10	0.020	99.995	1.758	98.720

If we apply the previous statistical analysis when n is equal to 250 trading days, we obtain the results given in Table 2.2. For instance, the probability to have zero exception is 8.106%, the probability to have one exception is 20.469%, etc. We retrieve the three color zones determined by the Basel Committee. The green zone corresponds to the interval $[0, 4]$, the yellow zone is defined by the interval $[5, 9]$ and the red zone involves the interval $[10, 250]$. We notice that the color zones can vary significantly if the confidence level of the value-at-risk is not equal to 99%. For instance, if it is equal to 98%, the green zone corresponds to less than 9 exceptions. In Figure 2.3, we have reported the color zones with respect to the size n of the sample.

Example 10 *Calculate the color zones when n is equal to 1 000 trading days and $\alpha = 99\%$.*

We have $\Pr\{N_e \leq 14\} = 91.759\%$ and $\Pr\{N_e \leq 15\} = 95.213\%$. This implies that the green zones ends at 14 exceptions whereas the yellow zone begins at 15 exceptions. Because $\Pr\{N_e \leq 23\} = 99.989\%$ and $\Pr\{N_e \leq 24\} = 99.996\%$, we also deduce that the red zone begins at 24 exceptions.

Remark 2 *The statistical approach of backtesting ignores the effects of intra-day trading. Indeed, we make the assumption that the portfolio remains unchanged from $t-1$ to t, which is not the case in practice. This is why the Basel Committee proposes to compute the loss*

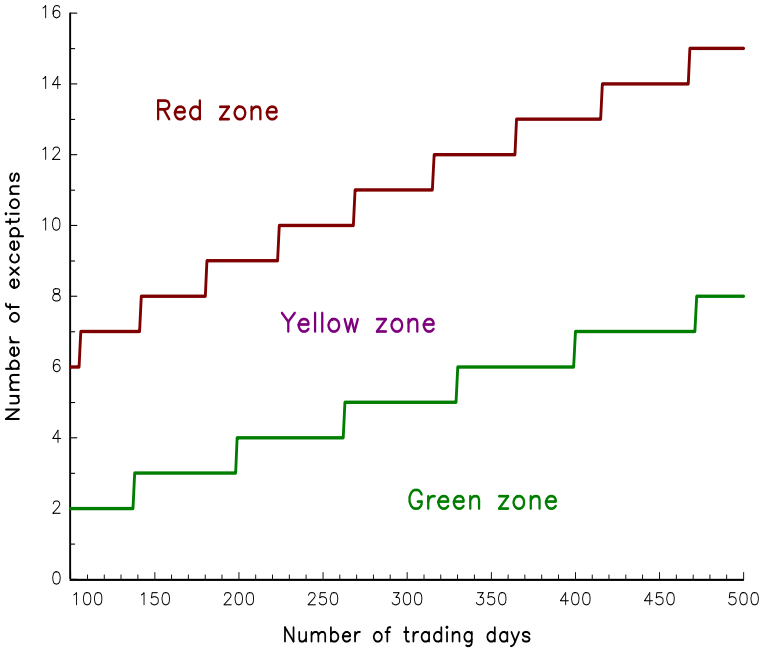


FIGURE 2.3: Color zones of the backtesting procedure ($\alpha = 99\%$)

in two different ways. The first approach uses the official realized P&L, whereas the second approach consists in separating the P&L of the previous's day portfolio and the P&L due to the intra-day trading activities.

2.1.2 The Basel III framework

The finalization of the reform for computing the market risk capital charge has taken considerable time. After the 2008 crisis, the market risk is revised by the Basel Committee, which adds new capital charges (Basel 2.5) in addition to those defined in the Basel I framework. In the same time, the Basel Committee published a new framework called Basel III, which focused on liquidity and leverage risks. In 2013, the Basel Committee launched a vast project called the fundamental review of the trading book (FRTB). During long time, the banking industry believed that these discussions were the basis of new reforms in order to prepare a Basel IV Accord. However, the Basel Committee argued that these changes are simply completing the Basel III reforms. As for the Basel I Accord, banks have the choice between two approaches for computing the capital charge:

1. a standardized method (SA-TB[21]);
2. an internal model-based approach (IMA).

Contrary to the previous framework, the SA-TB method is very important even if banks calculate the capital charge with the IMA method. Indeed, the bank must implement SA-TB in order to meet the output floor requirement[22], which is set at 72.5% in January 2027.

[21] TB means trading book.
[22] The mechanism of capital floor is explained on page 22.

2.1.2.1 Standardized approach

The standardized capital charge is the sum of three components: sensitivity-based method capital, the default risk capital (DRC) and the residual risk add-on (RRAO). The first component must be viewed as the pure market risk and is the equivalent of the capital charge for the general market risk in the Basel I Accord. The second component captures the jump-to-default risk (JTD) and replaces the specific risk that we find in the Basel I framework. The last component captures specific risks that are difficult to measure in practice.

Sensitivity-based capital requirement This method consists in calculating a capital charge for delta, vega and curvature risks, and then aggregating the three capital requirements:

$$\mathcal{K} = \mathcal{K}^{\text{Delta}} + \mathcal{K}^{\text{Vega}} + \mathcal{K}^{\text{Curvature}}$$

Seven risk classes are defined by the Basel Committee: (1) general interest rate risk (GIRR), (2) credit spread risk (CSR) on non-securitization products, (3) CSR on non-correlation trading portfolio (non-CTP), (4) CSR on correlation trading portfolio (CTP), (5) equity risk, (6) commodity risk and (7) foreign exchange risk. The sensitivities of the different instruments of one risk class are risk-weighted and then aggregated. The first level of aggregation concerns the risk buckets, defined as risk factors with common characteristics. For example, the bucket #1 for credit spread risk corresponds to all instruments that are exposed to the IG sovereign credit spread. The second level of aggregation is done by considering the different buckets that compose the risk class. For example, the credit spread risk is composed of 18 risk buckets (8 investment grade buckets, 7 high yield buckets, 2 index buckets and one other sector bucket).

For delta and vega components, we first begin to calculate the weighted sensitivity of each risk factor $\mathcal{F}_j$:

$$\text{WS}_j = S_j \cdot \text{RW}_j$$

where S_j and RW_j are the net sensitivity of the portfolio with respect to the risk factor and the risk weight of $\mathcal{F}_j$. More precisely, we have $S_j = \sum_i S_{i,j}$ where $S_{i,j}$ is the sensitivity of the instrument i with respect to $\mathcal{F}_j$. Second, we calculate the capital requirement for the risk bucket $\mathcal{B}_k$:

$$\mathcal{K}_{\mathcal{B}_k} = \sqrt{\max\left(\sum_j \text{WS}_j^2 + \sum_{j' \neq j} \rho_{j,j'}\, \text{WS}_j\, \text{WS}_{j'}, 0\right)}$$

where $\mathcal{F}_j \in \mathcal{B}_k$. We recognize the formula of a standard deviation[23]. Finally, we aggregate the different buckets for a given risk class[24]:

$$\mathcal{K}^{\text{Delta/Vega}} = \sqrt{\sum_k \mathcal{K}_{\mathcal{B}_k}^2 + \sum_{k' \neq k} \gamma_{k,k'}\, \text{WS}_{\mathcal{B}_k}\, \text{WS}_{\mathcal{B}_{k'}}}$$

where $\text{WS}_{\mathcal{B}_k} = \sum_{j \in \mathcal{B}_k} \text{WS}_j$ is the weighted sensitivity of the bucket $\mathcal{B}_k$. Again, we recognize the formula of a standard deviation. Therefore, the capital requirement for delta and vega risks can be viewed as a Gaussian risk measure with the following parameters:

1. the sensitivities S_j of the risk factors that are calculated by the bank;
2. the risk weights RW_j of the risk factors;

[23] The variance is floored at zero, because the correlation matrix formed by the cross-correlations $\rho_{j,j'}$ is not necessarily positive definite.

[24] If the term under the square root is negative, the Basel Committee proposes an alternative formula.

3. the correlation $\rho_{j,j'}$ between risk factors within a bucket;

4. the correlation $\gamma_{k,k'}$ between the risk buckets.

For the curvature risk, the methodology is different because it is based on two adverse scenarios. We note $P_i(\mathcal{F}_j)$ the price of the instrument i when the current level of the risk factor is $\mathcal{F}_j$. We calculate $P_i^+(\mathcal{F}_j) = P_i\left(\mathcal{F}_j + \Delta\mathcal{F}_j^+\right)$ and $P_i^-(\mathcal{F}_j) = P_i\left(\mathcal{F}_j - \Delta\mathcal{F}_j^-\right)$ the price of instrument i when the risk factor is shocked upward by $\Delta\mathcal{F}_j^+$ and downward by $\Delta\mathcal{F}_j^-$. The curvature risk capital requirement for the risk factor $\mathcal{F}_j$ is equal to:

$$\text{CVR}_j^{\pm} = -\sum_i \left(P_i^{\pm}(\mathcal{F}_j) - P_i(\mathcal{F}_j) - S_{i,j} \cdot \text{RW}_j^{\text{CRV}}\right)$$

where $S_{i,j}$ is the delta sensitivity[25] of instrument i with respect to the risk factor $\mathcal{F}_j$ and RW_j^{CRV} is the curvature risk weight of $\mathcal{F}_j$. CVR_j^+ and CVR_j^- play the role of WS_j in the delta/vega capital computation. The capital requirement for the bucket (or risk class) $\mathcal{B}_k$ is:

$$\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^{\pm} = \sqrt{\max\left(\sum_j \left(\max\left(\text{CVR}_j^{\pm}, 0\right)\right)^2 + \sum_{j' \neq j} \rho_{j,j'} \psi\left(\text{CVR}_j^{\pm}, \text{CVR}_{j'}^{\pm}\right), 0\right)}$$

where $\psi(\text{CVR}_j, \text{CVR}_{j'})$ is equal to 0 if the two arguments are both negative or is equal to $\text{CVR}_j \times \text{CVR}_{j'}$ otherwise. Then, the capital requirement for the risk bucket $\mathcal{B}_k$ is the maximum of the two adverse scenarios:

$$\boldsymbol{\mathcal{K}}_{\mathcal{B}_k} = \max\left(\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^+, \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^-\right)$$

At this stage, one scenario is selected: the upward scenario if $\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^+ > \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^-$ or the downward scenario if $\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^+ < \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^-$. And we define the curvature risk $\text{CVR}_{\mathcal{B}_k}$ for each bucket as follows:

$$\begin{aligned}\text{CVR}_{\mathcal{B}_k} &= \mathbb{1}\left\{\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^+ > \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^-\right\} \cdot \sum_{j \in \mathcal{B}_k} \text{CVR}_j^+ + \\ &\quad \mathbb{1}\left\{\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^+ < \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^-\right\} \cdot \sum_{j \in \mathcal{B}_k} \text{CVR}_j^-\end{aligned}$$

Finally, the capital requirement for the curvature risk is equal to:

$$\boldsymbol{\mathcal{K}}^{\text{Curvature}} = \sqrt{\max\left(\sum_k \boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^2 + \sum_{k' \neq k} \gamma_{k,k'} \psi\left(\text{CVR}_{\mathcal{B}_k}, \text{CVR}_{\mathcal{B}_{k'}}\right), 0\right)}$$

We conclude that we use the same methodology for delta, vega and curvature risks with three main differences: the computation of the sensitivities, the scale of risk weights, and the use of two scenarios for the curvature risk.

The first step consists in defining the risk factors. The Basel Committee gives a very precise list of risk factors by asset classes (BCBS, 2019). For instance, the equity delta risk factors are the equity spot prices and the equity repo rates, the equity vega risk factors

[25] For FX and equity risk classes, $S_{i,j}$ is the delta sensitivity of instrument i. For the other risk classes, $S_{i,j}$ is the sum of delta sensitivities of instrument i with respect to the risk factor $\mathcal{F}_j$.

are the implied volatilities of options, and the equity curvature risk factors are the equity spot prices. We retrieve the notions of delta, vega and gamma that we encounter in the theory of options. In the case of the interest rate risk class (GIRR), the risk factors include the yield curve[26], a flat curve of market-implied inflation rates for each currency and some cross-currency basis risks. For the other categories, the delta risk factors are credit spread curves, commodity spot prices and exchange rates. As for equities, vega and curvature risk factors correspond to implied volatilities of options and aggregated delta risk factors.

The second step consists in calculating the sensitivities. The equity delta sensitivity of the instrument i with respect to the equity risk factor $\mathcal{F}_j$ is given by:

$$S_{i,j} = \boldsymbol{\Delta}_i(\mathcal{F}_j) \cdot \mathcal{F}_j$$

where $\boldsymbol{\Delta}_i(\mathcal{F}_j)$ measures the (discrete) delta[27] of the instrument i by shocking the equity risk factor $\mathcal{F}_j$ by 1%. If the instrument i corresponds to a stock, the sensitivity is exactly the price of this stock when the risk factor is the stock price, and zero otherwise. If the instrument i corresponds to an European option on this stock, the sensitivity is the traditional delta of the option times the stock price. The previous formula is also valid for FX and commodity risks. For interest rate and credit risks, the delta corresponds to the PV01, that is a change of the interest rate and credit spread by 1 bp. For the vega sensitivity, we have:

$$S_{i,j} = \boldsymbol{\upsilon}_i(\mathcal{F}_j) \cdot \mathcal{F}_j$$

where $\mathcal{F}_j$ is the implied volatility.

The third step consists in calculating the risk-weighted sensitivities WS_j. For that, we use the tables given in BCBS (2019). For example, the risk weight for the 3M interest rate is equal to 1.7% while the risk weight for the 30Y interest rate is equal to 1.1% (BCBS, 2019, Table 1, page 38). For equity spot prices, the risk weight goes from 15% for large cap DM indices to 70% for small cap EM stocks (BCBS, 2019, Table 10, page 47). The fourth step computes the capital charge for each bucket. In this case, we need the '*factor*' correlations $\rho_{j,j'}$ between the risk factors within the same bucket. For example, the yield curve correlations between the 10 tenors of the same currency are given in Table 2 on page 38 in BCBS (2019). For the equity risk, $\rho_{j,j'}$ goes from 7.5% to 80%. Finally, we can compute the capital by considering the '*bucket*' correlations. For example, $\gamma_{k,k'}$ is set to 50% between the different currencies in the case of the interest rate risk. We must note that the values given by the Basel Committee correspond to a medium correlation scenario. The Basel Committee observes that correlations may increase or decrease in period of a stressed market, and impose that the bank must use the maximum of capital requirement under three correlation scenarios: medium, high and low. Under the high correlation scenario, the correlations are increased: $\rho_{j,j'}^{\text{High}} = \min(1.25 \times \rho_{j,j'}, 1)$ and $\gamma_{k,k'}^{\text{High}} = \min(1.25 \times \gamma_{k,k'}, 1)$. Under the low correlation scenario, the correlations are decreased: $\rho_{j,j'}^{\text{Low}} = \max(2 \times \rho_{j,j'} - 1, 0.75 \times \rho_{j,j'})$ and $\gamma_{k,k'}^{\text{Low}} = \max(2 \times \gamma_{k,k'} - 1, 0.75 \times \gamma_{k,k'})$. Figure 2.4 shows how the medium correlation is scaled to high and low correlation scenarios.

[26] The risk factors correspond to the following tenors of the yield curve: 3M, 6M, 1Y, 2Y, 3Y, 5Y, 10Y, 15Y, 20Y and 30Y.

[27] It follows that:

$$\begin{aligned} S_{i,j} &= \frac{P_i(1.01 \cdot \mathcal{F}_j) - P_i(\mathcal{F}_j)}{1.01 \cdot \mathcal{F}_j - \mathcal{F}_j} \cdot \mathcal{F}_j \\ &= \frac{P_i(1.01 \cdot \mathcal{F}_j) - P_i(\mathcal{F}_j)}{0.01} \end{aligned}$$

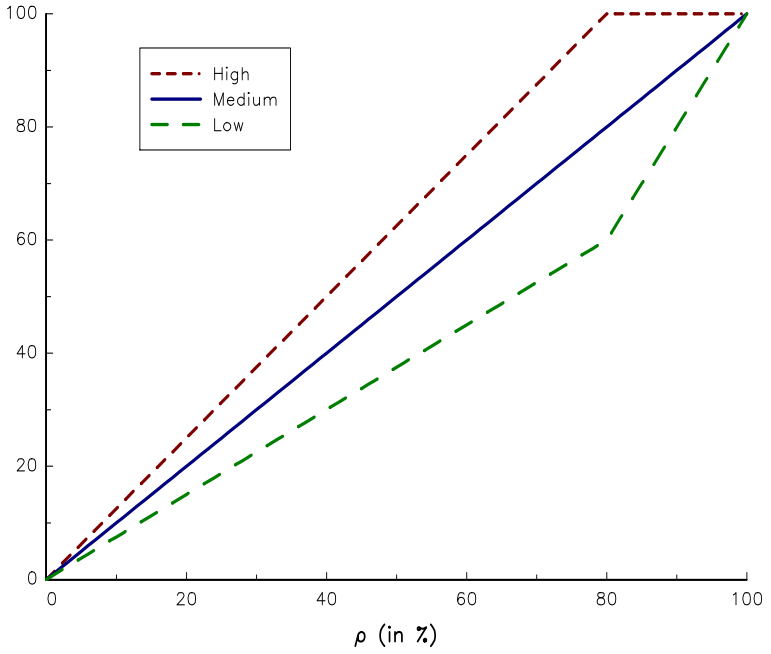


FIGURE 2.4: High, medium and low correlation scenarios

Default risk capital The gross jump-to-default (JTD) risk is computed by differentiating long and short exposures[28]:

$$\text{JTD}^{\mathcal{L}\text{ong}} = \max\left(N \cdot \text{LGD} + \Pi, 0\right)$$

and:

$$\text{JTD}^{\mathcal{S}\text{hort}} = \min\left(N \cdot \text{LGD} + \Pi, 0\right)$$

where N is the notional, LGD is the loss given default[29] and Π is the current P&L. Then, we offset long and short exposures to the same obligor under some conditions of seniority and maturity. At this stage, we obtain net JTD exposures, that can be positive (long) or negative (short). Three buckets are defined: (1) corporates, (2) sovereigns and (3) local governments and municipalities. For each bucket $\mathcal{B}_k$, the capital charge is calculated as follows:

$$\mathcal{K}_{\mathcal{B}_k}^{\text{DRC}} = \max\left(\sum_{i \in \mathcal{L}\text{ong}} \text{RW}_i \cdot \text{JTD}_i^{\text{Net}} - \text{HBR} \sum_{i \in \mathcal{S}\text{hort}} \text{RW}_i \cdot \left|\text{JTD}_i^{\text{Net}}\right|, 0\right) \tag{2.2}$$

where the risk weight depends on the rating of the obligor:

Rating	AAA	AA	A	BBB	BB	B	CCC	NR
RW	0.5%	2%	3%	6%	15%	30%	50%	15%

[28] A long exposure implies that the default results in a loss, whereas a short exposure implies that the default results in a gain.

[29] The default values are 100% for equity and non-senior debt instruments, 75% for senior debt instruments, 25% for covered bonds and 0% for FX instruments.

and HBR is the hedge benefit ratio:

$$\text{HBR} = \frac{\sum_{i \in \mathcal{L}\text{ong}} \text{JTD}_i^{\text{Net}}}{\sum_{i \in \mathcal{L}\text{ong}} \text{JTD}_i^{\text{Net}} + \sum_{i \in \mathcal{S}\text{hort}} \left|\text{JTD}_i^{\text{Net}}\right|}$$

At first sight, Equation (2.2) seems to be complicated. In order to better understand this formula, we assume that there is no short credit exposure and the P&L of each instrument is equal to zero. Therefore, the capital charge for the bucket $\mathcal{B}_k$ is equal to:

$$\boldsymbol{\mathcal{K}}_{\mathcal{B}_k}^{\text{DRC}} = \sum_{i \in \mathcal{B}_k} \underbrace{N_i \cdot \text{LGD}_i}_{\text{EAD}_i} \cdot \text{RW}_i$$

We recognize the formula for computing the credit risk capital when we replace the exposure at default by the product of the notional and the loss given default. In the case of a portfolio of loans, the exposures are always positive. In the case of a trading portfolio, we face more complex situations because we can have both long and short credit exposures. The introduction of the hedge benefit ratio allows to mitigate the risk of long credit exposures.

Remark 3 *The previous framework is valid for non-securitization instruments. For securitization, a similar approach is followed, but the LGD factor disappears in order to avoid double counting. Moreover, the treatment of offsetting differs for non-CTP and CTP products.*

Residual risk add-on The idea of this capital charge is to capture market risks which are not taken into account by the two previous methods. Residual risks concerns instruments with an exotic underlying (weather, natural disasters, longevity, etc.), payoffs that are not a linear combination of vanilla options (spread options, basket options, best-of, worst-of, etc.), or products that present significant gap, correlation or behavioral risks (digital options, barrier options, embedded options, etc.). We have:

$$\boldsymbol{\mathcal{K}}_i^{\text{RRAO}} = N_i \cdot \text{RW}_i$$

where RW_i is equal to 1% for instruments with an exotic underlying and 10 bps for the other residual risks.

2.1.2.2 Internal model-based approach

As in the first Basel Accord, the Basel III framework includes general criteria, qualitative standards, quantitative criteria, backtesting procedures and stress testing approaches. The main difference concerning general criteria is the introduction of trading desks. According to BCBS (2019), a trading desk is "*an unambiguously defined group of traders or trading accounts that implements a well-defined business strategy operating within a clear risk management structure*". Internal models are implemented at the trading desk level. Within a bank, some trading desks are then approved for the use of internal models, while other trading desks must use the SA-TB approach. The Basel Committee reinforces the role of the model validation unit, the process of the market risk measurement system (documentation, annual independent review, etc.) and the use of stress scenarios.

Capital requirement for modellable risk factors Concerning capital requirements, the value-at-risk at the 99% confidence level is replaced by the expected shortfall at the

TABLE 2.3: Liquidity horizon (Basel III)

Liquidity class k	Liquidity horizon h_k
1	10
2	20
3	40
4	60
5	120

97.5% confidence level. Moreover, the 10-day holding period is not valid for all instruments. Indeed, the expected shortfall must take into account the liquidity risk and we have:

$$\mathrm{ES}_\alpha(w) = \sqrt{\sum_{k=1}^{5}\left(\mathrm{ES}_\alpha(w;h_k)\sqrt{\frac{h_k-h_{k-1}}{h_1}}\right)^2}$$

where:

- $\mathrm{ES}_\alpha(w;h_1)$ is the expected shortfall of the portfolio w at horizon 10 days by considering all risk factors;
- $\mathrm{ES}_\alpha(w;h_k)$ is the expected shortfall of the portfolio w at horizon h_k days by considering the risk factors $\mathcal{F}_j$ that belongs to the liquidity class k;
- h_k is the horizon of the liquidity class k, which is given in Table 2.3 (h_0 is set to zero).

This expected shortfall framework is valid for modellable risk factors. Within this framework, all instruments are classified into 5 buckets (10, 20, 40, 60 and 120 days), which are defined by BCBS (2019) as follows:

1. Interest rates (specified currencies[30] and domestic currency of the bank), equity prices (large caps), FX rates (specified currency pairs[31]).
2. Interest rates (unspecified currencies), equity prices (small caps) and volatilities (large caps), FX rates (currency pairs), credit spreads (IG sovereigns), commodity prices (energy, carbon emissions, precious metals, non-ferrous metals).
3. FX rates (other types), FX volatilities, credit spreads (IG corporates and HY sovereigns).
4. Interest rates (other types), IR volatility, equity prices (other types) and volatilities (small caps), credit spreads (HY corporates), commodity prices (other types) and volatilities (energy, carbon emissions, precious metals, non-ferrous metals).
5. Credit spreads (other types) and credit spread volatilities, commodity volatilities and prices (other types).

The expected shortfall must reflect the risk measure for a period of stress. For that, the Basel Committee proposes an indirect approach:

$$\mathrm{ES}_\alpha(w;h) = \mathrm{ES}_\alpha^{(\text{reduced,stress})}(w;h)\cdot\min\left(\frac{\mathrm{ES}_\alpha^{(\text{full,current})}(w;h)}{\mathrm{ES}_\alpha^{(\text{reduced,current})}(w;h)},1\right)$$

[30] The specified currencies are composed of EUR, USD, GBP, AUD, JPY, SEK and CAD.

[31] They correspond to the 20 most liquid currencies: USD, EUR, JPY, GBP, AUD, CAD, CHF, MXN, CNY, NZD, RUB, HKD, SGD, TRY, KRW, SEK, ZAR, INR, NOK and BRL.

where $\mathrm{ES}_\alpha^{(\mathrm{full,current})}$ is the expected shortfall based on the current period with the full set of risk factors, $\mathrm{ES}_\alpha^{(\mathrm{reduced,current})}$ is the expected shortfall based on the current period with a restricted set of risk factors and $\mathrm{ES}_\alpha^{(\mathrm{reduced,stress})}$ is the expected shortfall based on the stress period[32] with the restricted set of risk factors. The Basel Committee recognizes that it is difficult to calculate directly $\mathrm{ES}_\alpha^{(\mathrm{full,stress})}(w;h)$ on the stress period with the full set of risk factors. Therefore, the previous formula assumes that there is a proportionality factor between the full set and the restricted set of risk factors[33]:

$$\frac{\mathrm{ES}_\alpha^{(\mathrm{full,stress})}(w;h)}{\mathrm{ES}_\alpha^{(\mathrm{full,current})}(w;h)} \approx \frac{\mathrm{ES}_\alpha^{(\mathrm{reduced,stress})}(w;h)}{\mathrm{ES}_\alpha^{(\mathrm{reduced,current})}(w;h)}$$

Example 11 *In the table below, we have calculated the 10-day expected shortfall for a given portfolio:*

Set of risk factors	*Period*	*Liquidity class*				
		1	2	3	4	5
Full	*Current*	100	75	34	12	6
Reduced	*Current*	88	63	30	7	5
Reduced	*Stress*	112	83	47	9	7

As expected, the expected shortfall decreases with the liquidity horizon, because there are less and less risk factors that belong to the liquidity class. We also verify that the ES for the reduced set of risk factors is lower than the ES for the full set of risk factors.

TABLE 2.4: Scaled expected shortfall

k	Sc_k	Full Current	Reduced Current	Reduced Stress	Full/Stress (not scaled)	Full Stress
1	1	100.00	88.00	112.00	127.27	127.27
2	1	75.00	63.00	83.00	98.81	98.81
3	$\sqrt{2}$	48.08	42.43	66.47	53.27	75.33
4	$\sqrt{2}$	16.97	9.90	12.73	15.43	21.82
5	$\sqrt{6}$	14.70	12.25	17.15	8.40	20.58
Total		135.80	117.31	155.91		180.38

Results are given in Table 2.4. For each liquidity class k, we have reported the scaling factor $Sc_k = \sqrt{(h_k - h_{k-1})/h_1}$, the scaled expected shortfall $\mathrm{ES}_\alpha^\star(w;h_k) = Sc_k \cdot \mathrm{ES}_\alpha(w;h_k)$ (columns 3, 4 and 5) and the total expected shortfall $\mathrm{ES}_\alpha(w) = \sqrt{\sum_{k=1}^5 \left(\mathrm{ES}_\alpha^\star(w;h_k)\right)^2}$. It is respectively equal to 135.80, 117.31 and 155.91 for the full/current, reduced/current and reduced/stress case. Since the proportionality factor is equal to $135.80/117.31 = 1.1576$, we deduce that the ES for the full set of risk factors and the stress period is equal to $1.1576 \times 155.91 = 180.48$. Another way to calculate the ES is first to compute the ES for the full set of risk factors and the stress period for each liquidity class k and deduce the scaled expected shortfall (columns 6 and 7). In this case, the ES for the full set of risk factors and the stress period is equal to 180.38.

[32] The bank must consider the most severe 12-month period of stress available.

[33] However, the Basel Committee indicates that the reduced set of risk factors must explain al leat 75% of the risk in periods of stress.

The final step for computing the capital requirement (also known as the 'internally modelled capital charge') is to apply this formula:

$$\text{IMCC} = \varrho \cdot \text{IMCC}_{global} + (1-\varrho) \cdot \sum_{k=1}^{5} \text{IMCC}_k$$

where ϱ is equal to 50%, IMCC_{global} is the stressed ES calculated with the internal model and cross-correlations between risk classes, IMCC_k is the stressed ES calculated at the risk class level (interest rate, equity, foreign exchange, commodity and credit spread). IMCC is then an average of two capital charges: one that takes into account cross-correlations and another one that ignores diversification effects.

Capital requirement for non-modellable risk factors Concerning non-modellable risk factors, the capital requirement is based on stress scenarios, that are equivalent to a stressed expected shortfall. The Basel Committee distinguish three types of non-modellable risk factors:

1. Non-modellable idiosyncratic credit spread risk factors ($i = 1, \ldots, m_{\text{c}}$);
2. Non-modellable idiosyncratic equity risk factors ($j = 1, \ldots, m_{\text{e}}$);
3. Remaining non-modellable risk factors ($k = 1, \ldots, m_{\text{o}}$).

The capital requirement for non-modellable risk factors is then equal to:

$$\text{SES} = \text{SES}^{\text{Credit}} + \text{SES}^{\text{Equity}} + \text{SES}^{\text{Other}}$$

where $\text{SES}^{\text{Credit}} = \sqrt{\sum_{i=1}^{m_{\text{c}}} \text{SES}_i^2}$, $\text{SES}^{\text{Equity}} = \sqrt{\sum_{j=1}^{m_e} \text{SES}_j^2}$ and:

$$\text{SES}^{\text{Other}} = \sqrt{\varrho^2 \cdot \left(\sum_{k=1}^{m_{\text{o}}} \text{SES}_k\right)^2 + (1-\varrho^2) \cdot \sum_{k=1}^{m_{\text{o}}} \text{SES}_k^2}$$

For non-modellable credit or equity risks, we assume a zero correlation. For the remaining non-modellable risks, the correlation ϱ is set to 60%. An important issue for computing SES is the liquidity horizon. The Basel Committee imposes to consider the same values used for modellable risk factors, with a floor of 20 days. For idiosyncratic credit spreads, the liquidity horizon is set to 120 days.

Capital requirement for default risk The default risk capital (DRC) is calculated using a value-at-risk model with a 99.9% confidence level. The computation must be done using the same default probabilities that are used for the IRB approach. This implies that default risk is calculated under the historical probability measure, and not under the risk-neutral probability measure. This is why market-implied default probabilities are prohibited.

Capital requirement for the market risk For eligible trading desks that are approved to use the IMA approach, the capital requirement for market risk is equal to:

$$\boldsymbol{\mathcal{K}}_t^{\text{IMA}} = \max\left(\text{IMCC}_{t-1} + \text{SES}_{t-1}, \frac{m_c \sum_{i=1}^{60} \text{IMCC}_{t-i} + \sum_{i=1}^{60} \text{SES}_{t-i}}{60}\right) + \text{DRC} \tag{2.3}$$

where $m_c = 1.5 + \xi$ and $0 \leq \xi \leq 0.5$. This formula is similar to the one defined in the Basel I Accord. We notice that the magnitude of the multiplication factor m_c has changed since we have $1.5 \leq m_c \leq 2$.

TABLE 2.5: Value of the penalty coefficient ξ in Basel III

Zone	Number of exceptions	ξ
Green	0 – 4	0.00
Amber	5	0.20
	6	0.26
	7	0.33
	8	0.38
	9	0.42
Red	10+	0.50

Backtesting The backtesting procedure continues to be based on the daily VaR with a 99% confidence level and a sample of the last 250 observations. Table 2.5 presents the definition of the color zones. We notice that the amber zone replaces the yellow zone, and the values of the penalty coefficient ξ have changed. The value of the multiplier $m_c = 1.5+\xi$ depends then on the one-year backtesting procedure at the bank-wide level. However, the bank must also conduct backtesting exercises for each eligible trading desk because of two reasons. First, the P&L attribution (PLA) is one of the pillars for the approval of trading desks by supervisory authorities. It is highly reinforced with several PLA tests, that distinguish actual P&L (including intra-day trading activities) and hypothetical P&L (static portfolio). Second, if one eligible trading desk is located in the amber zone, the formula (2.3) is modified in order to take into account a capital surcharge. Moreover, if one eligible trading desk has more than 12 exceptions[34], the bank must use the SA-TB approach for calculating the capital charge of this trading desk.

2.2 Statistical estimation methods of risk measures

We have seen that Basel I is based on the value-at-risk while Basel III uses the expected shortfall for computing the capital requirement for market risk. In this section, we define precisely what a risk measure is and we analyze the value-at-risk and the expected shortfall, which are the two regulatory risk measures. In particular, we present the three statistical approaches (historical, analytical and Monte Carlo) that are available. The last part of this section is dedicated to options and exotic products.

2.2.1 Definition

2.2.1.1 Coherent risk measures

Let $\mathcal{R}(w)$ be the risk measure of portfolio w. In this section, we define the different properties that should satisfy the risk measure $\mathcal{R}(w)$ in order to be acceptable in terms of capital allocation. Following Artzner *et al.* (1999), $\mathcal{R}$ is said to be '*coherent*' if it satisfies the following properties:

[34]The Basel Committee adds a second inclusive condition: the trading desk must have less than 30 exceptions at the 97.5% confidence level. This remark shows that the bank must in fact conduct two backtesting procedures at the trading desk level: one based at the 99% confidence level and another one based at the 97.5% confidence level.

1. Subadditivity
$$\mathcal{R}(w_1 + w_2) \leq \mathcal{R}(w_1) + \mathcal{R}(w_2)$$
The risk of two portfolios should be less than adding the risk of the two separate portfolios.

2. Homogeneity
$$\mathcal{R}(\lambda w) = \lambda \mathcal{R}(w) \quad \text{if } \lambda \geq 0$$
Leveraging or deleveraging of the portfolio increases or decreases the risk measure in the same magnitude.

3. Monotonicity
$$\text{if } w_1 \prec w_2, \text{ then } \mathcal{R}(w_1) \geq \mathcal{R}(w_2)$$
If portfolio w_2 has a better return than portfolio w_1 under all scenarios, risk measure $\mathcal{R}(w_1)$ should be higher than risk measure $\mathcal{R}(w_2)$.

4. Translation invariance
$$\text{if } m \in \mathbb{R}, \text{ then } \mathcal{R}(w + m) = \mathcal{R}(w) - m$$
Adding a cash position of amount m to the portfolio reduces the risk by m. This implies that we can hedge the risk of the portfolio by considering a capital that is equal to the risk measure:
$$\mathcal{R}(w + \mathcal{R}(w)) = \mathcal{R}(w) - \mathcal{R}(w) = 0$$

The definition of coherent risk measures led to a considerable interest in the quantitative risk management. Thus, Föllmer and Schied (2002) propose to replace the homogeneity and subadditivity conditions by a weaker condition called the convexity property:
$$\mathcal{R}(\lambda w_1 + (1-\lambda) w_2) \leq \lambda \mathcal{R}(w_1) + (1-\lambda) \mathcal{R}(w_2)$$
This condition means that diversification should not increase the risk.

We can write the loss of a portfolio as $L(w) = -P_t(w) R_{t+h}(w)$ where $P_t(w)$ and $R_{t+h}(w)$ are the current value and the future return of the portfolio. Without loss of generality[35], we assume that $P_t(w)$ is equal to 1. In this case, the expected loss $\mathbb{E}[L(w)]$ is the opposite of the expected return $\mu(w)$ of the portfolio and the standard deviation $\sigma(L(w))$ is equal to the portfolio volatility $\sigma(w)$. We consider then different risk measures:

- Volatility of the loss
$$\mathcal{R}(w) = \sigma(L(w)) = \sigma(w)$$
The volatility of the loss is the standard deviation of the portfolio loss.

- Standard deviation-based risk measure
$$\mathcal{R}(w) = \text{SD}_c(w) = \mathbb{E}[L(w)] + c \cdot \sigma(L(w)) = -\mu(w) + c \cdot \sigma(w)$$
To obtain this measure, we scale the volatility by factor $c > 0$ and subtract the expected return of the portfolio.

[35] The homogeneity property implies that:
$$\mathcal{R}\left(\frac{w}{P_t(w)}\right) = \frac{\mathcal{R}(w)}{P_t(w)}$$
We can therefore calculate the risk measure using the absolute loss (expressed in \$) or the relative loss (expressed in %). The two approaches are perfectly equivalent.

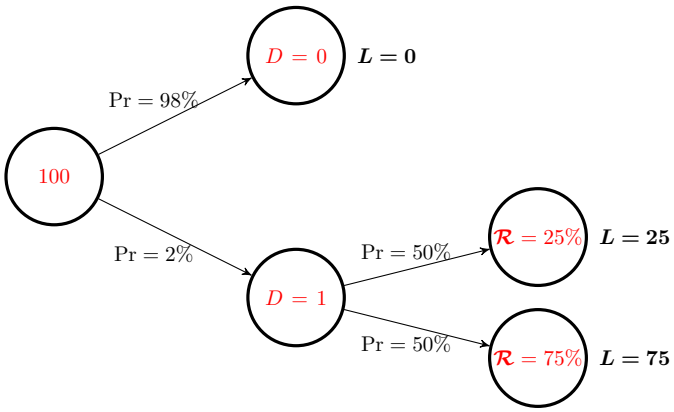
- Value-at-risk
$$\mathcal{R}(w) = \text{VaR}_\alpha(w) = \inf\{\ell : \Pr\{L(w) \leq \ell\} \geq \alpha\}$$
The value-at-risk is the α-quantile of the loss distribution $\mathbf{F}$ and we note it $\mathbf{F}^{-1}(\alpha)$.

- Expected shortfall
$$\mathcal{R}(w) = \text{ES}_\alpha(w) = \frac{1}{1-\alpha}\int_\alpha^1 \text{VaR}_u(w)\,\mathrm{d}u$$
The expected shortfall is the average of the VaRs at level α and higher (Acerbi and Tasche, 2002). We note that it is also equal to the expected loss given that the loss is beyond the value-at-risk:
$$\text{ES}_\alpha(w) = \mathbb{E}[L(w) \mid L(w) \geq \text{VaR}_\alpha(w)]$$
By definition, the expected shortfall is greater or equal than the value-at-risk for a given confidence level.

We can show that the standard deviation-based risk measure and the expected shortfall satisfy the previous coherency and convexity conditions. For the value-at-risk, the subadditivity property does not hold in general. This is a problem because the portfolio risk may have be meaningful in this case. More curiously, the volatility is not a coherent risk measure because it does not verify the translation invariance axiom.

Example 12 *We consider a* \$100 *defaultable zero-coupon bond, whose default probability is equal to* 200 *bps. We assume that the recovery rate* $\boldsymbol{\mathcal{R}}$ *is a binary random variable with* $\Pr\{\boldsymbol{\mathcal{R}} = 0.25\} = \Pr\{\boldsymbol{\mathcal{R}} = 0.75\} = 50\%$.

Below, we have represented the probability tree diagram of the loss L of the zero-coupon bond. We deduce that $\mathbf{F}(0) = \Pr\{L \leq 0\} = 98\%$, $\mathbf{F}(25) = \Pr\{L_i \leq 25\} = 99\%$ and $\mathbf{F}(75) = \Pr\{L_i \leq 75\} = 100\%$.



It follows that the 99% value-at-risk is equal to \$25, and we have:
$$\begin{aligned}\text{ES}_{99\%}(L) &= \mathbb{E}[L \mid L \geq 25] \\ &= \frac{25+75}{2} \\ &= \$50\end{aligned}$$

We assume now that the portfolio contains two zero-coupon bonds, whose default times are independent. The probability density function of (L_1, L_2) is given below:

	$L_1 = 0$	$L_1 = 25$	$L_1 = 75$	
$L_2 = 0$	96.04%	0.98%	0.98%	98.00%
$L_2 = 25$	0.98%	0.01%	0.01%	1.00%
$L_2 = 75$	0.98%	0.01%	0.01%	1.00%
	98.00%	1.00%	1.00%	

We deduce that the probability distribution function of $L = L_1 + L_2$ is:

ℓ	0	25	50	75	100	150
$\Pr\{L = \ell\}$	96.04%	1.96%	0.01%	1.96%	0.02%	0.01%
$\Pr\{L \leq \ell\}$	96.04%	98%	98.01%	99.97%	99.99%	100%

It follows that $\text{VaR}_{99\%}(L) = 75$ and:

$$\begin{aligned} \text{ES}_{99\%}(L) &= \frac{75 \times 1.96\% + 100 \times 0.02\% + 150 * 0.01\%}{1.96\% + 0.02\% + 0.01\%} \\ &= \$75.63 \end{aligned}$$

For this example, the value-at-risk does not satisfy the subadditivity property, which is not the case of the expected shortfall[36].

For this reason, the value-at-risk has been frequently criticized by academics. They also pointed out that it does not capture the tail risk of the portfolio. This led the Basel Committee to replace the 99% value-at-risk by the 97.5% expected shortfall for the internal model-based approach in Basel III (BCBS, 2019).

2.2.1.2 Value-at-risk

The value-at-risk $\text{VaR}_\alpha(w; h)$ is defined as the potential loss which the portfolio w can suffer for a given confidence level α and a fixed holding period h. Three parameters are necessary to compute this risk measure:

- the holding period h, which indicates the time period to calculate the loss;
- the confidence level α, which gives the probability that the loss is lower than the value-at-risk;
- the portfolio w, which gives the allocation in terms of risky assets and is related to the risk factors.

Without the first two parameters, it is not possible to interpret the amount of the value-at-risk, which is expressed in monetary units. For instance, a portfolio with a VaR of \$100 mn may be regarded as highly risky if the VaR corresponds to a 90% confidence level and a one-day holding period, but it may be a low risk investment if the confidence level is 99.9% and the holding period is one year.

We note $P_t(w)$ the mark-to-market value of the portfolio w at time t. The profit and loss between t and $t + h$ is equal to:

$$\Pi(w) = P_{t+h}(w) - P_t(w)$$

[36] We have $\text{VaR}_{99\%}(L_1) + \text{VaR}_{99\%}(L_2) = 50$, $\text{VaR}_{99\%}(L_1 + L_2) > \text{VaR}_{99\%}(L_1) + \text{VaR}_{99\%}(L_2)$, $\text{ES}_{99\%}(L_1) + \text{ES}_{99\%}(L_2) = 100$ and $\text{ES}_{99\%}(L_1 + L_2) < \text{ES}_{99\%}(L_1) + \text{ES}_{99\%}(L_2)$.

We define the loss of the portfolio as the opposite of the P&L: $L(w) = -\Pi(w)$. At time t, the loss is not known and is therefore random. From a statistical point of view, the value-at-risk $\text{VaR}_\alpha(w;h)$ is the quantile[37] of the loss for the probability α:

$$\Pr\{L(w) \leq \text{VaR}_\alpha(w;h)\} = \alpha$$

This means that the probability that the random loss is lower than the VaR is exactly equal to the confidence level. We finally obtain:

$$\text{VaR}_\alpha(w;h) = \mathbf{F}_L^{-1}(\alpha)$$

where $\mathbf{F}_L$ is the distribution function of the loss[38].

We notice that the previous analysis assumes that the portfolio remains unchanged between t and $t+h$. In practice, it is not the case because of trading and rebalancing activities. The holding period h depends then on the nature of the portfolio. The Basel Committee has set h to one trading day for performing the backtesting procedure in order to minimize rebalancing impacts. However, h is equal to 10 trading days for capital requirements in Basel I. It is the period which is considered necessary to ensure the rebalancing of the portfolio if it is too risky or if it costs too much regulatory capital. The confidence level α is equal to 99% meaning that there is an exception every 100 trading days. It is obvious that it does not correspond to an extreme risk measure. From the point of view of regulators, the 99% value-at-risk gives then a measure of the market risk in the case of normal conditions.

2.2.1.3 Expected shortfall

The expected shortfall $\text{ES}_\alpha(w;h)$ is defined as the expected loss beyond the value-at-risk of the portfolio:

$$\text{ES}_\alpha(w;h) = \mathbb{E}[L(w) \mid L(w) \geq \text{VaR}_\alpha(w;h)]$$

Therefore, it depends on the three parameters (h, α and w) of the VaR. Since we have $\text{ES}_\alpha(w;h) \geq \text{VaR}_\alpha(w;h)$, the expected shortfall is considered as a risk measure under more extreme conditions than the value-at-risk. By construction, we also have:

$$\alpha_1 > \alpha_2 \Rightarrow \text{ES}_{\alpha_1}(w;h) \geq \text{VaR}_{\alpha_2}(w;h)$$

However, it is impossible de compare the expected shortfall and the value-at-risk when the ES confidence level is lower than the VaR confidence level ($\alpha_1 < \alpha_2$). This is why it is difficult to compare the ES in Basel III ($\alpha = 97.5\%$) and the VaR in Basel I ($\alpha = 99\%$).

2.2.1.4 Estimator or estimate?

To calculate the value-at-risk or the expected shortfall, we first have to identify the risk factors that affect the future value of the portfolio. Their number can be large or small depending on the market, but also on the portfolio. For instance, in the case of an equity portfolio, we can use the one-factor model (CAPM), a multi-factor model (industry risk factors, Fama-French risk factors, etc.) or we can have a risk factor for each individual stock. For interest rate products, the Basel Committee imposes that the bank uses at least

[37] If the distribution of the loss is not continuous, the statistical definition of the quantile function is:

$$\text{VaR}_\alpha(w;h) = \inf\{x : \Pr\{L(w) \leq x\} \geq \alpha\}$$

[38] In a similar way, we have $\Pr\{\Pi(w) \geq -\text{VaR}_\alpha(w;h)\} = \alpha$ and $\text{VaR}_\alpha(w;h) = -\mathbf{F}_\Pi^{-1}(1-\alpha)$ where $\mathbf{F}_\Pi$ is the distribution function of the P&L.

six factors to model the yield curve risk in Basel I and ten factors in Basel III. This contrasts with currency and commodity portfolios where we must take into account one risk factor by exchange rate and by currency. Let $(\mathcal{F}_1, \ldots, \mathcal{F}_m)$ be the vector of risk factors. We assume that there is a pricing function g such that:

$$P_t(w) = g(\mathcal{F}_{1,t}, \ldots, \mathcal{F}_{m,t}; w)$$

We deduce that the expression of the random loss is the difference between the current value and the future value of the portfolio:

$$\begin{aligned} L(w) &= P_t(w) - g(\mathcal{F}_{1,t+h}, \ldots, \mathcal{F}_{m,t+h}; w) \\ &= \ell(\mathcal{F}_{1,t+h}, \ldots, \mathcal{F}_{m,t+h}; w) \end{aligned}$$

where ℓ is the loss function. The big issue is then to model the future values of risk factors. In practice, the distribution $\mathbf{F}_L$ is not known because the multidimensional distribution of the risk factors is not known. This is why we have to estimate $\mathbf{F}_L$ meaning that the calculated VaR and ES are also two estimated values:

$$\widehat{\mathrm{VaR}}_\alpha(w; h) = \hat{\mathbf{F}}_L^{-1}(\alpha) = -\hat{\mathbf{F}}_\Pi^{-1}(1-\alpha)$$

and:

$$\widehat{\mathrm{ES}}_\alpha(w; h) = \frac{1}{1-\alpha} \int_\alpha^1 \hat{\mathbf{F}}_L^{-1}(u)\, \mathrm{d}u$$

Therefore, we have to make the difference between the estimator and the estimate. Indeed, the calculated value-at-risk or expected shortfall is an estimate, meaning that it is a realization of the corresponding estimator. In practice, there are three approaches to calculate the risk measure depending on the method used to estimate $\hat{\mathbf{F}}_L$:

1. the historical value-at-risk/expected shortfall, which is also called the empirical or non-parametric VaR/ES;
2. the analytical (or parametric) value-at-risk/expected shortfall;
3. the Monte Carlo (or simulated) value-at-risk/expected shortfall.

The historical approach is the most widely used method by banks for computing the capital charge. This is an unbiased estimator, but with a large variance. On the contrary, the analytical estimator is biased, because it assumes a parametric function for the risk factors, but it has a lower variance than the historical estimator. Finally, the Monte Carlo estimator can produce an unbiased estimator with a small variance. However, it could be difficult to put in place because it requires large computational times.

Remark 4 *In this book, we use the statistical expressions* $\mathrm{VaR}_\alpha(w; h)$ *and* $\mathrm{ES}_\alpha(w; h)$ *in place of* $\widehat{\mathrm{VaR}}_\alpha(w; h)$ *and* $\widehat{\mathrm{ES}}_\alpha(w; h)$ *in order to reduce the amount of notation.*

2.2.2 Historical methods

The historical VaR corresponds to a non-parametric estimate of the value-at-risk. For that, we consider the empirical distribution of the risk factors observed in the past. Let $(\mathcal{F}_{1,s}, \ldots, \mathcal{F}_{m,s})$ be the vector of risk factors observed at time $s < t$. If we calculate the future P&L with this historical scenario, we obtain:

$$\Pi_s(w) = g(\mathcal{F}_{1,s}, \ldots, \mathcal{F}_{m,s}; w) - P_t(w)$$

If we consider n_S historical scenarios ($s = 1, \ldots, n_S$), the empirical distribution $\hat{\mathbf{F}}_\Pi$ is described by the following probability distribution:

$\Pi(w)$	$\Pi_1(w)$	$\Pi_2(w)$	$\cdots$	$\Pi_{n_S}(w)$
p_s	$1/n_S$	$1/n_S$		$1/n_S$

because each probability of occurrence is the same for all the historical scenarios. To calculate the empirical quantile $\hat{\mathbf{F}}_L^{-1}(\alpha)$, we can use two approaches: the order statistic approach and the kernel density approach.

2.2.2.1 The order statistic approach

Let $X_1, \ldots, X_n$ be a sample from a continuous distribution $\mathbf{F}$. Suppose that for a given scalar $\alpha \in]0,1[$, there exists a sequence $\{a_n\}$ such that $\sqrt{n}(a_n - n\alpha) \to 0$. Lehmann (1999) shows that:

$$\sqrt{n}\left(X_{(a_n:n)} - \mathbf{F}^{-1}(\alpha)\right) \to \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2\left(\mathbf{F}^{-1}(\alpha)\right)}\right) \tag{2.4}$$

This result implies that we can estimate the quantile $\mathbf{F}^{-1}(\alpha)$ by the mean of the $n\alpha^{\text{th}}$ order statistic. Let us apply the previous result to our problem. We calculate the order statistics associated to the P&L sample $\{\Pi_1(w), \ldots, \Pi_{n_S}(w)\}$:

$$\min_s \Pi_s(w) = \Pi_{(1:n_S)} \leq \Pi_{(2:n_S)} \leq \cdots \leq \Pi_{(n_S-1:n_S)} \leq \Pi_{(n_S:n_S)} = \max_s \Pi_s(w)$$

The value-at-risk with a confidence level α is then equal to the opposite of the $n_S(1-\alpha)^{\text{th}}$ order statistic of the P&L:

$$\text{VaR}_\alpha(w;h) = -\Pi_{(n_S(1-\alpha):n_S)} \tag{2.5}$$

If $n_S(1-\alpha)$ is not an integer, we consider the interpolation scheme:

$$\text{VaR}_\alpha(w;h) = -\left(\Pi_{(q:n_S)} + (n_S(1-\alpha) - q)\left(\Pi_{(q+1:n_S)} - \Pi_{(q:n_S)}\right)\right)$$

where $q = q_\alpha(n_S) = \lfloor n_S(1-\alpha) \rfloor$ is the integer part of $n_S(1-\alpha)$. For instance, if $n_S = 100$, the 99% value-at-risk corresponds to the largest loss. In the case where we use 250 historical scenarios, the 99% value-at-risk is the mean between the second and third largest losses:

$$\begin{aligned}
\text{VaR}_\alpha(w;h) &= -\left(\Pi_{(2:250)} + (2.5-2)\left(\Pi_{(3:250)} - \Pi_{(2:250)}\right)\right) \\
&= -\frac{1}{2}\left(\Pi_{(2:250)} + \Pi_{(3:250)}\right) \\
&= \frac{1}{2}\left(L_{(249:250)} + L_{(248:250)}\right)
\end{aligned}$$

Remark 5 *We reiterate that* $\text{VaR}_\alpha(w;h)$ *defined by Equation (2.5) is an estimator with an asymptotic variance given by Equation (2.4). Suppose that the loss of the portfolio is Gaussian and* $L(w) \sim \mathcal{N}(0,1)$. *The exact value-at-risk is* $\Phi^{-1}(\alpha)$ *and takes the values* 1.28 *or* 2.33 *if* α *is equal to 90% or 99%. The standard deviation of the estimator depends on the number* n_S *of historical scenarios:*

$$\sigma\left(\text{VaR}_\alpha(w;h)\right) \approx \frac{\sqrt{\alpha(1-\alpha)}}{\sqrt{n_S}\phi\left(\Phi^{-1}(\alpha)\right)}$$

In Figure 2.5, we have reported the density function of the VaR estimator. We notice that the estimation error decreases with n_S. *Moreover, it is lower for* $\alpha = 90\%$ *than for* $\alpha = 99\%$, *because the density of the Gaussian distribution at the point* $x = 1.28$ *is larger than at the point* $x = 2.33$.

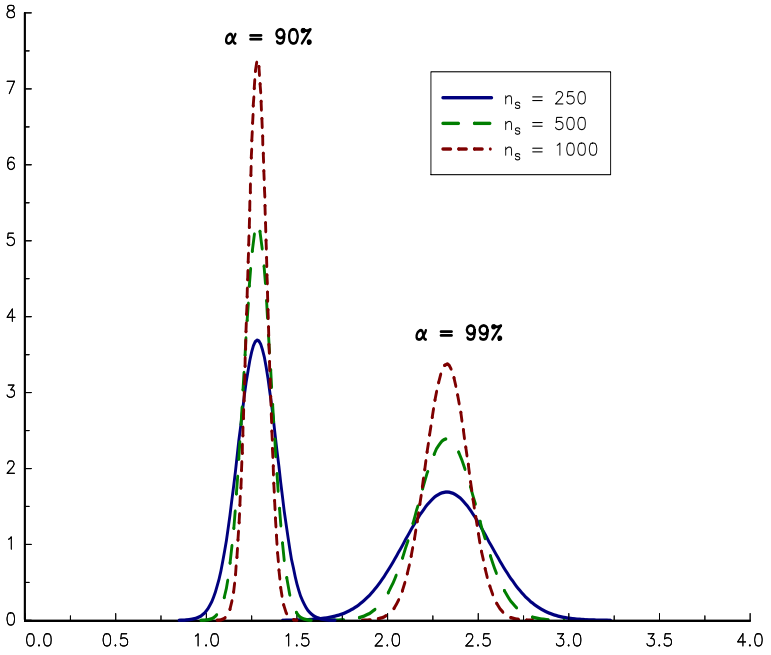


FIGURE 2.5: Density of the VaR estimator (Gaussian case)

Example 13 *We consider a portfolio composed of* 10 *stocks Apple and* 20 *stocks Coca-Cola. The current date is 2 January 2015.*

The mark-to-market of the portfolio is:

$$P_t(w) = 10 \times P_{1,t} + 20 \times P_{2,t}$$

where $P_{1,t}$ and $P_{2,t}$ are the stock prices of Apple and Coca-Cola. We assume that the market risk factors corresponds to the daily stock returns $R_{1,t}$ and $R_{2,t}$. We deduce that the P&L for the scenario s is equal to:

$$\Pi_s(w) = \underbrace{10 \times P_{1,s} + 20 \times P_{2,s}}_{g(R_{1,s},R_{2,s};w)} - P_t(w)$$

where $P_{i,s} = P_{i,t} \times (1 + R_{i,s})$ is the simulated price of stock i for the scenario s. In Table 2.6, we have reported the values of the first ten historical scenarios[39]. Using these scenarios, we can calculate the simulated price $P_{i,s}$ using the current price of the stocks (\$109.33 for Apple and \$42.14 for Coca-Cola). For instance, in the case of the 9^{th} scenario, we obtain:

$$\begin{aligned} P_{1,s} &= 109.33 \times (1 - 0.77\%) = \$108.49 \\ P_{2,s} &= 42.14 \times (1 - 1.04\%) = \$41.70 \end{aligned}$$

[39] For instance, the market risk factor for the first historical scenario and for Apple is calculated as follows:

$$R_{1,1} = \frac{109.33}{110.38} - 1 = -0.95\%$$

We then deduce the simulated mark-to-market $\mathrm{MtM}_s(w) = g(R_{1,s}, R_{2,s}; w)$, the current value of the portfolio[40] and the P&L $\Pi_s(w)$. These data are given in Table 2.7. In addition to the first ten historical scenarios, we also report the results for the six worst cases and the last scenario[41]. We notice that the largest loss is reached for the 236^{th} historical scenario at the date of 28 January 2014. If we rank the scenarios, the worst P&Ls are -84.34, -51.46, -43.31, -40.75, -35.91 and -35.42. We deduce that the daily historical VaR is equal to:

$$\mathrm{VaR}_{99\%}(w; \text{one day}) = \frac{1}{2}(51.46 + 43.31) = \$47.39$$

If we assume that $m_c = 3$, the corresponding capital charge represents 23.22% of the portfolio value:

$$\mathcal{K}_t^{\mathrm{VaR}} = 3 \times \sqrt{10} \times 47.39 = \$449.54$$

TABLE 2.6: Computation of the market risk factors $R_{1,s}$ and $R_{2,s}$

s	Date	Apple		Coca-Cola	
		Price	$R_{1,s}$	Price	$R_{2,s}$
1	2015-01-02	109.33	−0.95%	42.14	−0.19%
2	2014-12-31	110.38	−1.90%	42.22	−1.26%
3	2014-12-30	112.52	−1.22%	42.76	−0.23%
4	2014-12-29	113.91	−0.07%	42.86	−0.23%
5	2014-12-26	113.99	1.77%	42.96	0.05%
6	2014-12-24	112.01	−0.47%	42.94	−0.07%
7	2014-12-23	112.54	−0.35%	42.97	1.46%
8	2014-12-22	112.94	1.04%	42.35	0.95%
9	2014-12-19	111.78	−0.77%	41.95	−1.04%
10	2014-12-18	112.65	2.96%	42.39	2.02%

Under Basel 2.5, we have to compute a second capital charge for the stressed VaR. If we assume that the stressed period is from 9 October 2007 to 9 March 2009, we obtain 356 stressed scenarios. By applying the previous method, the six largest simulated losses are[42] 219.20 (29/09/2008), 127.84 (17/09/2008), 126.86 (07/10/2008), 124.23 (14/10/2008), 115.24 (23/01/2008) and 99.55 (29/09/2008). The 99% SVaR corresponds to the 3.56^{th} order statistic. We deduce that:

$$\begin{aligned}\mathrm{SVaR}_{99\%}(w; \text{one day}) &= 126.86 + (3.56 - 3) \times (124.23 - 126.86) \\ &= \$125.38\end{aligned}$$

It follows that:

$$\mathcal{K}_t^{\mathrm{SVaR}} = 3 \times \sqrt{10} \times 125.38 = \$1\,189.49$$

The total capital requirement under Basel 2.5 is then:

$$\mathcal{K}_t = \mathcal{K}_t^{\mathrm{VaR}} + \mathcal{K}_t^{\mathrm{SVaR}} = \$1\,639.03$$

It represents 84.6% of the current mark-to-market!

[40] We have:

$$P_t(w) = 10 \times 109.33 + 20 \times 42.14 = \$1\,936.10$$

[41] We assume that the value-at-risk is calculated using 250 historical scenarios (from 2015-01-02 to 2014-01-07).

[42] We indicate in brackets the scenario day of the loss.

TABLE 2.7: Computation of the simulated P&L $\Pi_s(w)$

s	Date	Apple		Coca-Cola		$\mathrm{MtM}_s(w)$	$\Pi_s(w)$
		$R_{1,s}$	$P_{1,s}$	$R_{2,s}$	$P_{2,s}$		
1	2015-01-02	−0.95%	108.29	−0.19%	42.06	1 924.10	−12.00
2	2014-12-31	−1.90%	107.25	−1.26%	41.61	1 904.66	−31.44
3	2014-12-30	−1.22%	108.00	−0.23%	42.04	1 920.79	−15.31
4	2014-12-29	−0.07%	109.25	−0.23%	42.04	1 933.37	−2.73
5	2014-12-26	1.77%	111.26	0.05%	42.16	1 955.82	19.72
6	2014-12-24	−0.47%	108.82	−0.07%	42.11	1 930.36	−5.74
7	2014-12-23	−0.35%	108.94	1.46%	42.76	1 944.57	8.47
8	2014-12-22	1.04%	110.46	0.95%	42.54	1 955.48	19.38
9	2014-12-19	−0.77%	108.49	−1.04%	41.70	1 918.91	−17.19
10	2014-12-18	2.96%	112.57	2.02%	42.99	1 985.51	49.41
23	2014-12-01	−3.25%	105.78	−0.62%	41.88	1 895.35	−40.75
69	2014-09-25	−3.81%	105.16	−1.16%	41.65	1 884.64	−51.46
85	2014-09-03	−4.22%	104.72	0.34%	42.28	1 892.79	−43.31
108	2014-07-31	−2.60%	106.49	−0.83%	41.79	1 900.68	−35.42
236	2014-01-28	−7.99%	100.59	0.36%	42.29	1 851.76	−84.34
242	2014-01-17	−2.45%	106.65	−1.08%	41.68	1 900.19	−35.91
250	2014-01-07	−0.72%	108.55	0.30%	42.27	1 930.79	−5.31

Remark 6 *As the previous example has shown, directional exposures are highly penalized under Basel 2.5. More generally, it is not always evident that capital requirements are lower with IMA than with SMM (Crouhy et al., 2013).*

Since the expected shortfall is the expected loss beyond the value-at-risk, it follows that the historical expected shortfall is given by:

$$\mathrm{ES}_\alpha(w;h) = \frac{1}{q_\alpha(n_S)} \sum_{s=1}^{n_S} \mathbb{1}\left\{L_s \geq \mathrm{VaR}_\alpha(w;h)\right\} \cdot L_s$$

or:

$$\mathrm{ES}_\alpha(w;h) = -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{n_S} \mathbb{1}\left\{\Pi_s \leq -\mathrm{VaR}_\alpha(w;h)\right\} \cdot \Pi_s$$

where $q_\alpha(n_S) = \lfloor n_s(1-\alpha) \rfloor$ is the integer part of $n_s(1-\alpha)$. We deduce that:

$$\mathrm{ES}_\alpha(w;h) = -\frac{1}{q_\alpha(n_S)} \sum_{i=1}^{q_\alpha(n_S)} \Pi_{(i:n_S)}$$

Computing the historical expected shortfall consists then in averaging the first $q_\alpha(n_S)$ order statistics of the P&L. For example, if n_S is equal to 250 scenarios and $\alpha = 97.5\%$, we obtain $n_s(1-\alpha) = 6.25$ and $q_\alpha(n_S) = 6$. In Basel III, computing the historical ES is then equivalent to average the 6 largest losses of the 250 historical scenarios. In the table below, we indicate the value of $q_\alpha(n_S)$ for different values of n_S and α:

α / n_S	100	150	200	250	300	350	400	450	500	1000
90.0%	9	14	19	24	29	34	39	44	49	99
95.0%	5	7	10	12	15	17	20	22	25	50
97.5%	2	3	5	6	7	8	10	11	12	25
99.0%	1	1	2	2	3	3	4	4	5	10

Let us consider Example 13 on page 68. We have found that the historical value-at-risk $\text{VaR}_{99\%}(w; \text{one day})$ of the Apple/Coca-Cola portfolio was equal to \$47.39. The 99% expected shortfall is the average of the two largest losses:

$$\text{ES}_{99\%}(w; \text{one day}) = \frac{84.34 + 51.46}{2} = \$67.90$$

However, the confidence level is set to 97.5% in Basel III, meaning that the expected shortfall is the average of the six largest losses:

$$\begin{aligned}\text{ES}_{97.5\%}(w; \text{one day}) &= \frac{84.34 + 51.46 + 43.31 + 40.75 + 35.91 + 35.42}{6} \\ &= \$48.53\end{aligned}$$

2.2.2.2 The kernel approach

Let $\{x_1, \ldots, x_n\}$ be a sample of the random variable X. In Section 10.1.4.1 on page 637, we show that we can estimate the empirical distribution $\hat{\mathbf{F}}(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\}$ by the kernel estimator:

$$\hat{\mathbf{F}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}\left(\frac{x - x_i}{\boldsymbol{h}}\right)$$

where $\mathcal{I}$ is the integrated kernel function and $\boldsymbol{h}$ is the bandwidth.

To estimate the value-at-risk with a confidence level α, Gouriéroux *et al.* (2000) solves the equation $\hat{\mathbf{F}}_L(\text{VaR}_\alpha(w; h)) = \alpha$ or:

$$\frac{1}{n_S} \sum_{s=1}^{n_S} \mathcal{I}\left(\frac{-\text{VaR}_\alpha(w; h) - \Pi_s(w)}{\boldsymbol{h}}\right) = 1 - \alpha$$

If we consider Example 13 on page 68 with the last 250 historical scenarios, we obtain the results given in Figure 2.6. We have reported the estimated distribution $\hat{\mathbf{F}}_\Pi$ of $\Pi(w)$ based on order statistic and Gaussian kernel methods[43]. We verify that the kernel approach produces a smoother distribution. If we zoom on the 1% quantile, we notice that the two methods give similar results. The daily VaR with the kernel approach is equal to \$47.44 whereas it was equal to \$47.39 with the order statistic approach.

For computing the non-parametric expected shortfall, we use the following result[44]:

$$\mathbb{E}[X \cdot \mathbb{1}\{X \leq x\}] \approx \frac{1}{n} \sum_{i=1}^{n} x_i \mathcal{I}\left(\frac{x - x_i}{\boldsymbol{h}}\right)$$

Therefore, Scaillet (2004) shows that the kernel estimator of the expected shortfall is equal to:

$$\text{ES}_\alpha(w; h) = -\frac{1}{(1-\alpha)\, n_S} \sum_{s=1}^{n_S} \Pi_s \mathcal{I}\left(\frac{-\text{VaR}_\alpha(w; h) - \Pi_s}{\boldsymbol{h}}\right)$$

In the case of the Apple/Coca-Cola example, we obtain $\text{ES}_{99\%}(w; h) = \$60.53$ and $\text{ES}_{97.5\%}(w; h) = \$45.28$. With the kernel approach, we can estimate the value-at-risk and the expected shortfall with a high confidence level α. For instance, if $\alpha = 99.25\%$, we have $(1 - \alpha)\, n_s = 0.625 < 1$. Therefore, it is impossible to estimate the VaR or the ES with 250 observations, which is not the case with the kernel estimator. In our example, we obtain $\text{VaR}_{99.75\%}(w; h) = \$58.27$ and $\text{ES}_{99.75\%}(w; h) = \$77.32$.

[43] We consider the Gaussian kernel defined by $\mathcal{K}(u) = \phi(u)$ and $\mathcal{I}(u) = \Phi(u)$. The estimated standard deviation $\hat{\sigma}(\Pi)$ is equal to 17.7147, while the bandwidth is $\boldsymbol{h} = 1.364 \times n^{-1/5} \times \hat{\sigma}(\Pi) = 8.0027$.

[44] See Exercise 2.4.12 on page 124.

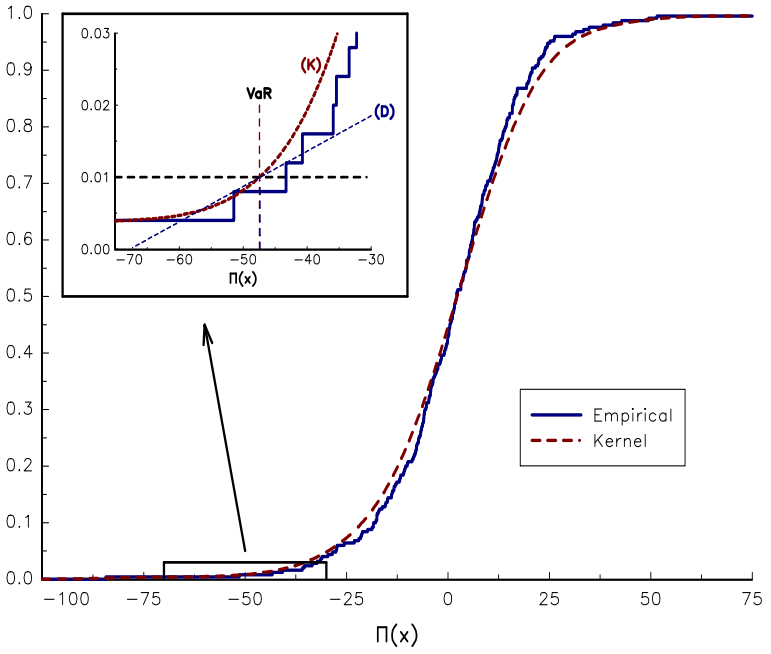


FIGURE 2.6: Kernel estimation of the historical VaR

Remark 7 *Monte Carlo simulations reveal that the kernel method reduces the variance of the VaR estimation, but not the variance of the ES estimation (Chen, 2007). In practice, the kernel approach gives similar figures than the order statistic approach, especially when the number of scenarios is large. However, the two estimators may differ in the presence of fat tails. For large confidence levels, the method based on order statistics seems to be more conservative.*

2.2.3 Analytical methods

2.2.3.1 Derivation of the closed-form formula

Gaussian value-at-risk We speak about analytical value-at-risk when we are able to find a closed-form formula of $\mathbf{F}_L^{-1}(\alpha)$. Suppose that $L(w) \sim \mathcal{N}\left(\mu(L), \sigma^2(L)\right)$. In this case, we have $\Pr\left\{L(w) \leq \mathbf{F}_L^{-1}(\alpha)\right\} = \alpha$ or:

$$\Pr\left\{\frac{L(w)-\mu(L)}{\sigma(L)} \leq \frac{\mathbf{F}_L^{-1}(\alpha)-\mu(L)}{\sigma(L)}\right\} = \alpha \Leftrightarrow \Phi\left(\frac{\mathbf{F}_L^{-1}(\alpha)-\mu(L)}{\sigma(L)}\right) = \alpha$$

We deduce that:

$$\frac{\mathbf{F}_L^{-1}(\alpha)-\mu(L)}{\sigma(L)} = \Phi^{-1}(\alpha) \Leftrightarrow \mathbf{F}_L^{-1}(\alpha) = \mu(L) + \Phi^{-1}(\alpha)\,\sigma(L)$$

The expression of the value-at-risk is then[45]:

$$\mathrm{VaR}_\alpha(w;h) = \mu(L) + \Phi^{-1}(\alpha)\,\sigma(L) \tag{2.6}$$

[45] We also have $\mathrm{VaR}_\alpha(w;h) = -\mu(\Pi) + \Phi^{-1}(\alpha)\,\sigma(\Pi)$ because the P&L $\Pi(x)$ is the opposite of the portfolio loss $L(x)$ meaning that $\mu(\Pi) = -\mu(L)$ and $\sigma(\Pi) = \sigma(L)$.

This formula is known as the Gaussian value-at-risk. For instance, if $\alpha = 99\%$ (resp. 95%), $\Phi^{-1}(\alpha)$ is equal to 2.33 (resp. 1.65) and we have:

$$\mathrm{VaR}_\alpha(w;h) = \mu(L) + 2.33 \times \sigma(L)$$

Remark 8 *We notice that the value-at-risk depends on the parameters $\mu(L)$ and $\sigma(L)$. This is why the analytical value-at-risk is also called the parametric value-at-risk. In practice, we don't know these parameters and we have to estimate them. This implies that the analytical value-at-risk is also an estimator. For the Gaussian distribution, we obtain:*

$$\widehat{\mathrm{VaR}}_\alpha(w;h) = \hat{\mu}(L) + \Phi^{-1}(\alpha)\,\hat{\sigma}(L)$$

In practice, it is extremely difficult to estimate the mean and we set $\hat{\mu}(L) = 0$.

Example 14 *We consider a short position of \$1 mn on the S&P 500 futures contract. We estimate that the annualized volatility $\hat{\sigma}_{\mathrm{SPX}}$ is equal to 35%. Calculate the daily value-at-risk with a 99% confidence level.*

The portfolio loss is equal to $L(w) = N \times R_{\mathrm{SPX}}$ where N is the exposure amount ($-\$1$ mn) and R_{SPX} is the (Gaussian) return of the S&P 500 index. We deduce that the annualized loss volatility is $\hat{\sigma}(L) = |N| \times \hat{\sigma}_{\mathrm{SPX}}$. The value-at-risk for a one-year holding period is:

$$\mathrm{VaR}_{99\%}(w;\text{one year}) = 2.33 \times 10^6 \times 0.35 = \$815\,500$$

By using the square-root-of-time rule, we deduce that:

$$\mathrm{VaR}_{99\%}(w;\text{one day}) = \frac{815\,500}{\sqrt{260}} = \$50\,575$$

This means that we have a 1% probability to lose more than \$50 575 per day.

In finance, the standard model is the Black-Scholes model where the price S_t of the asset is a geometric Brownian motion:

$$\mathrm{d}S_t = \mu_S S_t\,\mathrm{d}t + \sigma_S S_t\,\mathrm{d}W_t$$

and W_t is a Wiener process. We can show that:

$$\ln S_{t_2} - \ln S_{t_1} = \left(\mu_S - \frac{1}{2}\sigma_S^2\right)(t_2 - t_1) + \sigma_S(W_{t_2} - W_{t_1})$$

for $t_2 \geq t_1$. We have $W_{t_2} - W_{t_1} = \sqrt{t_2 - t_1}\varepsilon$ where $\varepsilon \sim \mathcal{N}(0,1)$. We finally deduce that $\mathrm{var}(\ln S_{t_2} - \ln S_{t_1}) = \sigma_S^2(t_2 - t_1)$. Let $R_S(\Delta t)$ be a sample of log-returns measured at a regular time interval Δt. It follows that:

$$\hat{\sigma}_S = \frac{1}{\sqrt{\Delta t}} \cdot \sigma(R_S(\Delta t))$$

If we consider two sample periods Δt and $\Delta t'$, we obtain the following relationship:

$$\sigma(R_S(\Delta t')) = \sqrt{\frac{\Delta t'}{\Delta t}} \cdot \sigma(R_S(\Delta t))$$

For the mean, we have $\hat{\mu}_S = \Delta t^{-1} \cdot \mathbb{E}[R_S(\Delta t)]$ and $\mathbb{E}(R_S(\Delta t')) = (\Delta t'/\Delta t) \cdot \mathbb{E}(R_S(\Delta t))$. We notice that the square-root-of-time rule is only valid for the volatility and therefore for risk measures that are linear with respect to the volatility. In practice, there is no other solution and this explains why this rule continues to be used even if we know that the approximation is poor when the portfolio loss is not Gaussian.

Gaussian expected shortfall By definition, we have:

$$\begin{aligned}\mathrm{ES}_\alpha(w) &= \mathbb{E}[L(w) \mid L(w) \geq \mathrm{VaR}_\alpha(w)] \\ &= \frac{1}{1-\alpha}\int_{\mathbf{F}_L^{-1}(\alpha)}^{\infty} x f_L(x)\,\mathrm{d}x\end{aligned}$$

where f_L and $\mathbf{F}_L$ are the density and distribution functions of the loss $L(w)$. In the Gaussian case $L(w) \sim \mathcal{N}\left(\mu(L), \sigma^2(L)\right)$, we have $\mathrm{VaR}_\alpha(w) = \mathbf{F}_L^{-1}(\alpha) = \mu(L) + \Phi^{-1}(\alpha)\sigma(L)$ and:

$$\mathrm{ES}_\alpha(w) = \frac{1}{1-\alpha}\int_{\mu(L)+\Phi^{-1}(\alpha)\sigma(L)}^{\infty} \frac{x}{\sigma(L)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu(L)}{\sigma(L)}\right)^2\right)\mathrm{d}x$$

With the variable change $t = \sigma(L)^{-1}(x - \mu(L))$, we obtain:

$$\begin{aligned}\mathrm{ES}_\alpha(w) &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} (\mu(L) + \sigma(L)t)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t \\ &= \frac{\mu(L)}{1-\alpha}\left[\Phi(t)\right]_{\Phi^{-1}(\alpha)}^{\infty} + \frac{\sigma(L)}{(1-\alpha)\sqrt{2\pi}}\int_{\Phi^{-1}(\alpha)}^{\infty} t\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t \\ &= \mu(L) + \frac{\sigma(L)}{(1-\alpha)\sqrt{2\pi}}\left[-\exp\left(-\frac{1}{2}t^2\right)\right]_{\Phi^{-1}(\alpha)}^{\infty} \\ &= \mu(L) + \frac{\sigma(L)}{(1-\alpha)\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\Phi^{-1}(\alpha)\right]^2\right)\end{aligned}$$

The expected shortfall of the portfolio w is then:

$$\mathrm{ES}_\alpha(w) = \mu(L) + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)}\sigma(L)$$

When the portfolio loss is Gaussian, the value-at-risk and the expected shortfall are both a standard deviation-based risk measure. They coincide when the scaling parameters $c_{\mathrm{VaR}} = \Phi^{-1}(\alpha_{\mathrm{VaR}})$ and $c_{\mathrm{ES}} = \phi\left(\Phi^{-1}(\alpha_{\mathrm{ES}})\right)/(1-\alpha_{\mathrm{ES}})$ are equal[46]. In Table 2.8, we report the values taken by c_{VaR} and c_{ES}. We notice that the 97.5% Gaussian expected shortfall is very close to the 99% Gaussian value-at-risk.

TABLE 2.8: Scaling factors c_{VaR} and c_{ES}

α (in %)	95.0	96.0	97.0	97.5	98.0	98.5	99.0	99.5
c_{VaR}	1.64	1.75	1.88	1.96	2.05	2.17	**2.33**	2.58
c_{ES}	2.06	2.15	2.27	**2.34**	2.42	2.52	2.67	2.89

Remark 9 *In the Gaussian case, the Basel III framework consists in replacing the scaling factor* 2.33 *by* 2.34. *In what follows, we focus on the VaR, because the ES figures can be directly deduced.*

[46]The equality is achieved when $(\alpha_{\mathrm{VaR}}, \alpha_{\mathrm{ES}})$ is equal to $(90\%, 75.44\%)$, $(95\%, 87.45\%)$, $(99\%, 97.42\%)$, $(99.9\%, 99.74\%)$, etc.

2.2.3.2 Linear factor models

We consider a portfolio of n assets and a pricing function g which is linear with respect to the asset prices. We have:

$$g(\mathcal{F}_t; w) = \sum_{i=1}^{n} w_i P_{i,t}$$

We deduce that the random P&L is:

$$\begin{aligned}\Pi(w) &= P_{t+h}(w) - P_t(w) \\ &= \sum_{i=1}^{n} w_i P_{i,t+h} - \sum_{i=1}^{n} w_i P_{i,t} \\ &= \sum_{i=1}^{n} w_i (P_{i,t+h} - P_{i,t})\end{aligned}$$

Here, $P_{i,t}$ is known whereas $P_{i,t+h}$ is stochastic. The first idea is to choose the factors as the future prices. The problem is that prices are far to be stationary meaning that we will face some issues to model the distribution $\mathbf{F}_\Pi$. Another idea is to write the future price as follows:

$$P_{i,t+h} = P_{i,t}(1 + R_{i,t+h})$$

where $R_{i,t+h}$ is the asset return between t and $t+h$. In this case, we obtain:

$$\Pi(w) = \sum_{i=1}^{n} w_i P_{i,t} R_{i,t+h}$$

In this approach, the asset returns are the market risk factors and each asset has its own risk factor.

The covariance model Let R_t be the vector of asset returns. We note $W_{i,t} = w_i P_{i,t}$ the wealth invested (or the nominal exposure) in asset i and $W_t = (W_{1,t}, \ldots, W_{n,t})$. It follows that:

$$\Pi(w) = \sum_{i=1}^{n} W_{i,t} R_{i,t+h} = W_t^\top R_{t+h}$$

If we assume that $R_{t+h} \sim \mathcal{N}(\mu, \Sigma)$, we deduce that $\mu(\Pi) = W_t^\top \mu$ and $\sigma^2(\Pi) = W_t^\top \Sigma W_t$. Using Equation (2.6), the expression of the value-at-risk is[47]:

$$\mathrm{VaR}_\alpha(w; h) = -W_t^\top \mu + \Phi^{-1}(\alpha) \sqrt{W_t^\top \Sigma W_t}$$

In this approach, we only need to estimate the covariance matrix of asset returns to compute the value-at-risk. This explains the popularity of this model, especially when the P&L of the portfolio is a linear function of the asset returns[48].

Let us consider our previous Apple/Coca-Cola example. The nominal exposures[49] are \$1 093.3 (Apple) and \$842.8 (Coca-Cola). If we consider the historical prices from 2014-01-07 to 2015-01-02, the estimated standard deviation of daily returns is equal to 1.3611% for

[47] For the expected shortfall formula, we replace $\Phi^{-1}(\alpha)$ by $\phi\left(\Phi^{-1}(\alpha)\right) / (1-\alpha)$.

[48] For instance, this approach is frequently used by asset managers to measure the risk of equity portfolios.

[49] These figures are equal to 10×109.33 and 20×42.14.

Apple and 0.9468% for Coca-Cola, whereas the cross-correlation is equal to 12.0787%. It follows that:

$$\begin{aligned}\sigma^2(\Pi) &= W_t^\top \Sigma W_t \\ &= 1\,093.3^2 \times \left(\frac{1.3611}{100}\right)^2 + 842.8^2 \times \left(\frac{0.9468}{100}\right)^2 + \\ &\quad 2 \times \frac{12.0787}{100} \times 1\,093.3 \times 842.8 \times \frac{1.3611}{100} \times \frac{0.9468}{100} \\ &= 313.80\end{aligned}$$

If we omit the term of expected return $-W_t^\top \mu$, we deduce that the 99% daily value-at-risk[50] is equal to \$41.21. We obtain a lower figure than with the historical value-at-risk, which was equal to \$47.39. We explain this result, because the Gaussian distribution underestimates the probability of extreme events and is not adapted to take into account tail risk.

The factor model We consider the standard linear factor model where asset returns R_t are related to a set of risk factors $\mathcal{F}_t = (\mathcal{F}_{1,t}, \ldots, \mathcal{F}_{m,t})$ in the following way:

$$R_t = B\mathcal{F}_t + \varepsilon_t$$

where $\mathbb{E}(\mathcal{F}_t) = \mu(\mathcal{F})$, $\text{cov}(\mathcal{F}_t) = \Omega$, $\mathbb{E}(\varepsilon_t) = \mathbf{0}$ and $\text{cov}(\varepsilon_t) = D$. $\mathcal{F}_t$ represents the common risks whereas ε_t is the vector of specific or idiosyncratic risks. This implies that $\mathcal{F}_t$ and ε_t are independent and D is a diagonal matrix[51]. B is a $(n \times m)$ matrix that measures the sensitivity of asset returns with respect to the risk factors. The first two moments of R_t are given by:

$$\mu = \mathbb{E}[R_t] = B\mu(\mathcal{F})$$

and[52]:

$$\Sigma = \text{cov}(R_t) = B\Omega B^\top + D$$

If we assume that asset returns are Gaussian, we deduce that[53]:

$$\text{VaR}_\alpha(w; h) = -W_t^\top B\mu(\mathcal{F}) + \Phi^{-1}(\alpha)\sqrt{W_t^\top (B\Omega B^\top + D) W_t}$$

The linear factor model plays a major role in financial modeling. The capital asset pricing model (CAPM) developed by Sharpe (1964) is a particular case of this model when there is a single factor, which corresponds to the market portfolio. In the arbitrage pricing theory (APT) of Ross (1976), $\mathcal{F}_t$ corresponds to a set of (unknown) arbitrage factors. They may be macro-economic, statistical or characteristic-based factors. The three-factor model of

[50] We have:

$$\text{VaR}_{99\%}(w; \text{one day}) = \Phi^{-1}(0.99)\sqrt{313.80} = \$41.21$$

[51] In the following, we note $D = \text{diag}\left(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2\right)$ where $\tilde{\sigma}_i$ is the idiosyncratic volatility of asset i.

[52] We have:

$$\begin{aligned}\Sigma &= \mathbb{E}\left[(R_t - \mu)(R_t - \mu)^\top\right] \\ &= \mathbb{E}\left[(B(\mathcal{F}_t - \mu(\mathcal{F}) + \varepsilon_t))(B(\mathcal{F}_t - \mu(\mathcal{F}) + \varepsilon_t))^\top\right] \\ &= B\mathbb{E}\left[(\mathcal{F}_t - \mu(\mathcal{F}))(\mathcal{F}_t - \mu(\mathcal{F}))^\top\right] B_t^\top + \mathbb{E}\left[\varepsilon_t \varepsilon_t^\top\right] \\ &= B\Omega B^\top + D\end{aligned}$$

[53] For the expected shortfall formula, we replace $\Phi^{-1}(\alpha)$ by $\phi\left(\Phi^{-1}(\alpha)\right)/(1-\alpha)$.

Fama and French (1993) is certainly the most famous application of APT. In this case, the factors are the market factor, the size factor corresponding to a long/short portfolio between small stocks and large stocks and the value factor, which is the return of stocks with high book-to-market values minus the return of stocks with low book-to-market values. Since its publication, the original Fama-French factor has been extended to many other factors including momentum, quality or liquidity factors[54].

BCBS (1996a) makes direct reference to CAPM. In this case, we obtain a single-factor model:

$$R_t = \alpha + \beta R_{m,t} + \varepsilon_t$$

where $R_{m,t}$ is the return of the market and $\beta = (\beta_1, \ldots, \beta_n)$ is the vector of beta coefficients. Let σ_m be the volatility of the market risk factor. We have $\operatorname{var}(R_{i,t}) = \beta_i^2 \sigma_m^2 + \tilde{\sigma}_i^2$ and $\operatorname{cov}(R_{i,t}, R_{j,t}) = \beta_i \beta_j \sigma_m^2$. By omitting the mean, we obtain:

$$\operatorname{VaR}_\alpha(w; h) = \Phi^{-1}(\alpha) \sqrt{\sigma_m^2 \left(\sum_{i=1}^n \tilde{\beta}_i^2 + 2 \sum_{j>i} \tilde{\beta}_i \tilde{\beta}_j \right) + \sum_{i=1}^n W_{i,t}^2 \tilde{\sigma}_i^2}$$

where $\tilde{\beta}_i = W_{i,t} \beta_i$ is the beta exposure of asset i expressed in \$. With the previous formula, we can calculate the VaR due to the market risk factor by omitting the specific risk[55].

If we consider our previous example, we can choose the S&P 500 index as the market risk factor. For the period 2014-01-07 to 2015-01-02, the beta coefficient is equal to 0.8307 for Apple and 0.4556 for Coca-Cola, whereas the corresponding idiosyncratic volatilities are 1.2241% (Apple) and 0.8887% (Coca-Cola). As the market volatility is estimated at 0.7165%, the daily value-at-risk is equal to \$41.68 if we include specific risks. Otherwise, it is equal to \$21.54 if we only consider the effect of the market risk factor.

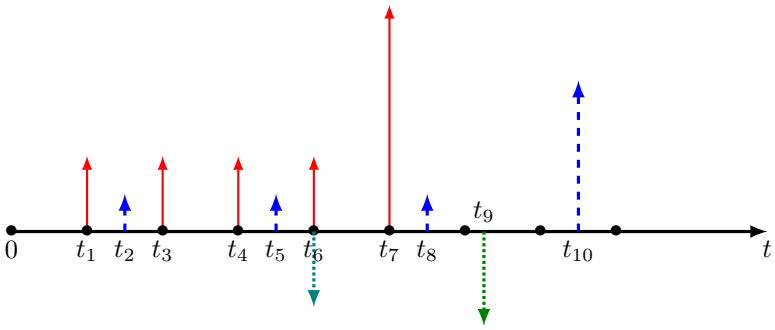


FIGURE 2.7: Cash flows of two bonds and two short exposures

Application to a bond portfolio We consider a portfolio of bonds from the same issuer. In this instance, we can model the bond portfolio by a stream of n_C coupons $C(t_m)$ with fixed dates $t_m \geq t$. Figure 2.7 presents an example of aggregating cash flows with two bonds with a fixed coupon rate and two short exposures. We note $B_t(T)$ the price of a zero-coupon bond at time t for the maturity T. We have $B_t(T) = e^{-(T-t)R_t(T)}$ where $R_t(T)$ is the zero-coupon rate. The sensitivity of the zero-coupon bond is:

$$\frac{\partial B_t(T)}{\partial R_t(T)} = -(T-t) B_t(T)$$

[54] See Cazalet and Roncalli (2014) for a survey.
[55] We set $\tilde{\sigma}_i$ to 0.

For a small change in yield, we obtain:

$$\Delta_h B_{t+h}(T) \approx -(T-t) B_t(T) \Delta_h R_{t+h}(T)$$

The value of the portfolio is:

$$P_t(w) = \sum_{m=1}^{n_C} C(t_m) B_t(t_m)$$

We deduce that:

$$\begin{aligned} \Pi(w) &= P_{t+h}(w) - P_t(w) \\ &= \sum_{m=1}^{n_C} C(t_m) (B_{t+h}(t_m) - B_t(t_m)) \end{aligned}$$

Let us consider the following approximation:

$$\begin{aligned} \Pi(w) &\approx -\sum_{m=1}^{n_C} C(t_m)(t_m - t) B_t(t_m) \Delta_h R_{t+h}(t_m) \\ &= \sum_{m=1}^{n_C} W_{i,t_m} \Delta_h R_{t+h}(t_m) \end{aligned}$$

where $W_{i,t_m} = -C(t_m)(t_m - t) B_t(t_m)$. This expression of the P&L is similar to this obtained with a portfolio of stocks. If we assume that the yield variations are Gaussian, the value-at-risk is equal to:

$$\mathrm{VaR}_\alpha(w;h) = -W_t^\top \mu + \Phi^{-1}(\alpha) \sqrt{W_t^\top \Sigma W_t}$$

where μ and Σ are the mean and the covariance matrix of the vector of yield changes $(\Delta_h R_{t+h}(t_1), \ldots, \Delta_h R_{t+h}(t_{n_c}))$.

Example 15 *We consider an exposure on a US bond at* 31 *December* 2014. *The notional of the bond is* 100 *whereas the annual coupons are equal to* 5. *The remaining maturity is five years and the fixing dates are at the end of December. The number of bonds held in the portfolio is* 10 000.

Using the US zero-coupon rates[56], we obtain the following figures for one bond at 31 December 2014:

$t_m - t$	$C(t_m)$	$R_t(t_m)$	$B_t(t_m)$	W_{t_m}
1	5	0.431%	0.996	−4.978
2	5	0.879%	0.983	−9.826
3	5	1.276%	0.962	−14.437
4	5	1.569%	0.939	−18.783
5	105	1.777%	0.915	−480.356

At the end of December 2014, the one-year zero-coupon rate is 0.431%, the two-year zero-coupon rate is 0.879%, etc. We deduce that the bond price is \$115.47 and the total exposure is \$1 154 706. Using the historical period of year 2014, we estimate the covariance matrix

[56] The data comes from the Datastream database. The zero-coupon interest rate of maturity yy years and mm months corresponds to the code USyyYmm.

between daily changes of the five zero-coupon rates[57]. We deduce that the Gaussian VaR of the bond portfolio is equal to \$4 971. If the multiplicative factor m_c is set to 3, the required capital $\mathcal{K}_t^{\text{VaR}}$ is equal to \$47 158 or 4.08% of the mark-to-market. We can compare these figures with those obtained with the historical value-at-risk. In this instance, the daily value-at-risk is higher and equal to \$5 302.

Remark 10 *The previous analysis assumes that the risk factors correspond to the yield changes, meaning that the calculated value-at-risk only concerns interest rate risk. Therefore, it cannot capture all the risks if the bond portfolio is subject to credit risk.*

Defining risk factors with the principal component analysis In the previous paragraph, the bond portfolio was very simple with only one bond and one yield curve. In practice, the bond portfolio contains streams of coupons for many maturities and yield curves. It is therefore necessary to reduce the dimension of the VaR calculation. The underlying idea is that we don't need to use the comprehensive set of zero-coupon rates to represent the set of risk factors that affects the yield curve. For instance, Nelson and Siegel (1987) propose a three-factor parametric model to define the yield curve. Another representation of the yield curve has been formulated by Litterman and Scheinkman (1991), who have proposed to characterize the factors using the principal component analysis (PCA).

Let Σ be the covariance matrix associated to the random vector X_t of dimension n. We consider the eigendecomposition $\Sigma = V\Lambda V^\top$ where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and V is an orthornormal matrix. In the principal component analysis, the (endogenous) risk factors are $\mathcal{F}_t = V^\top X_t$. The reduction method by PCA consists in selecting the first m risk factors with $m \leq n$. When applied to the value-at-risk calculation, it can be achieved in two different ways:

1. In the parametric approach, the covariance matrix Σ is replaced by $\Sigma^\star = V\Lambda^\star V^\top$ where $\Lambda^\star = \operatorname{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$.
2. In the historical method, we only consider the first m PCA factors $\mathcal{F}_t^\star = (\mathcal{F}_{1,t}, \ldots, \mathcal{F}_{m,t})$ or equivalently the modified random vector[58] $X_t^\star = V\mathcal{F}_t^\bullet$ where $\mathcal{F}_t^\bullet = (\mathcal{F}_t^\star, \mathbf{0}_{n-m})$.

If we apply this extracting method of risk factors to Example 15, the eigenvalues are equal to 47.299×10^8, 0.875×10^8, 0.166×10^8, 0.046×10^8, 0.012×10^8 whereas the matrix V of eigenvectors is:

$$V = \begin{pmatrix} 0.084 & -0.375 & -0.711 & 0.589 & 0.002 \\ 0.303 & -0.610 & -0.215 & -0.690 & -0.114 \\ 0.470 & -0.389 & 0.515 & 0.305 & 0.519 \\ 0.567 & 0.103 & 0.195 & 0.223 & -0.762 \\ 0.599 & 0.570 & -0.381 & -0.183 & 0.371 \end{pmatrix}$$

[57] The standard deviation is respectively equal to 0.746 bps for $\Delta_h R_t(t+1)$, 2.170 bps for $\Delta_h R_t(t+2)$, 3.264 bps for $\Delta_h R_t(t+3)$, 3.901 bps for $\Delta_h R_t(t+4)$ and 4.155 bps for $\Delta_h R_t(t+5)$ where h corresponds to one trading day. For the correlation matrix, we get:

$$\rho = \begin{pmatrix} 100.000 & & & & \\ 87.205 & 100.000 & & & \\ 79.809 & 97.845 & 100.000 & & \\ 75.584 & 95.270 & 98.895 & 100.000 & \\ 71.944 & 92.110 & 96.556 & 99.219 & 100.000 \end{pmatrix}$$

[58] Because we have $V^{-1} = V^\top$.

We deduce that:

$$\begin{aligned}\mathcal{F}_{1,t} &= 0.084 \times R_t(t+1) + 0.303 \times R_t(t+2) + \cdots + 0.599 \times R_t(t+5)\\ &\vdots\\ \mathcal{F}_{5,t} &= 0.002 \times R_t(t+1) - 0.114 \times R_t(t+2) + \cdots + 0.371 \times R_t(t+5)\end{aligned}$$

We retrieve the three factors of Litterman and Scheinkman, which are a level factor $\mathcal{F}_{1,t}$, a slope factor $\mathcal{F}_{2,t}$ and a convexity or curvature factor $\mathcal{F}_{3,t}$. In the following table, we report the incremental VaR of each risk factor, which is defined as difference between the value-at-risk including the risk factor and the value-at-risk excluding the risk factor:

VaR	$\mathcal{F}_{1,t}$	$\mathcal{F}_{2,t}$	$\mathcal{F}_{3,t}$	$\mathcal{F}_{4,t}$	$\mathcal{F}_{5,t}$	Sum
Gaussian	4934.71	32.94	2.86	0.17	0.19	4970.87
Historical	5857.39	−765.44	216.58	−7.98	1.41	5301.95

We notice that the value-at-risk is principally explained by the first risk factor, that is the general level of interest rates, whereas the contribution of the slope and convexity factors is small and the contribution of the remaining risk factors is marginal. This result can be explained by the long-only characteristics of the portfolio. Nevertheless, even if we consider a more complex bond portfolio, we generally observe that a number of factors is sufficient to model all the risk dimensions of the yield curve. An example is provided in Figure 2.8 with a stream of long and short exposures[59]. Using the period January 2014 – December 2014, the convergence of the value-at-risk is achieved with six factors. This result is connected to the requirement of the Basel Committee that "*banks must model the yield curve using a minimum of six risk factors*".

2.2.3.3 Volatility forecasting

The challenge of the Gaussian value-at-risk is the estimation of the loss volatility or the covariance matrix of asset returns/risk factors. The issue is not to consider the best estimate for describing the past, but to use the best estimate for forecasting the loss distribution. In the previous illustrations, we use the empirical covariance matrix or the empirical standard deviation. However, other estimators have been proposed by academics and professionals.

The original approach implemented in RiskMetrics used an exponentially weighted moving average (EWMA) for modeling the covariance between asset returns[60]:

$$\hat{\Sigma}_t = \lambda \hat{\Sigma}_{t-1} + (1-\lambda) R_{t-1} R_{t-1}^\top$$

where the parameter $\lambda \in [0,1]$ is the decay factor, which represents the degree of weighting decrease. Using a finite sample, the previous estimate is equivalent to a weighted estimator:

$$\hat{\Sigma}_t = \sum_{s=1}^{n_S} \omega_s R_{t-s} R_{t-s}^\top$$

where:

$$\omega_s = \frac{(1-\lambda)}{(1-\lambda^{n_S})} \lambda^{s-1}$$

In Figure 2.9, we represent the weights ω_s for different values of λ when the number n_S of historical scenarios is equal to 250. We verify that this estimator gives more importance to

[59] We have $C_t(t+1/2) = 400$, $C_t(t+1) = 300$, $C_t(t+3/2) = 200$, $C_t(t+2) = -200$, $C_t(t+3) = -300$, $C_t(t+4) = -500$, $C_t(t+5) = 500$, $C_t(t+6) = 400$, $C_t(t+7) = -300$, $C_t(t+10) = -700$, $C_t(t+10) = 300$ and $C_t(t+30) = 700$.

[60] We assume that the mean of expected returns is equal to $\mathbf{0}$.

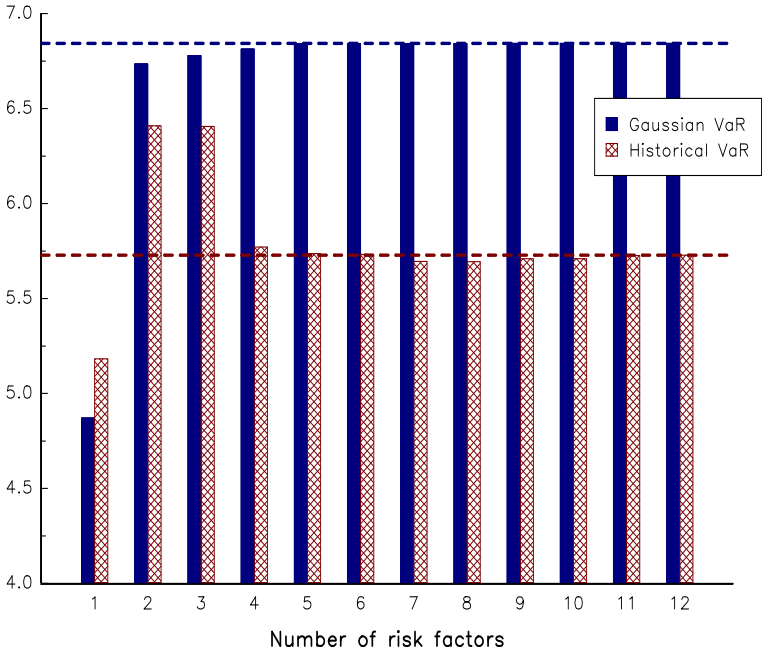


FIGURE 2.8: Convergence of the VaR with PCA risk factors

the current values than to the past values. For instance, if λ is equal to 0.94[61], 50% of the weights corresponds to the twelve first observations and the half-life is 16.7 days. We also observe that the case $\lambda = 1$ corresponds to the standard covariance estimator with uniform weights.

Another approach to model volatility in risk management is to consider that the volatility is time-varying. In 1982, Engle introduced a class of stochastic processes in order to take into account the heteroscedasticity of asset returns[62]:

$$R_{i,t} = \mu_i + \varepsilon_t \quad \text{where} \quad \varepsilon_t = \sigma_t e_t \quad \text{and} \quad e_t \sim \mathcal{N}(0,1)$$

The time-varying variance $h_t = \sigma_t^2$ satisfies the following equation:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

where $\alpha_j \geq 0$ for all $j \geq 0$. We note that the conditional variance of ε_t is not constant and depends on the past values of ε_t. A substantial impact on the asset return $R_{i,t}$ implies an increase of the conditional variance of ε_{t+1} at time $t+1$ and therefore an increase of the probability to observe another substantial impact on $R_{i,t+1}$. Therefore, this means that the volatility is persistent, which is a well-known stylized fact in finance (Chou, 1988). This type of stochastic processes, known as ARCH models (Autoregressive Conditional Heteroscedasticity), has been extended by Bollerslev (1986) in the following way:

$$h_t = \alpha_0 + \gamma_1 h_{t-1} + \gamma_2 h_{t-2} + \cdots + \gamma_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

In this case, the conditional variance depends also on its past values and we obtain a GARCH(p,q) model. If $\sum_{i=1}^{p} \gamma_i + \sum_{i=1}^{q} \alpha_i = 1$, we may show that the process ε_t^2 has a unit

[61] It was the original value of the RiskMetrics system (J.P. Morgan, 1996).
[62] See Section 10.2.4.1 on page 664 for a comprehensive presentation of ARCH and GARCH models.

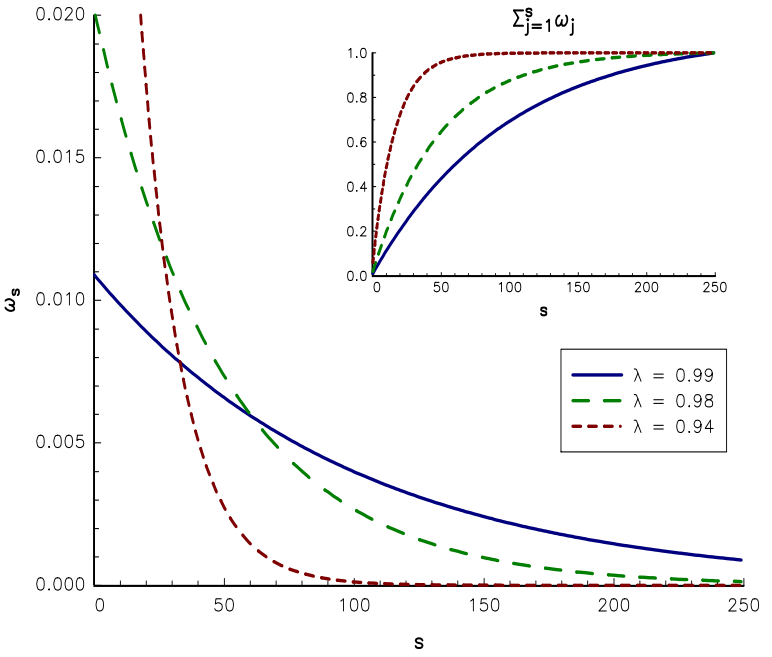


FIGURE 2.9: Weights of the EWMA estimator

root and the model is called an integrated GARCH (or IGARCH) process. If we neglect the constant term, the expression of the IGARCH(1,1) process is $h_t = (1-\alpha)\, h_{t-1} + \alpha R_{i,t-1}^2$ or equivalently:

$$\sigma_t^2 = (1-\alpha)\, \sigma_{t-1}^2 + \alpha R_{i,t-1}^2$$

This estimator is then an exponentially weighted moving average with a factor λ equal to $1-\alpha$.

In Figure 2.10, we have reported the annualized volatility of the S&P 500 index estimated using the GARCH model (first panel). The ML estimates of the parameters are $\hat{\gamma}_1 = 0.8954$ and $\hat{\alpha}_1 = 0.0929$. We verify that this estimated model is close to an IGARCH process. In the other panels, we compare the GARCH volatility with the empirical one-year historical volatility, the EWMA volatility (with $\lambda = 0.94$) and a short volatility based on 20 trading days. We observe large differences between the GARCH volatility and the one-year historical volatility, but the two others estimators (EWMA and short volatility) give similar results to the GARCH estimator. To compare the out-of-sample forecasting accuracy of these different models, we consider respectively a long and a short exposure on the S&P 500 index. At time t, we compute the value-at-risk for the next day and we compare this figure with the realized mark-to-market. Table 2.9 show the number of exceptions per year for the different models: (1) GARCH(1,1) model, (2) Gaussian value-at-risk with a one-year historical volatility, (3) EWMA model with $\lambda = 0.94$, (4) Gaussian value-at-risk with a twenty-day short volatility and (5) historical value-at-risk based on the last 260 trading days. We observe that the GARCH model produces the smallest number of exceptions, whereas the largest number of exceptions occurs in the case of the Gaussian value-at-risk with the one-year historical volatility. We also notice that the number of exceptions is smaller for the short exposure than for the long exposure. This is due to the asymmetry of returns, because extreme negative returns are larger than extreme positive returns on average.

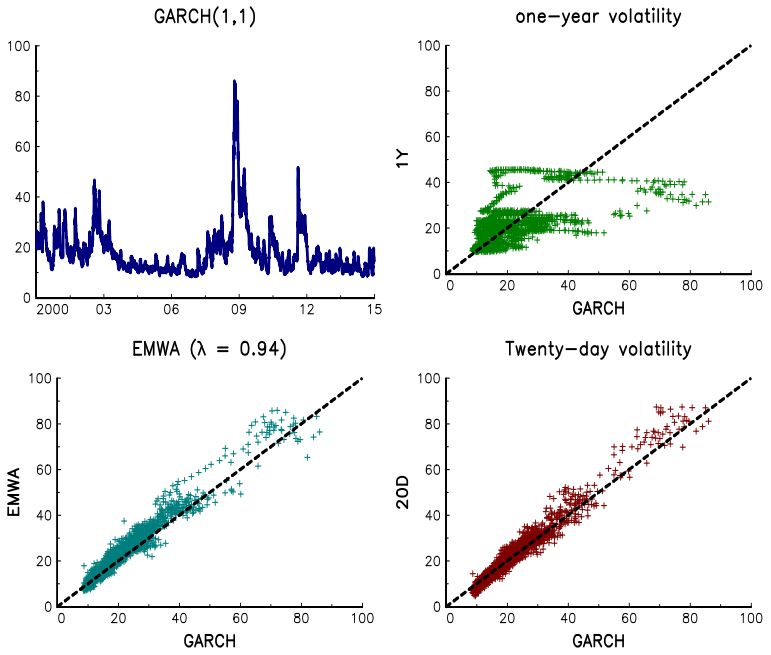


FIGURE 2.10: Comparison of GARCH and EWMA volatilities

TABLE 2.9: Number of exceptions per year for long and short exposures on the S&P 500 index

Year	Long exposure					Short exposure				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
2000	5	5	2	4	4	5	8	4	6	4
2001	4	3	2	3	2	2	4	2	5	2
2002	2	5	2	4	3	5	9	4	6	5
2003	1	0	0	2	0	1	0	1	4	0
2004	2	0	2	6	0	0	0	0	2	1
2005	1	1	2	4	3	1	4	1	6	3
2006	2	4	3	4	4	2	5	3	5	3
2007	6	15	6	10	7	1	9	0	3	7
2008	7	23	5	7	10	4	12	4	3	8
2009	5	0	1	6	0	2	2	2	3	0
2010	7	6	5	8	3	3	5	2	7	3
2011	6	8	6	7	4	2	8	1	6	3
2012	5	1	4	5	0	3	1	2	7	1
2013	4	2	3	9	2	2	2	2	4	1
2014	6	9	7	11	2	2	4	2	2	4

2.2.3.4 Extension to other probability distributions

The Gaussian value-at-risk has been strongly criticized because it depends only on the first two moments of the loss distribution. Indeed, there is a lot of evidence that asset returns and risk factors are not Gaussian (Cont, 2001). They generally present fat tails and skew effects. It is therefore interesting to consider alternative probability distributions, which are more appropriate to take into account these stylized facts.

Let $\mu_r = \mathbb{E}\left[(X - \mathbb{E}[X])^r\right]$ be the r-order centered moment of the random variable X. The skewness $\gamma_1 = \mu_3/\mu_2^{3/2}$ is the measure of the asymmetry of the loss distribution. If $\gamma_1 < 0$ (resp. $\gamma_1 > 0$), the distribution is left-skewed (resp. right-skewed) because the left (resp. right) tail is longer. For the Gaussian distribution, γ_1 is equal to zero. To characterize whether the distribution is peaked or flat relative to the normal distribution, we consider the excess kurtosis $\gamma_2 = \mu_4/\mu_2^2 - 3$. If $\gamma_2 > 0$, the distribution presents heavy tails. In the case of the Gaussian distribution, γ_2 is exactly equal to zero. We have illustrated the skewness and kurtosis statistics in Figure 2.11. Whereas we generally encounter skewness risk in credit and hedge fund portfolios, kurtosis risk has a stronger impact in equity portfolios. For example, if we consider the daily returns of the S&P 500 index, we obtain an empirical distribution[63] which has a higher kurtosis than the fitted Gaussian distribution (Figure 2.12).

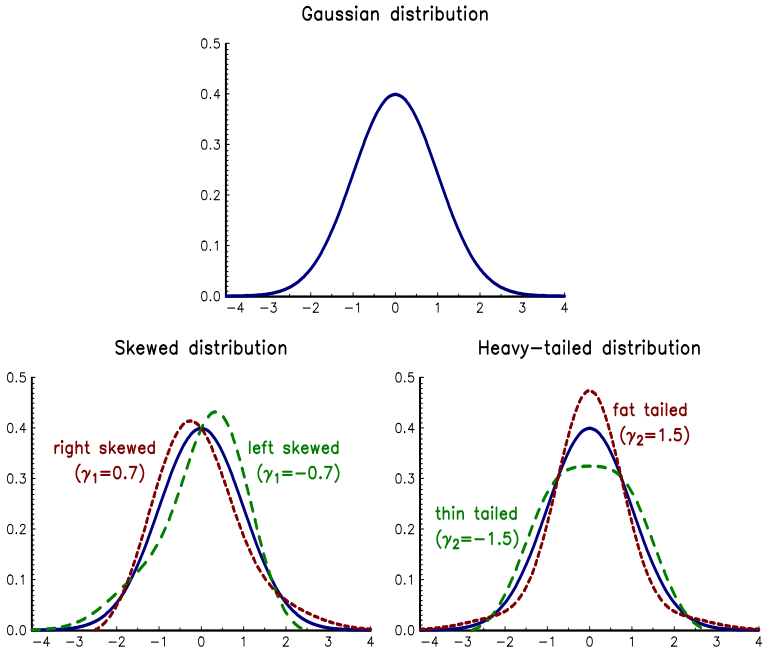


FIGURE 2.11: Examples of skewed and fat tailed distributions

An example of fat-tail distributions is the Student's t probability distribution. If $X \sim t_\nu$, we have $\mathbb{E}[X] = 0$ and $\text{var}(X) = \nu/(\nu - 2)$ for $\nu > 2$. Because X has a fixed mean and variance for a given degrees of freedom, we need to introduce location and scale parameters to model the future loss $L(w) = \xi + \omega X$. To calculate the value-at-risk, we proceed as in the Gaussian case. We have:

$$\Pr\left\{L(w) \leq \mathbf{F}_L^{-1}(\alpha)\right\} = \alpha \Leftrightarrow \Pr\left\{X \leq \frac{\mathbf{F}_L^{-1}(\alpha) - \xi}{\omega}\right\} = \alpha$$

[63] It is estimated using the kernel approach.

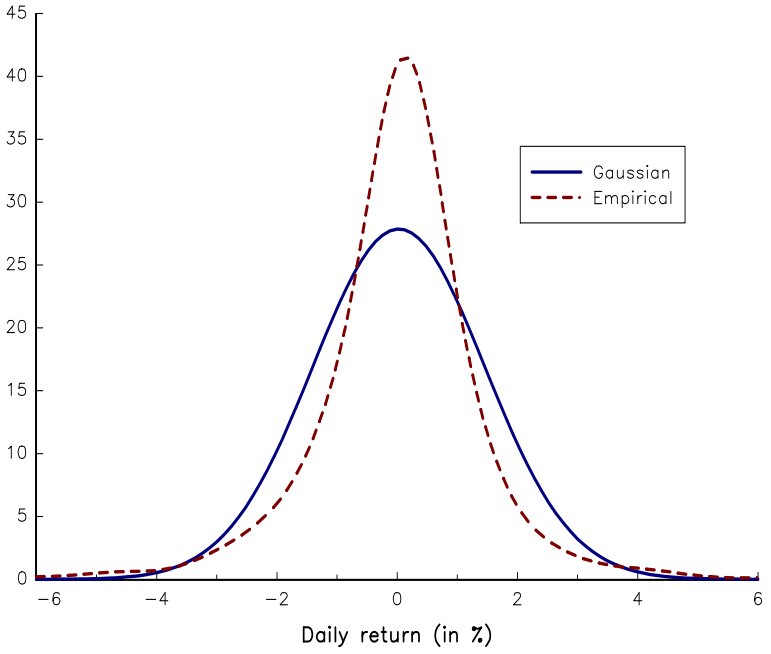


FIGURE 2.12: Estimated distribution of S&P 500 daily returns (2007-2014)

We deduce that:

$$\mathbf{T}\left(\frac{\mathbf{F}_L^{-1}(\alpha)-\xi}{\omega};\nu\right)=\alpha \Leftrightarrow \mathbf{F}_L^{-1}(\alpha)=\xi+\mathbf{T}^{-1}(\alpha;\nu)\,\omega$$

In practice, the parameters ξ and ω are estimated by the method of moments[64]. We finally deduce that:

$$\mathrm{VaR}_\alpha(w;h)=\mu(L)+\mathbf{T}^{-1}(\alpha;\nu)\,\sigma(L)\sqrt{\frac{\nu-2}{\nu}}$$

Let us illustrate the impact of the probability distribution with Example 13. By using different values of ν, we obtain the following daily VaRs:

ν	3.00	3.50	4.00	5.00	6.00	10.00	1000	∞
ω	10.23	11.60	12.53	13.72	14.46	15.84	17.70	17.71
$\mathrm{VaR}_\alpha(w;h)$	46.44	47.09	46.93	46.17	45.46	43.79	41.24	41.21

If $\nu \to \infty$, we verify that the Student's t value-at-risk converges to the Gaussian value-at-risk (\$41.21). If the degrees of freedom is equal to 4, it is closer to the historical value-at-risk (\$47.39).

We can derive closed-form formulas for several probability distributions. However, most of them are not used in practice, because these methods are not appealing from a professional point of view. Nevertheless, one approach is very popular among professionals. Using the Cornish-Fisher expansion of the normal distribution, Zangari (1996) proposes to estimate the value-at-risk in the following way:

$$\mathrm{VaR}_\alpha(w;h)=\mu(L)+\mathfrak{z}(\alpha;\gamma_1(L),\gamma_2(L))\times\sigma(L) \tag{2.7}$$

[64] We have $\mathbb{E}[\xi+\omega X]=\xi$ and $\mathrm{var}(\xi+\omega X)=\left(\omega^2\nu\right)/(\nu-2)$.

where:

$$\mathfrak{z}(\alpha;\gamma_1,\gamma_2) = z_\alpha + \frac{1}{6}\left(z_\alpha^2-1\right)\gamma_1 + \frac{1}{24}\left(z_\alpha^3-3z_\alpha\right)\gamma_2 - \frac{1}{36}\left(2z_\alpha^3-5z_\alpha\right)\gamma_1^2 \tag{2.8}$$

and $z_\alpha = \Phi^{-1}(\alpha)$. This is the same formula as the one used for the Gaussian value-at-risk but with another scaling parameter[65]. In Equation (2.7), the skewness and excess kurtosis coefficients are those of the loss distribution[66].

TABLE 2.10: Value of the Cornish-Fisher quantile $\mathfrak{z}(99\%;\gamma_1,\gamma_2)$

γ_1	γ_2							
	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00
−2.00								0.99
−1.00			1.68	1.92	2.15	2.38	2.62	2.85
−0.50		2.10	2.33	2.57	2.80	3.03	3.27	3.50
0.00	2.33	2.56	2.79	3.03	3.26	3.50	3.73	3.96
0.50		2.83	3.07	3.30	3.54	3.77	4.00	4.24
1.00			3.15	3.39	3.62	3.85	4.09	4.32
2.00								3.93

Table 2.10 shows the value of the Cornish-Fisher quantile $\mathfrak{z}(99\%;\gamma_1,\gamma_2)$ for different values of skewness and excess kurtosis. We cannot always calculate the quantile because Equation (2.8) does not define necessarily a probability distribution if the parameters γ_1 and γ_2 does not satisfy the following condition (Maillard, 2018):

$$\frac{\partial\,\mathfrak{z}(\alpha;\gamma_1,\gamma_2)}{\partial\,z_\alpha} \geq 0 \Leftrightarrow \frac{\gamma_1^2}{9} - 4\left(\frac{\gamma_2}{8}-\frac{\gamma_1^2}{6}\right)\left(1-\frac{\gamma_2}{8}+\frac{5\gamma_1^2}{36}\right) \leq 0$$

We have reported the domain of definition in the third panel in Figure 2.13. For instance, Equation (2.8) is not valid if the skewness is equal to 2 and the excess kurtosis is equal to 3. If we analyze results in Table 2.10, we do not observe that there is a monotone relationship between the skewness and the quantile. To understand this curious behavior, we report the partial derivatives of $\mathfrak{z}(\alpha;\gamma_1,\gamma_2)$ with respect to γ_1 and γ_2 in Figure 2.13. We notice that their signs depend on the confidence level α, but also on the skewness for $\partial_{\gamma_1}\mathfrak{z}(\alpha;\gamma_1,\gamma_2)$. Another drawback of the Cornish-Fisher approach concerns the statistical moments, which are not necessarily equal to the input parameters if the skewness and the kurtosis are not close to zero[67]. Contrary to what professionals commonly think, the Cornish-Fisher expansion is therefore difficult to implement.

When we consider other probability distribution than the normal distribution, the difficulty concerns the multivariate case. In the previous examples, we directly model the loss

[65] If $\gamma_1 = \gamma_2 = 0$, we retrieve the Gaussian value-at-risk because $\mathfrak{z}(\alpha;0,0) = \Phi^{-1}(\alpha)$.

[66] If we prefer to use the moments of the P&L, we have to consider the relationships $\gamma_1(Ł) = -\gamma_1(\Pi)$ and $\gamma_2(L) = \gamma_2(\Pi)$.

[67] Let Z be a Cornish-Fisher random variable satisfying $\mathbf{F}^{-1}(\alpha) = \mathfrak{z}(\alpha;\gamma_1,\gamma_2)$. A direct application of the result in Appendix A.2.2.3 gives:

$$\mathbb{E}[Z^r] = \int_0^1 \mathfrak{z}^r(\alpha;\gamma_1,\gamma_2)\,\mathrm{d}\alpha$$

Using numerical integration, we can show that $\gamma_1(Z) \neq \gamma_1$ and $\gamma_2(Z) \neq \gamma_2$ if γ_1 and γ_2 are large enough (Maillard, 2018).

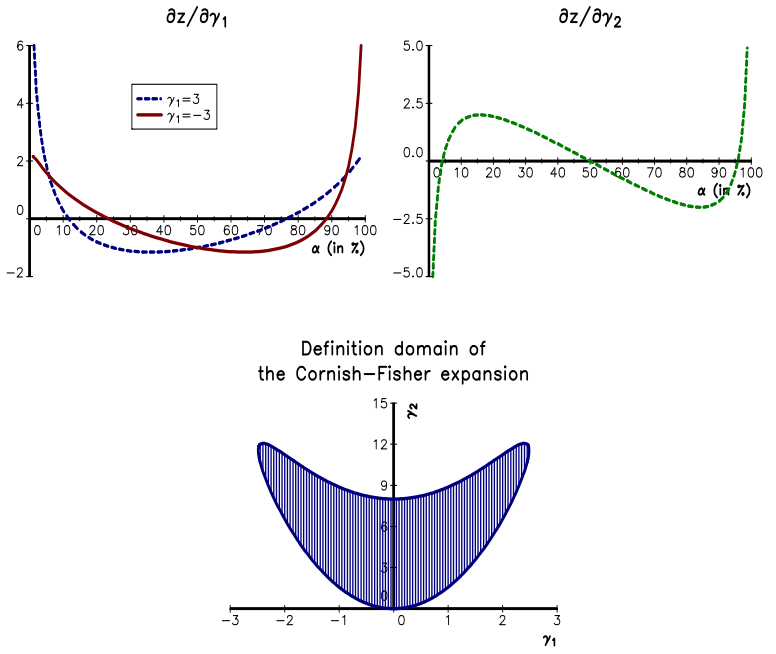


FIGURE 2.13: Derivatives and definition domain of the Cornish-Fisher expansion

distribution, that is the reduced form of the pricing system. To model the joint distribution of risk factors, two main approaches are available. The first approach considers copula functions and the value-at-risk is calculated using the Monte Carlo simulation method (see Chapters 11 and 13). The second approach consists in selecting a multivariate probability distribution, which has some appealing properties. For instance, it should be flexible enough to calibrate the first two moments of the risk factors and should also include asymmetry (positive and negative skewness) and fat tails (positive excess kurtosis) in a natural way. In order to obtain an analytical formula for the value-at-risk, it must be tractable and verify the closure property under affine transformation. This implies that if the random vector X follows a certain class of distribution, then the random vector $Y = A + BX$ belongs also to the same class. These properties reduce dramatically the set of eligible multivariate probability distributions, because the potential candidates are mostly elliptical distributions. Such examples are the skew normal and t distributions presented in Appendix A.2.1 on page 1057.

Example 16 *We consider a portfolio of three assets and assume that their annualized returns follows a multivariate skew normal distribution. The location parameters are equal to* 1%, −2% *and* 15% *whereas the scale parameters are equal to* 5%, 10% *and* 20%. *The correlation parameters to describe the dependence between the skew normal variables are given by the following matrix:*

$$\mathbb{C} = \begin{pmatrix} 1.00 & & \\ 0.35 & 1.00 & \\ 0.20 & -0.50 & 1.00 \end{pmatrix}$$

The three assets have different skewness profiles, and the shape parameters are equal to 0, 10 *and* −15.50.

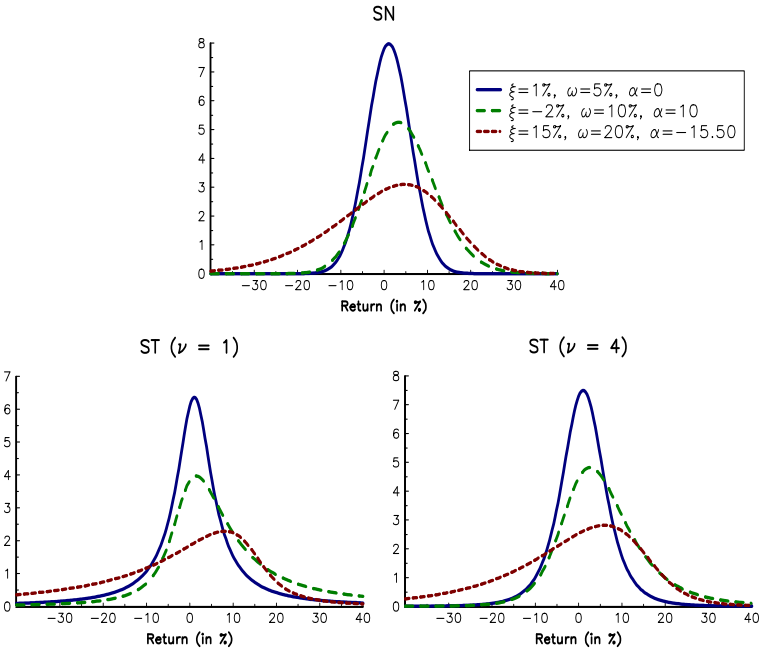


FIGURE 2.14: Skew normal and t distributions of asset returns

In Figure 2.14, we have reported the density function of the three asset returns[68]. The return of the first asset is close to be Gaussian whereas the two other assets exhibit respectively negative and positive skews. Moments are given in the table below:

Asset i	μ_i (in %)	σ_i (in %)	$\gamma_{1,i}$	$\gamma_{2,i}$
1	1.07	5.00	0.00	0.00
2	4.36	7.72	0.24	0.13
3	0.32	13.58	−0.54	0.39

Let us consider the nominal portfolio $w = (\$500, \$200, \$300)$. The annualized P&L $\Pi(w)$ is equal to $w^\top R$ where $R \sim \mathcal{SN}(\xi, \Omega, \eta)$. We deduce that $\Pi(w) \sim \mathcal{SN}(\xi_w, \omega_w, \eta_w)$ with $\xi_w = 46.00$, $\omega_w = 66.14$ and $\eta_w = -0.73$. We finally deduce that the one-year 99% value-at-risk is equal to \$123.91. If we use the multivariate skew t distribution in place of the multivariate skew normal distributions to model asset returns and if we use the same parameter values, the one-year 99% value-at-risk becomes \$558.35 for $\nu = 2$, \$215.21 for $\nu = 5$ and \$130.47 for $\nu = 50$. We verify that the skew t value-at-risk converges to the skew normal value-at-risk when the number of degrees of freedom ν tends to $+\infty$.

The choice of the probability distribution is an important issue and raises the question of model risk. In this instance, the Basel Committee justifies the introduction of the penalty coefficient in order to reduce the risk of a wrong specification (Stahl, 1997). For example, imagine that we calculate the value-at-risk with a probability distribution $\mathbf{F}$ while the true probability distribution of the portfolio loss is $\mathbf{H}$. The multiplication factor m_c defines then a capital buffer such that we are certain that the confidence level of the value-at-risk will be at least equal to α:

$$\Pr\{L(w) \leq \underbrace{m_c \cdot \mathrm{VaR}_\alpha^{(\mathbf{F})}(w)}_{\text{Capital}}\} \geq \alpha \tag{2.9}$$

[68] We also show the density function in the case of the skew t distribution with $\nu = 1$ and $\nu = 4$.

This implies that $\mathbf{H}\left(m_c \cdot \text{VaR}_\alpha^{(\mathbf{F})}(w)\right) \geq \alpha$ and $m_c \cdot \text{VaR}_\alpha^{(\mathbf{F})}(w) \geq \mathbf{H}^{-1}(\alpha)$. We finally deduce that:

$$m_c \geq \frac{\text{VaR}_\alpha^{(\mathbf{H})}(w)}{\text{VaR}_\alpha^{(\mathbf{F})}(w)}$$

In the case where $\mathbf{F}$ and $\mathbf{H}$ are the normal and Student's t distributions, we obtain[69]:

$$m_c \geq \sqrt{\frac{\nu-2}{\nu}} \frac{\mathbf{T}_\nu^{-1}(\alpha)}{\Phi^{-1}(\alpha)}$$

Below is the lower bound of m_c for different values of α and ν.

α/ν	3	4	5	6	10	50	100
90%	0.74	0.85	0.89	0.92	0.96	0.99	1.00
95%	1.13	1.14	1.12	1.10	1.06	1.01	1.01
99%	1.31	1.26	1.21	1.18	1.10	1.02	1.01
99.9%	1.91	1.64	1.48	1.38	1.20	1.03	1.02
99.99%	3.45	2.48	2.02	1.76	1.37	1.06	1.03

For instance, we have $m_c \geq 1.31$ when $\alpha = 99\%$ and $\nu = 3$.

Stahl (1997) considers the general case when $\mathbf{F}$ is the normal distribution and $\mathbf{H}$ is an unknown probability distribution. Let X be a given random variable. The Chebyshev's inequality states that:

$$\Pr\{(|X - \mu(X)| > k \cdot \sigma(X))\} \leq k^{-2}$$

for any real number $k > 0$. If we apply this theorem to the value-at-risk, we obtain[70]:

$$\Pr\left\{L(w) \leq \sqrt{\frac{1}{1-\alpha}}\sigma(L)\right\} \geq \alpha$$

Using Equation (2.9), we deduce that:

$$m_c = \sqrt{\frac{1}{1-\alpha}} \frac{\sigma(L)}{\text{VaR}_\alpha^{(\mathbf{F})}(w)}$$

In the case of the normal distribution, we finally obtain that the multiplicative factor is:

$$m_c = \frac{1}{\Phi^{-1}(\alpha)} \sqrt{\frac{1}{1-\alpha}}$$

This ratio is the multiplication factor to apply in order to be sure that the confidence level of the value-at-risk is at least equal to α if we use the normal distribution to model the portfolio loss. In the case where the probability distribution is symmetric, this ratio becomes:

$$m_c = \frac{1}{\Phi^{-1}(\alpha)} \sqrt{\frac{1}{2-2\alpha}}$$

In Table 2.11, we report the values of m_c for different confidence levels. If α is equal to 99%, the multiplication factor is equal to 3.04 if the distribution is symmetric and 4.30 otherwise.

[69] We recall that the Gaussian value-at-risk is equal to $\Phi^{-1}(\alpha)\sigma(L)$ whereas the Student's t value-at-risk is equal to $\sqrt{(\nu-2)/\nu} \cdot \mathbf{T}_\nu^{-1}(\alpha)\sigma(L)$.

[70] We set $\alpha = 1 - k^{-2}$.

TABLE 2.11: Value of the multiplication factor m_c deduced from the Chebyshev's inequality

α (in %)	90.00	95.00	99.00	99.25	99.50	99.75	99.99
Symmetric	1.74	1.92	3.04	3.36	3.88	5.04	19.01
Asymmetric	2.47	2.72	4.30	4.75	5.49	7.12	26.89

Remark 11 *Even if the previous analysis justifies the multiplication factor from a statistical point of view, we face two main issues. First, the multiplication factor assumes that the bank uses a Gaussian value-at-risk. It was the case for many banks in the early 1990s, but they use today historical value-at-risk measures. Some have suggested that the multiplication factor has been introduced in order to reduce the difference in terms of regulatory capital between SMM and IMA and it is certainly the case. The second issue concerns the specificity of the loss distribution. For many positions like long-only unlevered portfolios, the loss is bounded. If we use a Gaussian value-at-risk, the regulatory capital satisfies*[71] $\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{K}}^{\text{VaR}} + \boldsymbol{\mathcal{K}}^{\text{SVaR}} > 13.98 \cdot \sigma(L)$ *where* $\sigma(L)$ *is the non-stressed loss volatility. This implies that the value-at-risk is larger than the portfolio value if* $\sigma(L) > 7.2\%$*! There is a direct contradiction here.*

2.2.4 Monte Carlo methods

In this approach, we postulate a given probability distribution $\mathbf{H}$ for the risk factors:

$$(\mathcal{F}_{1,t+h}, \ldots, \mathcal{F}_{m,t+h}) \sim \mathbf{H}$$

Then, we simulate n_S scenarios of risk factors and calculate the simulated P&L $\Pi_s(w)$ for each scenario s. Finally, we estimate the risk measure (VaR/ES) by the method of order statistics. The Monte Carlo method to calculate the VaR/ES is therefore close to the historical method. The only difference is that it uses simulated scenarios instead of historical scenarios. This implies that the Monte Carlo approach is not limited by the number of scenarios. By construction, the Monte Carlo VaR/ES is also similar to the analytical VaR/ES, because they both specify the parametric probability distribution of risk factors. In summary, we can say that:

- the Monte Carlo VaR/ES is a historical VaR/ES with simulated scenarios;
- the Monte Carlo VaR/ES is a parametric VaR/ES for which it is difficult to find an analytical formula.

Let us consider Example 16 on page 87. The expression of the P&L is:

$$\Pi(w) = 500 \times R_1 + 200 \times R_2 + 300 \times R_3$$

Because we know that the combination of the components of a skew normal random vector is a skew normal random variable, we were able to compute the analytical quantile of $\Pi(w)$ at the 1% confidence level. Suppose now that we don't know the analytical distribution of $\Pi(w)$. We can repeat the exercise by using the Monte Carlo method. At each simulation s, we generate the random variates $(R_{1,s}, R_{2,s}, R_{3,s})$ such that:

$$(R_{1,s}, R_{2,s}, R_{3,s}) \sim SN(\xi, \Omega, \eta)$$

[71] Because we have $2 \times m_c \times 2.33 > 13.98$.

and the corresponding P&L $\Pi_s(w) = 500 \times R_{1,s} + 200 \times R_{2,s} + 300 \times R_{3,s}$. The Monte Carlo value-at-risk is the $n_s(1-\alpha)^{\text{th}}$ order statistic:

$$\widehat{\text{VaR}}_\alpha(n_S) = -\Pi_{(n_s(1-\alpha):n_s)}(w)$$

Using the law of large numbers, we can show that the MC estimator converges to the exact VaR:

$$\lim_{n_S \to \infty} \widehat{\text{VaR}}_\alpha(n_S) = \text{VaR}_\alpha$$

In Figure 2.15, we report four Monte Carlo runs with 10 000 simulated scenarios. We notice that the convergence of the Monte Carlo VaR to the analytical VaR is slow[72], because asset returns present high skewness. The convergence will be faster if the probability distribution of risk factors is close to be normal and has no fat tails.

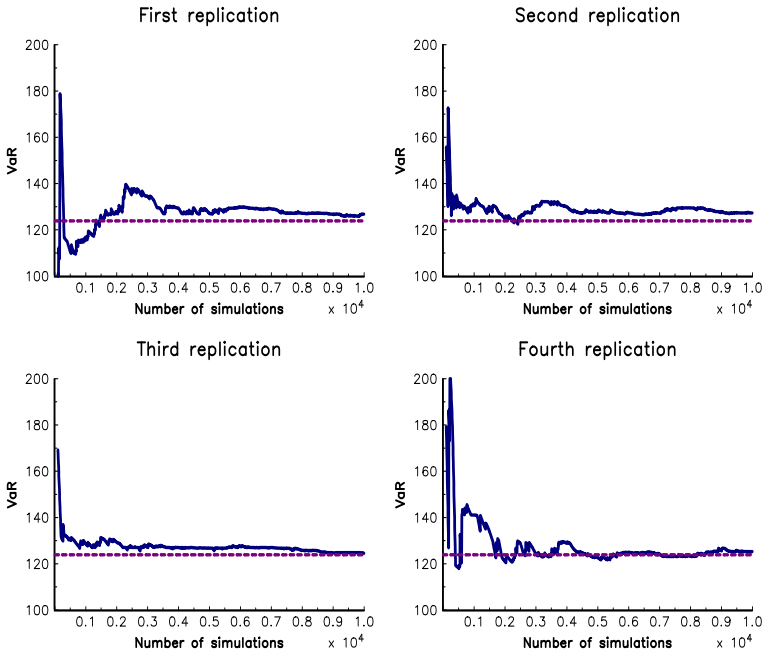


FIGURE 2.15: Convergence of the Monte Carlo VaR when asset returns are skew normal

Remark 12 *The Monte Carlo value-at-risk has been extensively studied with heavy-tailed risk factors (Dupire, 1998; Eberlein et al., 1998; Glasserman et al., 2002). In those cases, one needs to use advanced and specific methods to reduce the variance of the estimator*[73].

Example 17 *We use a variant of Example 15 on page 78. We consider that the bond is exposed to credit risk. In particular, we assume that the current default intensity of the bond issuer is equal to* 200 *bps whereas the recovery rate is equal to* 50%.

In the case of a defaultable bond, the coupons and the notional are paid until the issuer does not default whereas a recovery rate is applied if the issuer defaults before the maturity

[72] We have previously found that the exact VaR is equal to \$123.91.
[73] These techniques are presented in Chapter 13.

of the bond. If we assume that the recovery is paid at maturity, we can show that the bond price under default risk is:

$$P_t = \sum_{t_m \geq t} C(t_m) B_t(t_m) \mathbf{S}_t(t_m) + NB_t(T) (\mathbf{S}_t(T) + \boldsymbol{\mathcal{R}}_t (1 - \mathbf{S}_t(T)))$$

where $\mathbf{S}_t(t_m)$ is the survival function at time t_m and $\boldsymbol{\mathcal{R}}_t$ is the current recovery rate. We retrieve the formula of the bond price without default risk if $\mathbf{S}_t(t_m) = 1$. Using the numerical values of the parameters, the bond price is equal to \$109.75 and is lower than the non-defaultable bond price[74]. If we assume that the default time is exponential with $\mathbf{S}_t(t_m) = e^{-\lambda_t(t_m - t)}$, we have:

$$\begin{aligned} P_{t+h} &= \sum_{t_m \geq t} C(t_m) e^{(t_m - t - h) R_{t+h}(t_m)} e^{-\lambda_{t+h}(t_m - t - h)} + \\ &\quad Ne^{(T-t-h)R_{t+h}(T)} \left(\boldsymbol{\mathcal{R}}_{t+h} + (1 - \boldsymbol{\mathcal{R}}_{t+h}) e^{-\lambda_{t+h}(T-t-h)} \right) \end{aligned}$$

We define the risk factors as the zero-coupon rates, the default intensity and the recovery rate:

$$\begin{aligned} R_{t+h}(t_m) &\simeq R_t(t_m) + \Delta_h R_{t+h}(t_m) \\ \lambda_{t+h} &= \lambda_t + \Delta_h \lambda_{t+h} \\ \boldsymbol{\mathcal{R}}_{t+h} &= \boldsymbol{\mathcal{R}}_t + \Delta_h \boldsymbol{\mathcal{R}}_{t+h} \end{aligned}$$

We assume that the three risk factors are independent and follow the following probability distributions:

$$\begin{aligned} &(\Delta_h R_{t+h}(t_1), \ldots, \Delta_h R_{t+h}(t_n)) \sim \mathcal{N}(0, \Sigma) \\ &\Delta_h \lambda_{t+h} \sim \mathcal{N}(0, \sigma_\lambda^2) \\ &\Delta_h \boldsymbol{\mathcal{R}}_{t+h} \sim \mathcal{U}_{[a,b]} \end{aligned}$$

We can then simulate the daily P&L $\Pi(w) = w(P_{t+h} - P_t)$ using the above specifications. For the numerical application, we use the covariance matrix given in Footnote 57 whereas the values of σ_λ, a and b are equal to 20 bps, -10% and 10%. In Figure 2.16, we have estimated the density of the daily P&L using 100 000 simulations. IR corresponds to the case when risk factors are only the interest rates[75]. The case IR/S considers that both $R_t(t_m)$ and λ_t are risk factors whereas $\boldsymbol{\mathcal{R}}_t$ is assumed to be constant. Finally, we include the recovery risk in the case IR/S/RR. Using 10 million simulations, we find that the daily value-at-risk is equal to \$4 730 (IR), \$13 460 (IR/S) and \$18 360 (IR/S/RR). We see the impact of taking into account default risk in the calculation of the value-at-risk.

2.2.5 The case of options and derivatives

Special attention should be paid to portfolios of derivatives, because their risk management is much more complicated than a long-only portfolio of traditional assets (Duffie and Pan, 1997). They involve non-linear exposures to risk factors that are difficult to measure, they are sensitive to parameters that are not always observable and they are generally traded on OTC markets. In this section, we provide an overview of the challenges that arise when measuring and managing the risk of these assets. Chapter 9 complements it with a more exhaustive treatment of hedging and pricing issues as well as model risk.

[74] We recall that it was equal to \$115.47.

[75] This implies that we set $\Delta_h \lambda_{t+h}$ and $\Delta_h \boldsymbol{\mathcal{R}}_{t+h}$ to zero in the Monte Carlo procedure.

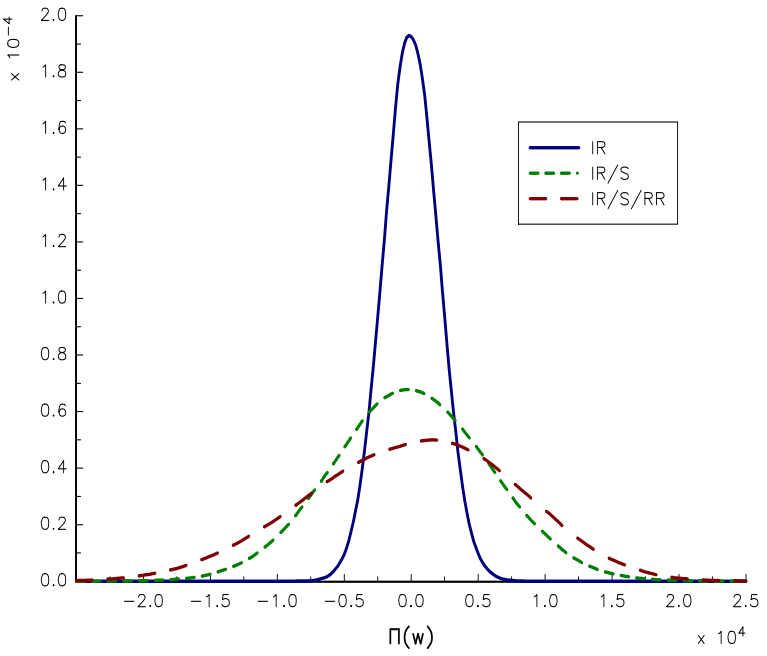


FIGURE 2.16: Probability density function of the daily P&L with credit risk

2.2.5.1 Identification of risk factors

Let us consider an example of a portfolio containing w_S stocks and w_C call options on this stock. We note S_t and $\boldsymbol{\mathcal{C}}_t$ the stock and option prices at time t. The P&L for the holding period h is equal to:

$$\Pi(w) = w_S (S_{t+h} - S_t) + w_C (\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t)$$

If we use asset returns as risk factors, we get:

$$\Pi(w) = w_S S_t R_{S,t+h} + w_C \boldsymbol{\mathcal{C}}_t R_{C,t+h}$$

where $R_{S,t+h}$ and $R_{C,t+h}$ are the returns of the stock and the option for the period $[t, t+h]$. In this approach, we identify two risk factors. The problem is that the option price $\boldsymbol{\mathcal{C}}_t$ is a non-linear function of the underlying price S_t:

$$\boldsymbol{\mathcal{C}}_t = f_C(S_t)$$

This implies that:

$$\begin{aligned} \Pi(w) &= w_S S_t R_{S,t+h} + w_C (f_C(S_{t+h}) - \boldsymbol{\mathcal{C}}_t) \\ &= w_S S_t R_{S,t+h} + w_C (f_C(S_t(1 + R_{S,t+h})) - \boldsymbol{\mathcal{C}}_t) \end{aligned}$$

The P&L depends then on a single risk factor R_S. We notice that we can write the return of the option price as a non-linear function of the stock return:

$$R_{C,t+h} = \frac{f_C(S_t(1 + R_{S,t+h})) - \boldsymbol{\mathcal{C}}_t}{\boldsymbol{\mathcal{C}}_t}$$

The problem is that the probability distribution of R_C is non-stationary and depends on the value of S_t. Therefore, the risk factors cannot be the random vector (R_S, R_C) because they require too complex modeling.

Risk factors are often explicit in primary financial assets (equities, bonds, currencies), which is not the case with derivatives. Previously, we have identified the return of the underlying asset as a risk factor for the call option. In the Black-Scholes model, the price of the call option is given by:

$$C_{\text{BS}}(S_t, K, \Sigma_t, T, b_t, r_t) = S_t e^{(b_t - r_t)\tau} \Phi(d_1) - K e^{-r_t \tau} \Phi(d_2) \tag{2.10}$$

where S_t is the current price of the underlying asset, K is the option strike, Σ_t is the volatility parameter, T is the maturity date, b_t is the cost-of-carry[76] and r_t is the interest rate. The parameter $\tau = T - t$ is the time to maturity whereas the coefficients d_1 and d_2 are defined as follows:

$$\begin{aligned} d_1 &= \frac{1}{\Sigma_t \sqrt{\tau}} \left(\ln \frac{S_t}{K} + b_t \tau \right) + \frac{1}{2} \Sigma_t \sqrt{\tau} \\ d_2 &= d_1 - \Sigma_t \sqrt{\tau} \end{aligned}$$

We can then write the option price as follows:

$$\boldsymbol{C}_t = f_{\text{BS}}(\theta_{\text{contract}}; \theta)$$

where θ_{contract} are the parameters of the contract (strike K and maturity T) and θ are the other parameters than can be objective as the underlying price S_t or subjective as the volatility Σ_t. Any one of these parameters θ may serve as risk factors:

- S_t is obviously a risk factor;
- if Σ_t is not constant, the option price may be sensitive to the volatility risk;
- the option may be impacted by changes in the interest rate or the cost-of-carry.

The risk manager faces here a big issue, because the risk measure will depend on the choice of the risk factors[77]. A typical example is the volatility parameter. We observe a difference between the historical volatility $\hat{\sigma}_t$ and the Black-Scholes volatility Σ_t. Because this implied volatility is not a market price, its value will depend on the option model and the assumptions which are required to calibrate it. For instance, it will be different if we use a stochastic volatility model or a local volatility model. Even if two banks use the same model, they will certainly obtain two different values of the implied volatility, because there is little possibility that they exactly follow the same calibration procedure.

With the underlying asset S_t, the implied volatility Σ_t is the most important risk factor, but other risk factors may be determinant. They concern the dividend risk for equity options, the yield curve risk for interest rate options, the term structure for commodity options or the correlation risk for basket options. In fact, the choice of risk factors is not always obvious because it is driven by the pricing model and the characteristics of the option. We will take a closer look at this point in Chapter 9.

2.2.5.2 Methods to calculate VaR and ES risk measures

The method of full pricing To calculate the value-at-risk or the expected shortfall of option portfolios, we use the same approaches as previously. The difference with primary

[76] The cost-of-carry depends on the underlying asset. We have $b_t = r_t$ for non-dividend stocks and total return indices, $b_t = r_t - d_t$ for stocks paying a continuous dividend yield d_t, $b_t = 0$ for forward and futures contracts and $b_t = r_t - r_t^\star$ for foreign exchange options where $r_t^\star$ is the foreign interest rate.

[77] We encounter the same difficulties for pricing and hedging purposes.

financial assets comes from the pricing function which is non-linear and more complex. In the case of historical and Monte Carlo methods, the P&L of the s^{th} scenario has the following expression:

$$\Pi_s(w) = g(\mathcal{F}_{1,s}, \ldots, \mathcal{F}_{m,s}; w) - P_t(w)$$

In the case of the introducing example, the P&L becomes then:

$$\Pi_s(w) = \begin{cases} w_S S_t R_s + w_C \left(f_C\left(S_t\left(1+R_s\right); \Sigma_t\right) - \boldsymbol{\mathcal{C}}_t\right) & \text{with one risk factor} \\ w_S S_t R_s + w_C \left(f_C\left(S_t\left(1+R_s\right), \Sigma_s\right) - \boldsymbol{\mathcal{C}}_t\right) & \text{with two risk factors} \end{cases}$$

where R_s and Σ_s are the asset return and the implied volatility generated by the s^{th} scenario. If we assume that the interest rate and the cost-of-carry are constant, the pricing function is:

$$f_C(S; \Sigma) = C_{\text{BS}}(S, K, \Sigma, T-h, b_t, r_t)$$

and we notice that the remaining maturity of the option decreases by h days. In the model with two risk factors, we have to simulate the underlying price and the implied volatility. For the single factor model, we use the current implied volatility Σ_t instead of the simulated value Σ_s.

Example 18 *We consider a long position on* 100 *call options with strike* $K = 100$. *The value of the call option is* \$4.14, *the residual maturity*[78] *is* 52 *days and the current price of the underlying asset is* \$100. *We assume that* $\Sigma_t = 20\%$ *and* $b_t = r_t = 5\%$. *The objective is to calculate the daily value-at-risk with a* 99% *confidence level and the daily expected shortfall with a* 97.5% *confidence level. For that, we consider* 250 *historical scenarios, whose first nine values are the following:*

s	1	2	3	4	5	6	7	8	9
R_s	−1.93	−0.69	−0.71	−0.73	1.22	1.01	1.04	1.08	−1.61
$\Delta\Sigma_s$	−4.42	−1.32	−3.04	2.88	−0.13	−0.08	1.29	2.93	0.85

TABLE 2.12: Daily P&L of the long position on the call option when the risk factor is the underlying price

s	R_s (in %)	S_{t+h}	$\boldsymbol{\mathcal{C}}_{t+h}$	Π_s
1	−1.93	98.07	3.09	−104.69
2	−0.69	99.31	3.72	−42.16
3	−0.71	99.29	3.71	−43.22
4	−0.73	99.27	3.70	−44.28
5	1.22	101.22	4.81	67.46
6	1.01	101.01	4.68	54.64
7	1.04	101.04	4.70	56.46
8	1.08	101.08	4.73	58.89
9	−1.61	98.39	3.25	−89.22

Using the price and the characteristics of the call option, we can show that the implied volatility Σ_t is equal to 19.99% (rounded to 20%). We first consider the case of the single risk factor. In Table 2.12, we show the values of the P&L for the first nine scenarios. As an illustration, we provide the detailed calculation for the first scenario. The asset return R_s

[78] We assume that there are 252 trading days per year.

is equal to -1.93%, thus implying that the asset price S_{t+h} is equal to $100 \times (1 - 1.93\%) = 98.07$. The residual maturity τ is equal to $51/252$ years. It follows that:

$$\begin{aligned} d_1 &= \frac{1}{20\% \times \sqrt{51/252}} \left(\ln \frac{98.07}{100} + 5\% \times \frac{51}{252} \right) + \frac{1}{2} \times 20\% \times \sqrt{\frac{51}{252}} \\ &= -0.0592 \end{aligned}$$

and:

$$d_2 = -0.0592 - 20\% \times \sqrt{\frac{51}{252}} = -0.1491$$

We deduce that:

$$\begin{aligned} \boldsymbol{\mathcal{C}}_{t+h} &= 98.07 \times e^{(5\%-5\%)\frac{51}{252}} \times \Phi(-0.0592) - 100 \times e^{5\% \times \frac{51}{252}} \times \Phi(-0.1491) \\ &= 98.07 \times 1.00 \times 0.4764 - 100 \times 1.01 \times 0.4407 \\ &= 3.093 \end{aligned}$$

The simulated P&L for the first historical scenario is then equal to:

$$\Pi_s = 100 \times (3.093 - 4.14) = -104.69$$

Based on the 250 historical scenarios, the 99% value-at-risk is equal to \$154.79, whereas the 97.5% expected shortfall is equal to \$150.04.

Remark 13 *In Figure 2.17, we illustrate that the option return R_C is not a new risk factor. We plot R_S against R_C for the 250 historical scenarios. The points are on the curve of the Black-Scholes formula. The correlation between the two returns is equal to 99.78%, which indicates that R_S and R_C are highly dependent. However, this dependence is non-linear for large positive or negative asset returns. The figure shows also the leverage effect of the call option, because R_C is not of the same order of magnitude as R_S. This illustrates the non-linear characteristic of options. A linear position with a volatility equal to 20% implies a daily VaR around 3%. In our example, the VaR is equal to 37.4% of the portfolio value, which corresponds to a linear exposure in a stock with a volatility of 259%!*

Let us consider the case with two risk factors when the implied volatility changes from t to $t + h$. We assume that the absolute variation of the implied volatility is the right risk factor to model the future implied volatility. It follows that:

$$\Sigma_{t+h} = \Sigma_t + \Delta\Sigma_s$$

In Table 2.13, we indicate the value taken by Σ_{t+h} for the first nine scenarios. This allows us to price the call option and deduce the P&L. For instance, the call option becomes[79] \$2.32 instead of \$3.09 for $s = 1$ because the implied volatility has decreased. Finally, the 99% value-at-risk is equal to \$181.70 and is larger than the previous one due to the second risk factor[80].

The method of sensitivities The previous approach is called *full pricing*, because it consists in re-pricing the option. In the method based on the Greek coefficients, the idea is to approximate the change in the option price by the Taylor expansion. For instance, we define the delta approach as follows[81]:

$$\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t \simeq \boldsymbol{\Delta}_t (S_{t+h} - S_t)$$

[79] We have $d_1 = -0.0986$, $d_2 = -0.1687$, $\Phi(d_1) = 0.4607$, $\Phi(d_2) = 0.4330$ and $\boldsymbol{\mathcal{C}}_{t+h} = 2.318$.
[80] For the expected shortfall, we have $\text{ES}_{97.5\%}(w; \text{one day}) = \$172.09$.
[81] We write the call price as the function $C_{\text{BS}}(S_t, \Sigma_t, T)$.

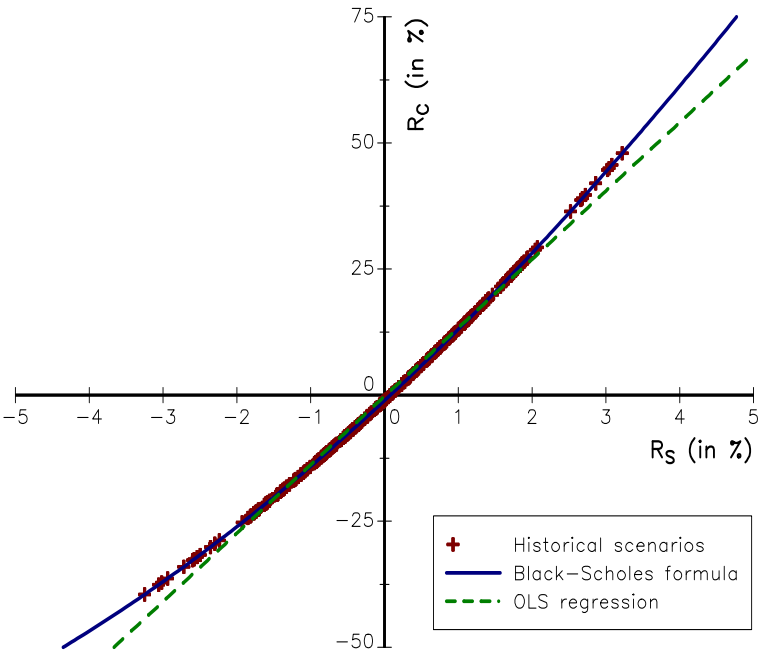


FIGURE 2.17: Relationship between the asset return R_S and the option return R_C

where $\boldsymbol{\Delta}_t$ is the option delta:

$$\boldsymbol{\Delta}_t = \frac{\partial C_{\text{BS}}(S_t, \Sigma_t, T)}{\partial S_t}$$

This approximation consists in replacing the non-linear exposure by a linear exposure with respect to the underlying price. As noted by Duffie and Pan (1997), this approach is not satisfactory because it is not accurate for large changes in the underlying price that are the most useful scenarios for calculating the risk measure. The delta approach may be implemented for the three VaR/ES methods. For instance, the Gaussian VaR of the call option is:

$$\text{VaR}_\alpha(w; h) = \Phi^{-1}(\alpha) \times |\boldsymbol{\Delta}_t| \times S_t \times \sigma(R_{S,t+h})$$

TABLE 2.13: Daily P&L of the long position on the call option when the risk factors are the underlying price and the implied volatility

s	R_s (in %)	S_{t+h}	$\Delta\Sigma_s$ (in %)	Σ_{t+h}	$\boldsymbol{\mathcal{C}}_{t+h}$	Π_s
1	−1.93	98.07	−4.42	15.58	2.32	−182.25
2	−0.69	99.31	−1.32	18.68	3.48	−65.61
3	−0.71	99.29	−3.04	16.96	3.17	−97.23
4	−0.73	99.27	2.88	22.88	4.21	6.87
5	1.22	101.22	−0.13	19.87	4.79	65.20
6	1.01	101.01	−0.08	19.92	4.67	53.24
7	1.04	101.04	1.29	21.29	4.93	79.03
8	1.08	101.08	2.93	22.93	5.24	110.21
9	−1.61	98.39	0.85	20.85	3.40	−74.21

whereas the Gaussian ES of the call option is:

$$\mathrm{ES}_\alpha(w;h) = \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \times |\boldsymbol{\Delta}_t| \times S_t \times \sigma\left(R_{S,t+h}\right)$$

If we consider the introductory example, we have:

$$\begin{aligned}\Pi(w) &= w_S\left(S_{t+h}-S_t\right) + w_C\left(\boldsymbol{\mathcal{C}}_{t+h}-\boldsymbol{\mathcal{C}}_t\right)\\ &\simeq \left(w_S + w_C\boldsymbol{\Delta}_t\right)\left(S_{t+h}-S_t\right)\\ &= \left(w_S + w_C\boldsymbol{\Delta}_t\right)S_t R_{S,t+h}\end{aligned}$$

With the delta approach, we aggregate the risk by netting the different delta exposures[82]. In particular, the portfolio is delta neutral if the net exposure is zero:

$$w_S + w_C\boldsymbol{\Delta}_t = 0 \Leftrightarrow w_S = -w_C\boldsymbol{\Delta}_t$$

With the delta approach, the VaR/ES of delta neutral portfolios is then equal to zero.

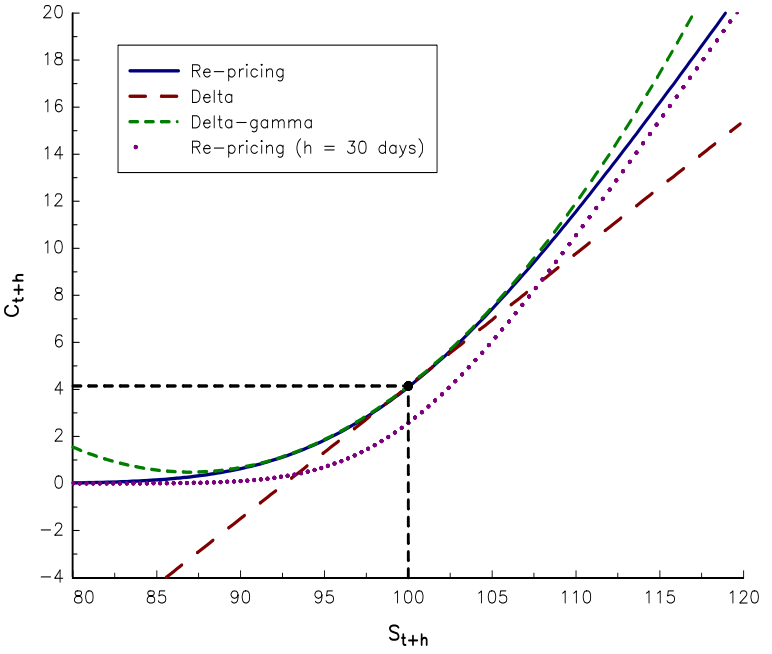


FIGURE 2.18: Approximation of the option price with the Greek coefficients

To overcome this drawback, we can use the second-order approximation or the delta-gamma approach:

$$\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t \simeq \boldsymbol{\Delta}_t\left(S_{t+h}-S_t\right) + \frac{1}{2}\boldsymbol{\Gamma}_t\left(S_{t+h}-S_t\right)^2$$

where $\boldsymbol{\Gamma}_t$ is the option gamma:

$$\boldsymbol{\Gamma}_t = \frac{\partial^2 C_{\mathrm{BS}}(S_t,\Sigma_t,T)}{\partial S_t^2}$$

[82] A long (or short) position on the underlying asset is equivalent to $\boldsymbol{\Delta}_t = 1$ (or $\boldsymbol{\Delta}_t = -1$).

In Figure 2.18, we compare the two Taylor expansions with the re-pricing method when h is equal to one trading day. We observe that the delta approach provides a bad approximation if the future price S_{t+h} is far from the current price S_t. The inclusion of the gamma helps to correct the pricing error. However, if the time period h is high, the two approximations may be inaccurate even in the neighborhood de S_t (see the case $h = 30$ days in Figure 2.18). It is therefore important to take into account the time or maturity effect:

$$\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t \simeq \boldsymbol{\Delta}_t \left(S_{t+h} - S_t\right) + \frac{1}{2}\boldsymbol{\Gamma}_t \left(S_{t+h} - S_t\right)^2 + \boldsymbol{\Theta}_t h$$

where $\boldsymbol{\Theta}_t = \partial_t C_{\text{BS}}\left(S_t, \Sigma_t, T\right)$ is the option theta[83].

The Taylor expansion can be generalized to a set of risk factors $\mathcal{F}_t = (\mathcal{F}_{1,t}, \ldots, \mathcal{F}_{m,t})$:

$$\begin{aligned}\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t \simeq & \sum_{j=1}^{m} \frac{\partial \boldsymbol{\mathcal{C}}_t}{\partial \mathcal{F}_{j,t}} \left(\mathcal{F}_{j,t+h} - \mathcal{F}_{j,t}\right) + \\ & \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} \frac{\partial^2 \boldsymbol{\mathcal{C}}_t}{\partial \mathcal{F}_{j,t}\, \partial \mathcal{F}_{k,t}} \left(\mathcal{F}_{j,t+h} - \mathcal{F}_{j,t}\right)\left(\mathcal{F}_{k,t+h} - \mathcal{F}_{k,t}\right)\end{aligned}$$

The delta-gamma-theta approach consists in considering the underlying price and the maturity as risk factors. If we add the implied volatility as a new risk factor, we obtain:

$$\begin{aligned}\boldsymbol{\mathcal{C}}_{t+h} - \boldsymbol{\mathcal{C}}_t \simeq & \boldsymbol{\Delta}_t \left(S_{t+h} - S_t\right) + \frac{1}{2}\boldsymbol{\Gamma}_t \left(S_{t+h} - S_t\right)^2 + \boldsymbol{\Theta}_t h + \\ & \boldsymbol{\upsilon}_t \left(\Sigma_{t+h} - \Sigma_t\right)\end{aligned}$$

where $\boldsymbol{\upsilon}_t = \partial_{\Sigma_t} C_{\text{BS}}\left(S_t, \Sigma_t, T\right)$ is the option vega. Here, we have considered that only the second derivative of $\boldsymbol{\mathcal{C}}_t$ with respect to S_t is significant, but we could also include the vanna or volga effect[84].

In the case of the call option, the Black-Scholes sensitivities are equal to:

$$\begin{aligned}\boldsymbol{\Delta}_t &= e^{(b_t - r_t)\tau}\Phi\left(d_1\right) \\ \boldsymbol{\Gamma}_t &= \frac{e^{(b_t - r_t)\tau}\phi\left(d_1\right)}{S_t \Sigma_t \sqrt{\tau}} \\ \boldsymbol{\Theta}_t &= -r_t K e^{-r_t \tau}\Phi\left(d_2\right) - \frac{1}{2\sqrt{\tau}} S_t \Sigma_t e^{(b_t - r_t)\tau}\phi\left(d_1\right) - \\ & \quad \left(b_t - r_t\right) S_t e^{(b_t - r_t)\tau}\Phi\left(d_1\right) \\ \boldsymbol{\upsilon}_t &= e^{(b_t - r_t)\tau} S_t \sqrt{\tau}\phi\left(d_1\right)\end{aligned}$$

If we consider again Example 18 on page 95, we obtain[85] $\boldsymbol{\Delta}_t = 0.5632$, $\boldsymbol{\Gamma}_t = 0.0434$, $\boldsymbol{\Theta}_t = -11.2808$ and $\boldsymbol{\upsilon}_t = 17.8946$. In Table 2.14, we have reported the approximated P&Ls for the first nine scenarios and the one-factor model. The fourth column indicates the P&L obtained by the full pricing method, which were already reported in Table 2.12. $\Pi_s^{\boldsymbol{\Delta}}(w)$, $\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}}(w)$ and $\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}+\boldsymbol{\Theta}}(w)$ correspond respectively to delta, delta-gamma, delta-gamma-theta approaches. For example, we have $\Pi_1^{\boldsymbol{\Delta}}(w) = 100 \times 0.5632 \times (98.07 - 100) = -108.69$, $\Pi_1^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}}(w) = -108.69 + 100 \times \frac{1}{2} \times 0.0434 \times (98.07 - 100)^2 = -100.61$ and $\Pi_1^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}+\boldsymbol{\Theta}}(w) =$

[83] An equivalent formula is $\boldsymbol{\Theta}_t = -\partial_T C_{\text{BS}}\left(S_t, \Sigma_t, T\right) = -\partial_\tau C_{\text{BS}}\left(S_t, \Sigma_t, T\right)$ because the maturity T (or the time to maturity τ) is moving in the opposite way with respect to the time t.

[84] The vanna coefficient corresponds to the cross-derivative of $\boldsymbol{\mathcal{C}}_t$ with respect to S_t and Σ_t whereas the volga effect is the second derivative of $\boldsymbol{\mathcal{C}}_t$ with respect to Σ_t.

[85] We have $d_1 = 0.1590$, $\Phi\left(d_1\right) = 0.5632$, $\phi\left(d_1\right) = 0.3939$, $d_2 = 0.0681$ and $\Phi\left(d_2\right) = 0.5272$.

$-100.61 - 11.2808 \times 1/252 = -105.09$. We notice that we obtain a good approximation with the delta, but it is more accurate to combine delta, gamma and theta sensibilities. Finally, the 99% VaRs for a one-day holding period are \$171.20 and \$151.16 and \$155.64. This is the delta-gamma-theta approach which gives the closest result[86]. If the set of risk factors includes the implied volatility, we obtain the results in Table 2.15. We notice that the vega effect is very significant (fifth column). As an illustration, we have $\Pi_1^{\boldsymbol{\upsilon}}(w) = 100 \times 17.8946 \times (15.58\% - 20\%) = -79.09$, implying that the volatility risk explains 43.4% of the loss of \$182.25 for the first scenario. Finally, the VaR is equal to \$183.76 with the delta-gamma-theta-vega approach whereas we found previously that it was equal to \$181.70 with the full pricing method.

TABLE 2.14: Calculation of the P&L based on the Greek sensitivities

s	R_s (in %)	S_{t+h}	Π_s	$\Pi_s^{\boldsymbol{\Delta}}$	$\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}}$	$\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}+\boldsymbol{\Theta}}$
1	−1.93	98.07	−104.69	−108.69	−100.61	−105.09
2	−0.69	99.31	−42.16	−38.86	−37.83	42.30
3	−0.71	99.29	−43.22	−39.98	−38.89	−43.37
4	−0.73	99.27	−44.28	−41.11	−39.96	−44.43
5	1.22	101.22	67.46	68.71	71.93	67.46
6	1.01	101.01	54.64	56.88	59.09	54.61
7	1.04	101.04	56.46	58.57	60.91	56.44
8	1.08	101.08	58.89	60.82	63.35	58.87
9	−1.61	98.39	−89.22	−90.67	−85.05	−89.53
$\text{VaR}_{99\%}(w;\text{one day})$			154.79	171.20	151.16	155.64
$\text{ES}_{97.5\%}(w;\text{one day})$			150.04	165.10	146.37	150.84

TABLE 2.15: Calculation of the P&L using the vega coefficient

s	S_{t+h}	Σ_{t+h}	Π_s	$\Pi_s^{\boldsymbol{\upsilon}}$	$\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\upsilon}}$	$\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}+\boldsymbol{\upsilon}}$	$\Pi_s^{\boldsymbol{\Delta}+\boldsymbol{\Gamma}+\boldsymbol{\Theta}+\boldsymbol{\upsilon}}$
1	98.07	15.58	−182.25	−79.09	−187.78	−179.71	−184.19
2	99.31	18.68	−65.61	−23.62	−62.48	−61.45	−65.92
3	99.29	16.96	−97.23	−54.40	−94.38	−93.29	−97.77
4	99.27	22.88	6.87	51.54	10.43	11.58	7.10
5	101.22	19.87	65.20	−2.33	66.38	69.61	65.13
6	101.01	19.92	53.24	−1.43	55.45	57.66	53.18
7	101.04	21.29	79.03	23.08	81.65	84.00	79.52
8	101.08	22.93	110.21	52.43	113.25	115.78	111.30
9	98.39	20.85	−74.21	15.21	−75.46	−69.84	−74.32
$\text{VaR}_{99\%}(w;\text{one day})$			181.70	77.57	190.77	179.29	183.76
$\text{ES}_{97.5\%}(w;\text{one day})$			172.09	73.90	184.90	169.34	173.81

Remark 14 *We do not present here the non-linear quadratic VaR, which consists in computing the VaR of option portfolios with the Cornish-Fisher expansion (Zangari, 1996; Britten-Jones and Schaefer, 1999). It is called 'quadratic' because it uses the delta-gamma approximation and requires calculating the moments of the quadratic form* $(S_{t+h} - S_t)^2$. *The treatment of this approach is left as Exercise 2.4.8 on page 123.*

[86] We found previously that the VaR was equal to \$154.79 with the full pricing method.

The hybrid method On the one hand, the full pricing method has the advantage to be accurate, but also the drawback to be time-consuming because it performs a complete revaluation of the portfolio for each scenario. On the other hand, the method based on the sensitivities is less accurate, but also faster than the re-pricing approach. Indeed, the Greek coefficients are calculated once and for all, and their values do not depend on the scenario. The hybrid method consists of combining the two approaches:

1. we first calculate the P&L for each (historical or simulated) scenario with the method based on the sensitivities;
2. we then identify the worst scenarios;
3. we finally revalue these worst scenarios by using the full pricing method.

The underlying idea is to consider the faster approach to locate the value-at-risk, and then to use the most accurate approach to calculate the right value.

TABLE 2.16: The 10 worst scenarios identified by the hybrid method

i	Full pricing		Greeks					
			$\boldsymbol{\Delta}-\boldsymbol{\Gamma}-\boldsymbol{\Theta}-\boldsymbol{\upsilon}$		$\boldsymbol{\Delta}-\boldsymbol{\Theta}$		$\boldsymbol{\Delta}-\boldsymbol{\Theta}-\boldsymbol{\upsilon}$	
	s	Π_s	s	Π_s	s	Π_s	s	Π_s
1	100	−183.86	100	−186.15	182	−187.50	134	−202.08
2	1	−182.25	1	−184.19	169	−176.80	100	−198.22
3	134	−181.15	134	−183.34	27	−174.55	1	−192.26
4	27	−163.01	27	−164.26	134	−170.05	169	−184.32
5	169	−162.82	169	−164.02	69	−157.66	27	−184.04
6	194	−159.46	194	−160.93	108	−150.90	194	−175.36
7	49	−150.25	49	−151.43	194	−149.77	49	−165.41
8	245	−145.43	245	−146.57	49	−147.52	182	−164.96
9	182	−142.21	182	−142.06	186	−145.27	245	−153.37
10	79	−135.55	79	−136.52	100	−137.38	69	−150.68

In Table 2.16, we consider the previous example with the implied volatility as a risk factor. We have reported the worst scenarios corresponding to the order statistic $i : n_S$ with $i \leq 10$. In the case of the full pricing method, the five worst scenarios are the 100^{th}, 1^{st}, 134^{th}, 27^{th} and 169^{th}. This implies that the hybrid method will give the right result if it is able to select the 100^{th}, 1^{st} and 134^{th} scenarios to compute the value-at-risk which corresponds to the average of the second and third order statistics. If we consider the $\boldsymbol{\Delta}-\boldsymbol{\Gamma}-\boldsymbol{\Theta}-\boldsymbol{\upsilon}$ approximation, we identify the same ten worst scenarios. It is perfectly normal, as it is easy to price an European call option. It will not be the case with exotic options, because the approximation may not be accurate. For instance, if we consider our example with the $\boldsymbol{\Delta}-\boldsymbol{\Theta}$ approximation, the five worst scenarios becomes the 182^{th}, 169^{st}, 27^{th}, 134^{th} and 69^{th}. If we revaluate these 5 worst scenarios, the 99% value-at-risk is equal to:

$$\text{VaR}_{99\%}(w; \text{one day}) = \frac{1}{2}(163.01 + 162.82) = \$162.92$$

which is a result far from the value of \$180.70 found with the full pricing method. With the 10 worst scenarios, we obtain:

$$\text{VaR}_{99\%}(w; \text{one day}) = \frac{1}{2}(181.15 + 163.01) = \$172.08$$

Once again, we do not find the exact value, because the $\boldsymbol{\Delta}-\boldsymbol{\Theta}$ approximation fails to detect the first scenario among the 10 worst scenarios. This problem vanishes with the $\boldsymbol{\Delta}-\boldsymbol{\Theta}-\boldsymbol{\upsilon}$ approximation, even if it gives a ranking different than this obtained with the full pricing method. In practice, the hybrid approach is widespread and professionals generally use the identification method with 10 worst scenarios[87].

2.2.5.3 Backtesting

When we consider a model to price a product, the valuation is known as '*mark-to-model*' and requires more attention than the mark-to-market approach. In this last case, the simulated P&L is the difference between the mark-to-model value at time $t+1$ and the current mark-to-market value:

$$\Pi_s(w) = \underbrace{P_{t+1}(w)}_{\text{mark-to-model}} - \underbrace{P_t(w)}_{\text{mark-to-market}}$$

At time $t+1$, the realized P&L is the difference between two mark-to-market values:

$$\Pi(w) = \underbrace{P_{t+1}(w)}_{\text{mark-to-market}} - \underbrace{P_t(w)}_{\text{mark-to-market}}$$

For exotic options and OTC derivatives, we don't have market prices and the portfolio is valuated using the mark-to-model approach. This means that the simulated P&L is the difference between two mark-to-model values:

$$\Pi_s(w) = \underbrace{P_{t+1}(w)}_{\text{mark-to-model}} - \underbrace{P_t(w)}_{\text{mark-to-model}}$$

and the realized P&L is also the difference between two mark-to-model values:

$$\Pi(w) = \underbrace{P_{t+1}(w)}_{\text{mark-to-model}} - \underbrace{P_t(w)}_{\text{mark-to-model}}$$

In the case of the mark-to-model valuation, we see the relevance of the pricing model in terms of risk management. Indeed, if the pricing model is wrong, the value-at-risk is wrong too and this cannot be detected by the backtesting procedure, which has little signification. This is why the supervisory authority places great importance on model risk.

2.2.5.4 Model risk

Model risk cannot be summarized in a unique definition due to its complexity. For instance, Derman (1996, 2001) considers six types of model risk (inapplicability of modeling, incorrect model, incorrect solutions, badly approximated solution, bugs and unstable data). Rebonato (2001) defines model risk as "*the risk of a significant difference between the mark-to-model value of an instrument, and the price at which the same instrument is revealed to have traded in the market*". According to Morini (2001), these two approaches are different. For Riccardo Rebonato, there is not a true value of an instrument before it will be traded on the market. Model risk can therefore be measured by selling the instrument in the market. For Emanuel Derman, an instrument has an intrinsic true value, but it is unknown. The proposition of Rebonato is certainly the right way to define model risk, but it does not help to measure model risk from an ex-ante point of view. Moreover, this approach does

[87] Its application is less frequent than in the past because computational times have dramatically decreased with the evolution of technology, in particular the development of parallel computing.

not distinguish between model risk and liquidity risk. The conception of Derman is more adapted to manage model risk and calibrate the associated provisions. This is the approach that has been adopted by banks and regulators. Nevertheless, the multifaceted nature of this approach induces very different implementations across banks, because it appears as a catalogue with an infinite number of rules.

We consider a classification with four main types of model risk:

1. the operational risk;
2. the parameter risk;
3. the risk of mis-specification;
4. the hedging risk.

The operational risk is the risk associated to the implementation of the pricer. It concerns programming mistakes or bugs, but also mathematical errors in closed-form formulas, approximations or numerical methods. A typical example is the use of a numerical scheme for solving a partial differential equation. The accuracy of the option price and the Greek coefficients will depend on the specification of the numerical algorithm (explicit, implicit or mixed scheme) and the discretization parameters (time and space steps). Another example is the choice of the Monte Carlo method and the number of simulations.

The parameter risk is the risk associated to the input parameters, in particular those which are difficult to estimate. A wrong value of one parameter can lead to a mis-pricing, even though the model is right and well implemented. In this context, the question of available and reliable data is a key issue. It is particularly true when the parameters are unobservable and are based on an expert's opinion. A typical example concerns the value of correlations in multi-asset options. Even if there is no problem with data, some parameters are indirectly related to market data via a calibration set. In this case, they may change with the specification of the calibration set. For instance, the pricing of exotic interest rate options is generally based on parameters calibrated from prices of plain vanilla instruments (caplets and swaptions). The analysis of parameter risk consists then of measuring the impact of parameter changes on the price and the hedging portfolio of the exotic option.

The risk of mis-specification is the risk associated to the mathematical model, because it may not include all risk factors, the dynamics of the risk factors is not adequate or the dependence between them is not well defined. It is generally easy to highlight this risk, because various models calibrated with the same set of instruments can produce different prices for the same exotic option. The big issue is to define what is the least bad model. For example, in the case of equity options, we have the choice between many models: Black-Scholes, local volatility, Heston model, other stochastic volatility models, jump-diffusion, etc. In practice, the frontier between the risk of parameters and the risk of mis-specification may be unclear as shown by the seminal work of uncertainty on pricing and hedging by Avellaneda *et al.* (1995). Moreover, a model which appears to be good for pricing may not be well adapted for risk management. This explains that the trader and the risk manager can use sometimes two different models for the same option payoff.

The hedging risk is the risk associated to the trading management of the option portfolio. The sales margin corresponds to the difference between the transaction price and the mark-to-model price. The sales margin is calculated at the inception date of the transaction. To freeze the margin, we have to hedge the option. The mark-to-model value is then transferred to the option trader and represents the hedging cost. We face here the risk that the realized hedging cost will be larger than the mark-to-model price. A typical example is a put option, which has a negative delta. The hedging portfolio corresponds then to a short selling on

the underlying asset. Sometimes, this short position may be difficult to implement (e.g. a ban on short selling) or may be very costly (e.g. due to a change in the bank funding condition). Some events may also generate a rebalancing risk. The most famous example is certainly the hedge fund crisis in October 2008, which has imposed redemption restrictions or gates. This caused difficulties to traders, who managed call options on hedge funds and were unable to reduce their deltas at this time. The hedging risk does not only concern the feasibility of the hedging implementation, but also its adequacy with the model. As an illustration, we suppose that we use a stochastic volatility model for an option, which is sensitive to the vanna coefficient. The risk manager can then decide to use this model for measuring the value-at-risk, but the trader can also prefer to implement a Black-Scholes hedging portfolio[88]. This is not a problem that the risk manager uses a different model than the trader if the model risk only includes the first three categories. However, it will be a problem if it also concerns hedging risk.

In the Basel III framework, the Basel Committee highlights the role of the model validation team:

> "*A distinct unit of the bank that is separate from the unit that designs and implements the internal models must conduct the initial and ongoing validation of all internal models used to determine market risk capital requirements. The model validation unit must validate all internal models used for purposes of the IMA on at least an annual basis. [...] Banks must maintain a process to ensure that their internal models have been adequately validated by suitably qualified parties independent of the model development process to ensure that each model is conceptually sound and adequately reflects all material risks. Model validation must be conducted both when the model is initially developed and when any significant changes are made to the model*" (BCBS, 2019, pages 68-69).

Therefore, model risk justifies that model validation is an integral part of the risk management process for exotic options. The tasks of a model validation team are multiple and concern reviewing the programming code, checking mathematical formulas and numerical approximations, validating market data, testing the calibration stability, challenging the pricer with alternative models, proposing provision buffers, etc. This team generally operates at the earliest stages of the pricer development (or when the pricer changes), whereas the risk manager is involved to follow the product on a daily basis. In Chapter 9, we present the different tools available for the model validation unit in order to assess the robustness of risk measures that are based on mark-to-model prices.

Remark 15 *It is a mistake to think that model risk is an operational risk. Model risk is intrinsically a market risk. Indeed, it exists because exotic options are difficult to price and hedge, implying that commercial risk is high. This explains that sales margins are larger than for vanilla options and implicitly include model risk, which is therefore inherent to the business of exotic derivatives.*

2.3 Risk allocation

Measuring the risk of a portfolio is a first step to manage it. In particular, a risk measure is a single number that is not very helpful for understanding the sources of the portfolio risk.

[88] There may be many reasons for implementing more simple hedging portfolios: the trader may be more confident in the robustness, there is no market instrument to replicate the vanna position, etc.

To go further, we must define precisely the notion of risk contribution in order to propose risk allocation principles.

Let us consider two trading desks A and B, whose risk measure is respectively $\mathcal{R}(w_A)$ and $\mathcal{R}(w_B)$. At the global level, the risk measure is equal to $\mathcal{R}(w_{A+B})$. The question is then how to allocate $\mathcal{R}(w_{A+B})$ to the trading desks A and B:

$$\mathcal{R}(w_{A+B}) = \mathcal{RC}_A(w_{A+B}) + \mathcal{RC}_B(w_{A+B})$$

There is no reason that $\mathcal{RC}_A(w_{A+B}) = \mathcal{R}(w_A)$ and $\mathcal{RC}_B(w_{A+B}) = \mathcal{R}(w_B)$ except if there is no diversification. This question is an important issue for the bank because risk allocation means capital allocation:

$$\mathcal{K}(w_{A+B}) = \mathcal{K}_A(w_{A+B}) + \mathcal{K}_B(w_{A+B})$$

Capital allocation is not neutral, because it will impact the profitability of business units that compose the bank.

Remark 16 *This section is based on Chapter 2 of the book of Roncalli (2013).*

2.3.1 Euler allocation principle

According to Litterman (1996), risk allocation consists in decomposing the risk portfolio into a sum of risk contributions by sub-portfolios (assets, trading desks, etc.). The concept of risk contribution is key in identifying concentrations and understanding the risk profile of the portfolio, and there are different methods for defining them. As illustrated by Denault (2001), some methods are more pertinent than others and the Euler principle is certainly the most used and accepted one.

We decompose the P&L as follows:

$$\Pi = \sum_{i=1}^{n} \Pi_i$$

where Π_i is the P&L of the i^{th} sub-portfolio. We note $\mathcal{R}(\Pi)$ the risk measure associated with the P&L[89]. Let us consider the risk-adjusted performance measure (RAPM) defined by[90]:

$$\text{RAPM}(\Pi) = \frac{\mathbb{E}[\Pi]}{\mathcal{R}(\Pi)}$$

Tasche (2008) considers the portfolio-related RAPM of the i^{th} sub-portfolio defined by:

$$\text{RAPM}(\Pi_i \mid \Pi) = \frac{\mathbb{E}[\Pi_i]}{\mathcal{R}(\Pi_i \mid \Pi)}$$

Based on the notion of RAPM, Tasche (2008) states two properties of risk contributions that are desirable from an economic point of view:

1. Risk contributions $\mathcal{R}(\Pi_i \mid \Pi)$ to portfolio-wide risk $\mathcal{R}(\Pi)$ satisfy the full allocation property if:

$$\sum_{i=1}^{n} \mathcal{R}(\Pi_i \mid \Pi) = \mathcal{R}(\Pi) \tag{2.11}$$

[89] We recall that $\mathcal{R}(\Pi) = \mathcal{R}(-L)$.

[90] This concept is close to the RAROC measure introduced by Banker Trust (see page 2).

2. Risk contributions $\mathcal{R}(\Pi_i \mid \Pi)$ are RAPM compatible if there are some $\varepsilon_i > 0$ such that[91]:

$$\text{RAPM}(\Pi_i \mid \Pi) > \text{RAPM}(\Pi) \Rightarrow \text{RAPM}(\Pi + h\Pi_i) > \text{RAPM}(\Pi) \tag{2.12}$$

for all $0 < h < \varepsilon_i$.

Tasche (2008) shows therefore that if there are risk contributions that are RAPM compatible in the sense of the two previous properties (2.11) and (2.12), then $\mathcal{R}(\Pi_i \mid \Pi)$ is uniquely determined as:

$$\mathcal{R}(\Pi_i \mid \Pi) = \frac{\mathrm{d}}{\mathrm{d}h}\mathcal{R}(\Pi + h\Pi_i)\Big|_{h=0} \tag{2.13}$$

and the risk measure is homogeneous of degree 1. In the case of a subadditive risk measure, one can also show that:

$$\mathcal{R}(\Pi_i \mid \Pi) \leq \mathcal{R}(\Pi_i) \tag{2.14}$$

This means that the risk contribution of the sub-portfolio i is always smaller than its stand-alone risk measure. The difference is related to the risk diversification.

Let us return to risk measure $\mathcal{R}(w)$ defined in terms of weights. The previous framework implies that the risk contribution of sub-portfolio i is uniquely defined as:

$$\mathcal{RC}_i = w_i \frac{\partial \mathcal{R}(w)}{\partial w_i} \tag{2.15}$$

and the risk measure satisfies the Euler decomposition:

$$\mathcal{R}(w) = \sum_{i=1}^{n} w_i \frac{\partial \mathcal{R}(w)}{\partial w_i} = \sum_{i=1}^{n} \mathcal{RC}_i \tag{2.16}$$

This relationship is also called the Euler allocation principle.

Remark 17 *We can always define the risk contributions of a risk measure by using Equation (2.15). However, this does not mean that the risk measure satisfies the Euler decomposition (2.16).*

Remark 18 *Kalkbrener (2005) develops an axiomatic approach to risk contribution. In particular, he shows that the Euler allocation principle is the only risk allocation method compatible with diversification principle (2.14) if the risk measure is subadditive.*

If we assume that the portfolio return $R(w)$ is a linear function of the weights w, the expression of the standard deviation-based risk measure becomes:

$$\begin{aligned} \mathcal{R}(w) &= -\mu(w) + c \cdot \sigma(w) \\ &= -w^\top \mu + c \cdot \sqrt{w^\top \Sigma w} \end{aligned}$$

where μ and Σ are the mean vector and the covariance matrix of sub-portfolios. It follows that the vector of marginal risks is:

$$\begin{aligned} \frac{\partial \mathcal{R}(w)}{\partial w} &= -\mu + c \cdot \frac{1}{2}\left(w^\top \Sigma w\right)^{-1}(2\Sigma w) \\ &= -\mu + c \cdot \frac{\Sigma w}{\sqrt{w^\top \Sigma w}} \end{aligned}$$

[91] This property means that assets with a better risk-adjusted performance than the portfolio continue to have a better RAPM if their allocation increases in a small proportion.

The risk contribution of the i^{th} sub-portfolio is then:

$$\mathcal{RC}_i = w_i \cdot \left(-\mu_i + c \cdot \frac{(\Sigma w)_i}{\sqrt{w^\top \Sigma w}} \right)$$

We verify that the standard deviation-based risk measure satisfies the full allocation property:

$$\begin{aligned}
\sum_{i=1}^n \mathcal{RC}_i &= \sum_{i=1}^n w_i \cdot \left(-\mu_i + c \cdot \frac{(\Sigma w)_i}{\sqrt{w^\top \Sigma w}} \right) \\
&= w^\top \left(-\mu + c \cdot \frac{\Sigma w}{\sqrt{w^\top \Sigma w}} \right) \\
&= -w^\top \mu + c \cdot \sqrt{w^\top \Sigma w} \\
&= \mathcal{R}(w)
\end{aligned}$$

Because Gaussian value-at-risk and expected shortfall are two special cases of the standard deviation-based risk measure, we conclude that they also satisfy the Euler allocation principle. In the case of the value-at-risk, the risk contribution becomes:

$$\mathcal{RC}_i = w_i \cdot \left(-\mu_i + \Phi^{-1}(\alpha) \cdot \frac{(\Sigma w)_i}{\sqrt{w^\top \Sigma w}} \right) \tag{2.17}$$

whereas in the case of the expected shortfall, it is equal to:

$$\mathcal{RC}_i = w_i \cdot \left(-\mu_i + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)} \cdot \frac{(\Sigma w)_i}{\sqrt{w^\top \Sigma w}} \right) \tag{2.18}$$

Remark 19 *Even if the risk measure is convex, it does not necessarily satisfy the Euler allocation principle. The most famous example is the variance of the portfolio return. We have* $\operatorname{var}(w) = w^\top \Sigma w$ *and* $\partial_w \operatorname{var}(w) = 2\Sigma w$*. It follows that* $\sum_{i=1}^n w_i \cdot \partial_{w_i} \operatorname{var}(w) = \sum_{i=1}^n w_i \cdot (2\Sigma w)_i = 2w^\top \Sigma w = 2\operatorname{var}(w) > \operatorname{var}(w)$*. In the case of the variance, the sum of the risk contributions is then always larger than the risk measure itself, because the variance does not satisfy the homogeneity property.*

Example 19 *We consider the Apple/Coca-Cola portfolio that has been used for calculating the Gaussian VaR on page 68. We recall that the nominal exposures were* \$1 093.3 *(Apple) and* \$842.8 *(Coca-Cola), the estimated standard deviation of daily returns was equal to* 1.3611% *for Apple and* 0.9468% *for Coca-Cola and the cross-correlation of stock returns was equal to* 12.0787%*.*

In the two-asset case, the expression of the value-at-risk or the expected shortfall is:

$$\mathcal{R}(w) = -w_1\mu_1 - w_2\mu_2 + c\sqrt{w_1^2\sigma_1^2 + 2w_1w_2\rho\sigma_1\sigma_2 + w_2^2\sigma_2^2}$$

It follows that the marginal risk of the first asset is:

$$\mathcal{MR}_1 = -\mu_1 + c\frac{w_1\sigma_1^2 + w_2\rho\sigma_1\sigma_2}{\sqrt{w_1^2\sigma_1^2 + 2w_1w_2\rho\sigma_1\sigma_2 + w_2^2\sigma_2^2}}$$

We then deduce that the risk contribution of the first asset is:

$$\mathcal{RC}_1 = -w_1\mu_1 + c\frac{w_1^2\sigma_1^2 + w_1w_2\rho\sigma_1\sigma_2}{\sqrt{w_1^2\sigma_1^2 + 2w_1w_2\rho\sigma_1\sigma_2 + w_2^2\sigma_2^2}}$$

By using the numerical values[92] of Example 19, we obtain the results given in Tables 2.17 and 2.18. We verify that the sum of risk contributions is equal to the risk measure. We notice that the stock Apple explains 75.14% of the risk whereas it represents 56.47% of the allocation.

TABLE 2.17: Risk decomposition of the 99% Gaussian value-at-risk

Asset	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
Apple	1093.3	2.83%	30.96	75.14%
Coca-Cola	842.8	1.22%	10.25	24.86%
$\mathcal{R}(w)$			41.21	

TABLE 2.18: Risk decomposition of the 99% Gaussian expected shortfall

Asset	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
Apple	1093.3	3.24%	35.47	75.14%
Coca-Cola	842.8	1.39%	11.74	24.86%
$\mathcal{R}(w)$			47.21	

2.3.2 Application to non-normal risk measures

2.3.2.1 Main results

In the previous section, we provided formulas for when asset returns are normally distributed. However, the previous expressions can be extended in the general case. For the value-at-risk, Gouriéroux *et al.* (2000) show that the risk contribution is equal to[93]:

$$\begin{aligned}\mathcal{RC}_i &= \mathcal{R}(\Pi_i \mid \Pi) \\ &= -\mathbb{E}[\Pi_i \mid \Pi = -\operatorname{VaR}_\alpha(\Pi)] \\ &= \mathbb{E}[L_i \mid L(w) = \operatorname{VaR}_\alpha(L)] \end{aligned} \tag{2.19}$$

Formula (2.19) is more general than Equation (2.17) obtained in the Gaussian case. Indeed, we can retrieve the latter if we assume that the returns are Gaussian. We recall that the portfolio return is $R(w) = \sum_{i=1}^n w_i R_i = w^\top R$. The portfolio loss is defined by $L(w) = -R(w)$. We deduce that:

$$\begin{aligned}\mathcal{RC}_i &= \mathbb{E}[-w_i R_i \mid -R(w) = \operatorname{VaR}_\alpha(w;h)] \\ &= -w_i \mathbb{E}[R_i \mid R(w) = -\operatorname{VaR}_{\alpha;h}(w)]\end{aligned}$$

Because $R(w)$ is a linear combination of R, the random vector $(R, R(w))$ is Gaussian and we have:

$$\begin{pmatrix} R \\ R(w) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ w^\top \mu \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma w \\ w^\top \Sigma & w^\top \Sigma w \end{pmatrix}\right)$$

[92] We set $\mu_1 = \mu_2 = 0$.
[93] See also Hallerbach (2003).

We know that $\text{VaR}_\alpha(w;h) = -w^\top\mu + \Phi^{-1}(\alpha)\sqrt{w^\top\Sigma w}$. It follows that[94]:

$$\begin{aligned}\mathbb{E}[R|R(w) = -\text{VaR}_\alpha(w;h)] &= \mathbb{E}\left[R \mid R(w) = w^\top\mu - \Phi^{-1}(\alpha)\sqrt{w^\top\Sigma w}\right] \\ &= \mu + \Sigma w\left(w^\top\Sigma w\right)^{-1} \cdot \\ &\quad \left(w^\top\mu - \Phi^{-1}(\alpha)\sqrt{w^\top\Sigma w} - w^\top\mu\right)\end{aligned}$$

and:

$$\begin{aligned}\mathbb{E}[R|R(w) = -\text{VaR}_\alpha(w;h)] &= \mu - \Phi^{-1}(\alpha)\Sigma w\frac{\sqrt{w^\top\Sigma w}}{(w^\top\Sigma w)^{-1}} \\ &= \mu - \Phi^{-1}(\alpha)\frac{\Sigma w}{\sqrt{w^\top\Sigma w}}\end{aligned}$$

We finally obtain the same expression as Equation (2.17):

$$\begin{aligned}\mathcal{RC}_i &= -w_i\left(\mu - \Phi^{-1}(\alpha)\frac{\Sigma w}{\sqrt{w^\top\Sigma w}}\right)_i \\ &= -w_i\mu_i + \Phi^{-1}(\alpha)\frac{w_i\cdot(\Sigma w)_i}{\sqrt{w^\top\Sigma w}}\end{aligned}$$

In the same way, Tasche (2002) shows that the general expression of the risk contributions for the expected shortfall is:

$$\begin{aligned}\mathcal{RC}_i &= \mathcal{R}(\Pi_i \mid \Pi) \\ &= -\mathbb{E}[\Pi_i \mid \Pi \leq -\text{VaR}_\alpha(\Pi)] \\ &= \mathbb{E}[L_i \mid L(w) \geq \text{VaR}_\alpha(L)]\end{aligned} \tag{2.20}$$

Using Bayes' theorem, it follows that:

$$\mathcal{RC}_i = \frac{\mathbb{E}[L_i\cdot\mathbb{1}\{L(w)\geq\text{VaR}_\alpha(L)\}]}{1-\alpha}$$

If we apply the previous formula to the Gaussian case, we obtain:

$$\mathcal{RC}_i = -\frac{w_i}{1-\alpha}\mathbb{E}[R_i\cdot\mathbb{1}\{R(w)\leq-\text{VaR}_\alpha(L)\}]$$

After some tedious computations, we retrieve the same expression as found previously[95].

2.3.2.2 Calculating risk contributions with historical and simulated scenarios

The case of value-at-risk When using historical or simulated scenarios, the VaR is calculated as follows:

$$\text{VaR}_\alpha(w;h) = -\Pi_{((1-\alpha)n_S:n_S)} = L_{(\alpha n_S:n_S)}$$

Let $\mathfrak{R}_\Pi(s)$ be the rank of the P&L associated to the s^{th} observation meaning that:

$$\mathfrak{R}_\Pi(s) = \sum_{j=1}^{n_S}\mathbb{1}\{\Pi_j\leq\Pi_s\}$$

[94] We use the formula of the conditional expectation presented in Appendix A.2.2.4 on page 1062.
[95] The derivation of the formula is left as an exercise (Section 2.4.9 on page 123).

We deduce that:

$$\Pi_s = \Pi_{(\mathfrak{R}_\Pi(s):n_S)}$$

Formula (2.19) is then equivalent to decompose $\Pi_{((1-\alpha)n_S:n_S)}$ into individual P&Ls. We have $\Pi_s = \sum_{i=1}^n \Pi_{i,s}$ where $\Pi_{i,s}$ is the P&L of the i^{th} sub-portfolio for the s^{th} scenario. It follows that:

$$\begin{aligned}
\text{VaR}_\alpha(w;h) &= -\Pi_{((1-\alpha)n_S:n_S)} \\
&= -\Pi_{\mathfrak{R}_\Pi^{-1}((1-\alpha)n_S)} \\
&= -\sum_{i=1}^n \Pi_{i,\mathfrak{R}_\Pi^{-1}((1-\alpha)n_S)}
\end{aligned}$$

where $\mathfrak{R}_\Pi^{-1}$ is the inverse function of the rank. We finally deduce that:

$$\begin{aligned}
\mathcal{RC}_i &= -\Pi_{i,\mathfrak{R}_\Pi^{-1}((1-\alpha)n_S)} \\
&= L_{i,\mathfrak{R}_\Pi^{-1}((1-\alpha)n_S)}
\end{aligned}$$

The risk contribution of the i^{th} sub-portfolio is the loss of the i^{th} sub-portfolio corresponding to the scenario $\mathfrak{R}_\Pi^{-1}((1-\alpha)n_S)$. If $(1-\alpha)n_S$ is not an integer, we have:

$$\mathcal{RC}_i = -\left(\Pi_{i,\mathfrak{R}_\Pi^{-1}(q)} + ((1-\alpha)n_S - q)\left(\Pi_{i,\mathfrak{R}_\Pi^{-1}(q+1)} - \Pi_{i,\mathfrak{R}_\Pi^{-1}(q)}\right)\right)$$

where $q = q_\alpha(n_S)$ is the integer part of $(1-\alpha)n_S$.

Let us consider Example 13 on page 68. We have found that the historical value-at-risk is \$47.39. It corresponds to the linear interpolation between the second and third largest loss. Using results in Table 2.7 on page 70, we notice that $\mathfrak{R}_\Pi^{-1}(1) = 236$, $\mathfrak{R}_\Pi^{-1}(2) = 69$, $\mathfrak{R}_\Pi^{-1}(3) = 85$, $\mathfrak{R}_\Pi^{-1}(4) = 23$ and $\mathfrak{R}_\Pi^{-1}(5) = 242$. We deduce that the second and third order statistics correspond to the 69^{th} and 85^{th} historical scenarios. The risk decomposition is reported in Table 2.19. Therefore, we calculate the risk contribution of the Apple stock as follows:

$$\begin{aligned}
\mathcal{RC}_1 &= -\frac{1}{2}(\Pi_{1,69} + \Pi_{1,85}) \\
&= -\frac{1}{2}(10 \times (105.16 - 109.33) + 10 \times (104.72 - 109.33)) \\
&= \$43.9
\end{aligned}$$

For the Coca-Cola stock, we obtain:

$$\begin{aligned}
\mathcal{RC}_2 &= -\frac{1}{2}(\Pi_{2,69} + \Pi_{2,85}) \\
&= -\frac{1}{2}(20 \times (41.65 - 42.14) + 20 \times (42.28 - 42.14)) \\
&= \$3.5
\end{aligned}$$

If we compare these results with those obtained with the Gaussian VaR, we observe that the risk decomposition is more concentrated for the historical VaR. Indeed, the exposure on Apple represents 96.68% whereas it was previously equal to 75.14%. The problem is that the estimator of the risk contribution only uses two observations, implying that its variance is very high.

TABLE 2.19: Risk decomposition of the 99% historical value-at-risk

Asset	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
Apple	56.47%	77.77	43.92	92.68%
Coca-Cola	43.53%	7.97	3.47	7.32%
$\mathcal{R}(w)$			47.39	

We can consider three techniques to improve the efficiency of the estimator $\mathcal{RC}_i = L_{i,\mathfrak{R}_\Pi^{-1}(n_S(1-a))}$. The first approach is to use a regularization method (Scaillet, 2004). The idea is to estimate the value-at-risk by weighting the order statistics:

$$\begin{aligned}\mathrm{VaR}_\alpha(w;h) &= -\sum_{s=1}^{n_S} \varpi_\alpha(s;n_S)\,\Pi_{(s:n_S)} \\ &= -\sum_{s=1}^{n_S} \varpi_\alpha(s;n_S)\,\Pi_{\mathfrak{R}_\Pi^{-1}(s)}\end{aligned}$$

where $\varpi_\alpha(s;n_S)$ is a weight function dependent on the confidence level α. The expression of the risk contribution then becomes:

$$\mathcal{RC}_i = -\sum_{s=1}^{n_S} \varpi_\alpha(s;n_S)\,\Pi_{i,\mathfrak{R}_\Pi^{-1}(s)}$$

Of course, this naive method can be improved by using more sophisticated approaches such as importance sampling (Glasserman, 2005).

In the second approach, asset returns are assumed to be elliptically distributed. In this case, Carroll *et al.* (2001) show that[96]:

$$\mathcal{RC}_i = \mathbb{E}[L_i] + \frac{\mathrm{cov}(L,L_i)}{\sigma^2(L)}\left(\mathrm{VaR}_\alpha(L) - \mathbb{E}[L]\right) \tag{2.21}$$

Estimating the risk contributions with historical scenarios is then straightforward. It suffices to apply Formula (2.21) by replacing the statistical moments by their sample statistics:

$$\mathcal{RC}_i = \bar{L}_i + \frac{\sum_{s=1}^{n_S}\left(L_s - \bar{L}\right)\left(L_{i,s} - \bar{L}_i\right)}{\sum_{s=1}^{n_S}\left(L_s - \bar{L}\right)^2}\left(\mathrm{VaR}_\alpha(L) - \bar{L}\right)$$

where $\bar{L}_i = n_S^{-1}\sum_{s=1}^{n_S} L_{i,s}$ and $\bar{L} = n_S^{-1}\sum_{s=1}^{n_S} L_s$. Equation (2.21) can be viewed as the estimation of the conditional expectation $\mathbb{E}[L_i|\,L = \mathrm{VaR}_\alpha(L)]$ in a linear regression framework:

$$L_i = \beta L + \varepsilon_i$$

[96] We verify that the sum of the risk contributions is equal to the value-at-risk:

$$\begin{aligned}\sum_{i=1}^n \mathcal{RC}_i &= \sum_{i=1}^n \mathbb{E}[L_i] + \left(\mathrm{VaR}_\alpha(L) - \mathbb{E}[L]\right)\sum_{i=1}^n \frac{\mathrm{cov}(L,L_i)}{\sigma^2(L)} \\ &= \mathbb{E}[L] + \left(\mathrm{VaR}_\alpha(L) - \mathbb{E}[L]\right) \\ &= \mathrm{VaR}_\alpha(L)\end{aligned}$$

Because the least squares estimator is $\hat{\beta} = \text{cov}(L, L_i)/\sigma^2(L)$, we deduce that:

$$\begin{aligned}\mathbb{E}[L_i | L = \text{VaR}_\alpha(L)] &= \hat{\beta}\,\text{VaR}_\alpha(L) + \mathbb{E}[\varepsilon_i] \\ &= \hat{\beta}\,\text{VaR}_\alpha(L) + \left(\mathbb{E}[L_i] - \hat{\beta}\mathbb{E}[L]\right) \\ &= \mathbb{E}[L_i] + \hat{\beta}\left(\text{VaR}_\alpha(L) - \mathbb{E}[L]\right)\end{aligned}$$

Epperlein and Smillie (2006) extend Formula (2.21) in the case of non-elliptical distributions. If we consider the generalized conditional expectation $\mathbb{E}[L_i | L = x] = f(x)$ where the function f is unknown, the estimator is given by the kernel regression[97]:

$$\hat{f}(x) = \frac{\sum_{s=1}^{n_S} \mathcal{K}(L_s - x) L_{i,s}}{\sum_{s=1}^{n_S} \mathcal{K}(L_s - x)}$$

where $\mathcal{K}(u)$ is the kernel function. We deduce that:

$$\mathcal{RC}_i = \hat{f}(\text{VaR}_\alpha(L))$$

Epperlein and Smillie (2006) note however that this risk decomposition does not satisfy the Euler allocation principle. This is why they propose the following correction:

$$\begin{aligned}\mathcal{RC}_i &= \frac{\text{VaR}_\alpha(L)}{\sum_{i=1}^{n} \mathcal{RC}_i} \hat{f}(\text{VaR}_\alpha(L)) \\ &= \text{VaR}_\alpha(L) \frac{\sum_{s=1}^{n_S} \mathcal{K}(L_s - \text{VaR}_\alpha(L)) L_{i,s}}{\sum_{i=1}^{n}\sum_{s=1}^{n_S} \mathcal{K}(L_s - \text{VaR}_\alpha(L)) L_{i,s}} \\ &= \text{VaR}_\alpha(L) \frac{\sum_{s=1}^{n_S} \mathcal{K}(L_s - \text{VaR}_\alpha(L)) L_{i,s}}{\sum_{s=1}^{n_S} \mathcal{K}(L_s - \text{VaR}_\alpha(L)) L_s}\end{aligned}$$

In Table 2.20, we have reported the risk contributions of the 99% value-at-risk for Apple and Coca-Cola stocks. The case **G** corresponds to the Gaussian value-at-risk whereas all the other cases correspond to the historical value-at-risk. For the case **R1**, the regularization weights are $\varpi_{99\%}(2;250) = \varpi_{99\%}(3;250) = \frac{1}{2}$ and $\varpi_{99\%}(s;250) = 0$ when $s \neq 2$ or $s \neq 3$. It corresponds to the classical interpolation method between the second and third order statistics. For the case **R2**, we have $\varpi_{99\%}(s;250) = \frac{1}{4}$ when $s \leq 4$ and $\varpi_{99\%}(s;250) = 0$ when $s > 4$. The value-at-risk is therefore estimated by averaging the first four order statistics. The cases **E** and **K** correspond to the methods based on the elliptical and kernel approaches. For these two cases, we obtain a risk decomposition, which is closer to this obtained with the Gaussian method. This is quite logical as the Gaussian distribution is a special case of elliptical distributions and the kernel function is also Gaussian.

TABLE 2.20: Risk contributions calculated with regularization techniques

Asset	**G**	**R1**	**R2**	**E**	**K**
Apple	30.97	43.92	52.68	35.35	39.21
Coca-Cola	10.25	3.47	2.29	12.03	8.17
$\mathcal{R}(w)$	41.21	47.39	54.96	47.39	47.39

[97] $\hat{f}(x)$ is called the Nadaraya-Watson estimator (see Section 10.1.4.2 on page 641).

Example 20 *Let $L = L_1 + L_2$ be the portfolio loss where L_i $(i = 1, 2)$ is defined as follows:*

$$L_i = w_i \left(\mu_i + \sigma_i T_i\right)$$

and T_i has a Student's t distribution with the number of degrees of freedom ν_i. The dependence function between the losses (L_1, L_2) is given by the Clayton copula:

$$\mathbf{C}\left(u_1, u_2\right) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$$

For the numerical illustration, we consider the following values: $w_1 = 100$, $\mu_1 = 10\%$, $\sigma_1 = 20\%$, $\nu_1 = 6$, $w_2 = 200$, $\mu_2 = 10\%$, $\sigma_2 = 25\%$, $\nu_2 = 4$ and $\theta = 2$. The confidence level α of the value-at-risk is set to 90%.

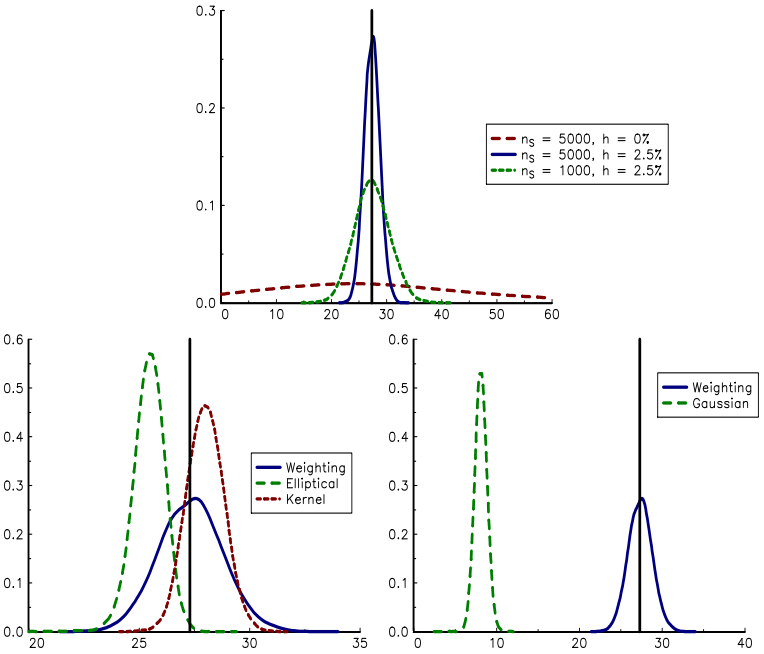


FIGURE 2.19: Density function of the different risk contribution estimators

In Figure 2.19, we compare the different statistical estimators of the risk contribution $\mathcal{RC}_1$ when we use $n_S = 5\,000$ simulations. Concerning the regularization method, we consider the following weight function applied to the order statistics of losses[98]:

$$\varpi_\alpha^L\left(s; n_S\right) = \frac{1}{2hn_S + 1} \cdot \mathbb{1}\left\{\frac{\left|s - q_\alpha\left(n_S\right)\right|}{n_S} \leq h\right\}$$

It corresponds to a uniform kernel on the range $[q_\alpha(n_S) - hn_S, q_\alpha(n_S) + hn_S]$. In the first panel, we report the probability density function of $\mathcal{RC}_1$ when h is equal to 0% and 2.5%. The case $h = 0\%$ is the estimator based on only one observation. We verify that the variance

[98] This is equivalent to use this weight function applied to the order statistics of P&Ls:

$$\varpi_\alpha\left(s; n_S\right) = \frac{1}{2hn_S + 1} \cdot \mathbb{1}\left\{\frac{\left|s - q_{\bar{\alpha}}\left(n_S\right)\right|}{n_S} \leq h\right\}$$

of this estimator is larger for $h = 0\%$ than for $h = 2.5\%$. However, we notice that this last estimator is a little biased, because we estimate the quantile 90% by averaging the order statistics corresponding to the range $[87.5\%, 92.5\%]$. In the second panel, we compare the weighting method with the elliptical and kernel approaches. These two estimators have a smaller variance, but a larger bias because they assume that the loss distribution is elliptical or may be estimated using a Gaussian kernel. Finally, the third panel shows the probability density function of $\mathcal{RC}_1$ estimated with the Gaussian value-at-risk.

The case of expected shortfall On page 70, we have shown that the expected shortfall is estimated as follows:

$$\mathrm{ES}_\alpha(L) = \frac{1}{q_\alpha(n_S)} \sum_{s=1}^{n_S} \mathbb{1}\{L_s \geq \mathrm{VaR}_\alpha(L)\} \cdot L_s$$

or:

$$\mathrm{ES}_\alpha(L) = -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{n_S} \mathbb{1}\{\Pi_s \leq -\mathrm{VaR}_\alpha(L)\} \cdot \Pi_s$$

It corresponds to the average of the losses larger or equal than the value-at-risk. It follows that:

$$\begin{aligned}
\mathrm{ES}_\alpha(L) &= -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{q_\alpha(n_S)} \Pi_{(s:n_S)} \\
&= -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{q_\alpha(n_S)} \Pi_{\mathfrak{R}_\Pi^{-1}(s)} \\
&= -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{q_\alpha(n_S)} \sum_{i=1}^{n} \Pi_{i,\mathfrak{R}_\Pi^{-1}(s)}
\end{aligned}$$

We deduce that:

$$\begin{aligned}
\mathcal{RC}_i &= -\frac{1}{q_\alpha(n_S)} \sum_{s=1}^{q_\alpha(n_S)} \Pi_{i,\mathfrak{R}_\Pi^{-1}(s)} \\
&= \frac{1}{q_\alpha(n_S)} \sum_{s=1}^{q_\alpha(n_S)} L_{i,\mathfrak{R}_\Pi^{-1}(s)}
\end{aligned}$$

In the Apple/Coca-Cola example, we recall that the 99% daily value-at-risk is equal to \$47.39. The corresponding expected shortfall is then the average of the two largest losses:

$$\mathrm{ES}_\alpha(w; \text{one day}) = \frac{84.34 + 51.46}{2} = \$67.90$$

For the risk contribution, we obtain[99]:

$$\mathcal{RC}_1 = \frac{87.39 + 41.69}{2} = \$64.54$$

[99] Because we have:

$$\Pi_{(1:250)} = -87.39 + 3.05 = -84.34$$

and:

$$\Pi_{(2:250)} = -41.69 - 9.77 = -51.46$$

and:

$$\mathcal{RC}_2 = \frac{-3.05 + 9.77}{2} = \$3.36$$

The corresponding risk decomposition is given in Tables 2.21 and 2.22 for $\alpha = 99\%$ and $\alpha = 97.5\%$. With the new rules of Basel III, the capital is higher for this example.

TABLE 2.21: Risk decomposition of the 99% historical expected shortfall

Asset	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
Apple	56.47%	114.29	64.54	95.05%
Coca-Cola	43.53%	7.72	3.36	4.95%
$\mathcal{R}(w)$			67.90	

TABLE 2.22: Risk decomposition of the 97.5% historical expected shortfall

Asset	w_i	$\mathcal{MR}_i$	$\mathcal{RC}_i$	$\mathcal{RC}_i^\star$
Apple	56.47%	78.48	44.32	91.31%
Coca-Cola	43.53%	9.69	4.22	8.69%
$\mathcal{R}(w)$			48.53	

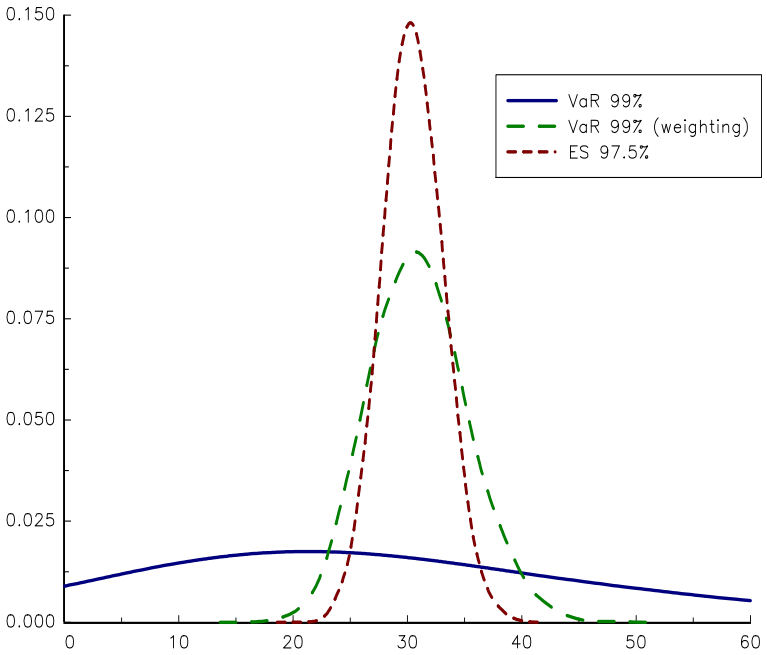


FIGURE 2.20: Probability density function of the $\mathcal{RC}_1$ estimator for the 99% VaR and 97.5% ES

In Figure 2.20, we report the probability density function of the $\mathcal{RC}_1$ estimator in the case of Example 20. We consider the 99% value-at-risk and the 97.5% expected shortfall with $n_S = 5\,000$ simulated scenarios. For the VaR risk measure, the risk contribution is estimated using respectively only one single observation and a weighting function corresponding to a

uniform window[100]. We notice that the estimator has a smaller variance with the expected shortfall risk measure. Of course, we can always reduce the variance of ES risk contributions by using the previous smoothing techniques (Scaillet, 2004), but this is less of an issue than for the value-at-risk measure.

2.4 Exercises

2.4.1 Calculating regulatory capital with the Basel I standardized measurement method

1. We consider an interest rate portfolio with the following exposures: a long position of \$100 mn on four-month instruments, a short position of \$50 mn on five-month instruments, a long position of \$10 mn on fifteen-year instruments and a short position of \$50 mn on twelve-year instruments.

   (a) Using BCBS (1996a), explain the maturity approach for computing the capital requirement due to the interest rate risk.

   (b) By assuming that the instruments correspond to bonds with coupons larger than 3%, calculate the capital requirement of the trading portfolio.

2. We consider the following portfolio of stocks:

Stock	3M	Exxon	IBM	Pfizer	AT&T	Cisco	Oracle
$\mathcal{L}_i$	100	100	10	50	60	90	
$\mathcal{S}_i$		50					80

   where $\mathcal{L}_i$ and $\mathcal{S}_i$ indicate the long and short exposures on stock i expressed in \$ mn.

   (a) Calculate the capital charge for the specific risk.

   (b) Calculate the capital charge for the general market risk.

   (c) How can the investor hedge the market risk of his portfolio by using S&P 500 futures contracts? What is the corresponding capital charge? Verify that the investor minimizes the total capital charge in this case.

3. We consider a net exposure $\mathcal{N}_w$ on an equity portfolio w. We note $\sigma(w)$ the annualized volatility of the portfolio return.

   (a) Calculate the required capital under the standardized measurement method.

   (b) Calculate the required capital under the internal model method if we assume that the bank uses a Gaussian value-at-risk[101].

   (c) Deduce an upper bound $\sigma(w) \leq \sigma^+$ under which the required capital under SMM is higher than the required capital under IMA.

   (d) Comment on these results.

[100] We set $h = 0.5\%$ meaning that the risk contribution is estimated with 51 observations for the 99% value-at-risk.

[101] We consider the Basel II capital requirement rules.

4. We consider the portfolio with the following long and short positions expressed in \$ mn:

Asset	EUR	JPY	CAD	Gold	Sugar	Corn	Cocoa
$\mathcal{L}_i$	100	50		50	50	60	90
$\mathcal{S}_i$	100	100	50			80	110

   (a) How do you explain that some assets present both long and short positions?

   (b) Calculate the required capital under the simplified approach.

5. We consider the following positions (in \$) of the commodity i:

Time band	0−1M	1M−3M	6M−1Y	1Y−2Y	2Y−3Y	3Y+
$\mathcal{L}_i(t)$	500	0	1 800	300	0	0
$\mathcal{S}_i(t)$	300	900	100	600	100	200

   (a) Using BCBS (1996a), explain the maturity ladder approach for commodities.

   (b) Compute the capital requirement.

2.4.2 Covariance matrix

We consider a universe of there stocks A, B and C.

1. The covariance matrix of stock returns is:

$$\Sigma = \begin{pmatrix} 4\% & & \\ 3\% & 5\% & \\ 2\% & -1\% & 6\% \end{pmatrix}$$

   (a) Calculate the volatility of stock returns.

   (b) Deduce the correlation matrix.

2. We assume that the volatilities are 10%, 20% and 30%. whereas the correlation matrix is equal to:

$$\rho = \begin{pmatrix} 100\% & & \\ 50\% & 100\% & \\ 25\% & 0\% & 100\% \end{pmatrix}$$

   (a) Write the covariance matrix.

   (b) Calculate the volatility of the portfolio $(50\%, 50\%, 0)$.

   (c) Calculate the volatility of the portfolio $(60\%, -40\%, 0)$. Comment on this result.

   (d) We assume that the portfolio is long \$150 on stock A, long \$500 on stock B and short \$200 on stock C. Find the volatility of this long/short portfolio.

3. We consider that the vector of stock returns follows a one-factor model:

$$R = \beta \mathcal{F} + \varepsilon$$

   We assume that $\mathcal{F}$ and ε are independent. We note $\sigma_{\mathcal{F}}^2$ the variance of $\mathcal{F}$ and $D = \text{diag}\left(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \tilde{\sigma}_3^2\right)$ the covariance matrix of idiosyncratic risks ε_t. We use the following numerical values: $\sigma_{\mathcal{F}} = 50\%$, $\beta_1 = 0.9$, $\beta_2 = 1.3$, $\beta_3 = 0.1$, $\tilde{\sigma}_1 = 5\%$, $\tilde{\sigma}_2 = 5\%$ and $\tilde{\sigma}_3 = 15\%$.

(a) Calculate the volatility of stock returns.

(b) Calculate the correlation between stock returns.

4. Let X and Y be two independent random vectors. We note $\mu(X)$ and $\mu(Y)$ the vector of means and $\Sigma(X)$ and $\Sigma(Y)$ the covariance matrices. We define the random vector $Z = (Z_1, Z_2, Z_3)$ where Z_i is equal to the product $X_i Y_i$.

   (a) Calculate $\mu(Z)$ and $\text{cov}(Z)$.

   (b) We consider that $\mu(X)$ is equal to zero and $\Sigma(X)$ corresponds to the covariance matrix of Question 2. We assume that Y_1, Y_2 and Y_3 are three independent uniform random variables $\mathcal{U}_{[0,1]}$. Calculate the 99% Gaussian value-at-risk of the portfolio corresponding to Question 2(d) when Z is the random vector of asset returns. Compare this value with the Monte Carlo VaR.

2.4.3 Risk measure

1. We denote $\mathbf{F}$ the cumulative distribution function of the loss L.

   (a) Give the mathematical definition of the value-at-risk and expected shortfall risk measures.

   (b) Show that:

$$\text{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathbf{F}^{-1}(t)\,\mathrm{d}t$$

   (c) We assume that L follows a Pareto distribution $\mathcal{P}(\theta, x_-)$ defined by:

$$\Pr\{L \le x\} = 1 - \left(\frac{x}{x_-}\right)^{-\theta}$$

   where $x \ge x_-$ and $\theta > 1$. Calculate the moments of order one and two. Interpret the parameters x_- and θ. Calculate $\text{ES}_\alpha(L)$ and show that:

$$\text{ES}_\alpha(L) > \text{VaR}_\alpha(L)$$

   (d) Calculate the expected shortfall when L is a Gaussian random variable $\mathcal{N}(\mu, \sigma^2)$. Show that:

$$\Phi(x) = -\frac{\phi(x)}{x^1} + \frac{\phi(x)}{x^3} + \ldots$$

   Deduce that:

$$\text{ES}_\alpha(L) \to \text{VaR}_\alpha(L) \text{ when } \alpha \to 1$$

   (e) Comment on these results in a risk management perspective.

2. Let $\mathcal{R}(L)$ be a risk measure of the loss L.

   (a) Is $\mathcal{R}(L) = \mathbb{E}[L]$ a coherent risk measure?

   (b) Same question if $\mathcal{R}(L) = \mathbb{E}[L] + \sigma(L)$.

3. We assume that the probability distribution $\mathbf{F}$ of the loss L is defined by:

$$\Pr\{L = \ell_i\} = \begin{cases} 20\% & \text{if} \quad \ell_i = 0 \\ 10\% & \text{if} \quad \ell_i \in \{1,2,3,4,5,6,7,8\} \end{cases}$$

   (a) Calculate ES_α for $\alpha = 50\%$, $\alpha = 75\%$ and $\alpha = 90\%$.

   (b) Let us consider two losses L_1 and L_2 with the same distribution $\mathbf{F}$. Build a joint distribution of (L_1, L_2) which does not satisfy the subadditivity property when the risk measure is the value-at-risk.

2.4.4 Value-at-risk of a long/short portfolio

We consider a long/short portfolio composed of a long position on asset A and a short position on asset B. The long exposure is equal to \$2 mn whereas the short exposure is equal to \$1 mn. Using the historical prices of the last 250 trading days of assets A and B, we estimate that the asset volatilities σ_A and σ_B are both equal to 20% per year and that the correlation $\rho_{A,B}$ between asset returns is equal to 50%. In what follows, we ignore the mean effect.

1. Calculate the Gaussian VaR of the long/short portfolio for a one-day holding period and a 99% confidence level.

2. How do you calculate the historical VaR? Using the historical returns of the last 250 trading days, the five worst scenarios of the 250 simulated daily P&L of the portfolio are $-58\,700$, $-56\,850$, $-54\,270$, $-52\,170$ and $-49\,231$. Calculate the historical VaR for a one-day holding period and a 99% confidence level.

3. We assume that the multiplication factor m_c is 3. Deduce the required capital if the bank uses an internal model based on the Gaussian value-at-risk. Same question when the bank uses the historical VaR. Compare these figures with those calculated with the standardized measurement method.

4. Show that the Gaussian VaR is multiplied by a factor equal to $\sqrt{7/3}$ if the correlation $\rho_{A,B}$ is equal to -50%. How do you explain this result?

5. The portfolio manager sells a call option on the stock A. The delta of the option is equal to 50%. What does the Gaussian value-at-risk of the long/short portfolio become if the nominal of the option is equal to \$2 mn? Same question when the nominal of the option is equal to \$4 mn. How do you explain this result?

6. The portfolio manager replaces the short position on the stock B by selling a call option on the stock B. The delta of the option is equal to 50%. Show that the Gaussian value-at-risk is minimum when the nominal is equal to four times the correlation $\rho_{A,B}$. Deduce then an expression of the lowest Gaussian VaR. Comment on these results.

2.4.5 Value-at-risk of an equity portfolio hedged with put options

We consider two stocks A and B and an equity index I. We assume that the risk model corresponds to the CAPM and we have:

$$R_j = \beta_j R_I + \varepsilon_j$$

where R_j and R_I are the returns of stock j and the index. We assume that R_I and ε_j are independent. The covariance matrix of idiosyncratic risks is diagonal and we note $\tilde{\sigma}_j$ the volatility of ε_j.

1. The parameters are the following: $\sigma^2(R_I) = 4\%$, $\beta_A = 0.5$, $\beta_B = 1.5$, $\tilde{\sigma}_A^2 = 3\%$ and $\tilde{\sigma}_B^2 = 7\%$.

   (a) Calculate the volatility of stocks A and B and the cross-correlation.

   (b) Find the correlation between the stocks and the index.

   (c) Deduce the covariance matrix.

2. The current price of stocks A and B is equal to \$100 and \$50 whereas the value of the index is equal to \$50. The composition of the portfolio is 4 shares of A, 10 shares of B and 5 shares of I.

   (a) Determine the Gaussian value-at-risk for a confidence level of 99% and a 10-day holding period.

   (b) Using the historical returns of the last 260 trading days, the five lowest simulated daily P&Ls of the portfolio are -62.39, -55.23, -52.06, -51.52 and -42.83. Calculate the historical VaR for a confidence level of 99% and a 10-day holding period.

   (c) What is the regulatory capital[102] if the bank uses an internal model based on the Gaussian value-at-risk? Same question when the bank uses the historical value-at-risk. Compare these figures with those calculated with the standardized measurement method.

3. The portfolio manager would like to hedge the directional risk of the portfolio. For that, he purchases put options on the index I at a strike of \$45 with a delta equal to -25%. Write the expression of the P&L using the delta approach.

   (a) How many options should the portfolio manager purchase for hedging 50% of the index exposure? Deduce the Gaussian value-at-risk of the corresponding portfolio?

   (b) The portfolio manager believes that the purchase of 96 put options minimizes the value-at-risk. What is the basis for his reasoning? Do you think that it is justified? Calculate then the Gaussian VaR of this new portfolio.

2.4.6 Risk management of exotic options

Let us consider a short position on an exotic option, whose its current value $\boldsymbol{\mathcal{C}}_t$ is equal to \$6.78. We assume that the price S_t of the underlying asset is \$100 and the implied volatility Σ_t is equal to 20%.

1. At time $t+1$, the value of the underlying asset is \$97 and the implied volatility remains constant. We find that the P&L of the trader between t and $t+1$ is equal to \$1.37. Can we explain the P&L by the sensitivities knowing that the estimates of delta $\boldsymbol{\Delta}_t$, gamma $\boldsymbol{\Gamma}_t$ and vega[103] $\boldsymbol{\upsilon}_t$ are respectively equal to 49%, 2% and 40%?

2. At time $t+2$, the price of the underlying asset is \$97 while the implied volatility increases from 20% to 22%. The value of the option $\boldsymbol{\mathcal{C}}_{t+2}$ is now equal to \$6.17. Can we explain the P&L by the sensitivities knowing that the estimates of delta $\boldsymbol{\Delta}_{t+1}$, gamma $\boldsymbol{\Gamma}_{t+1}$ and vega $\boldsymbol{\upsilon}_{t+1}$ are respectively equal to 43%, 2% and 38%?

3. At time $t+3$, the price of the underlying asset is \$95 and the value of the implied volatility is 19%. We find that the P&L of the trader between $t+2$ and $t+3$ is equal to \$0.58. Can we explain the P&L by the sensitivities knowing that the estimates of delta $\boldsymbol{\Delta}_{t+2}$, gamma $\boldsymbol{\Gamma}_{t+2}$ and vega $\boldsymbol{\upsilon}_{t+2}$ are respectively equal to 44%, 1.8% and 38%.

4. What can we conclude in terms of model risk?

[102] We assume that the multiplication factor m_c is equal to 3.
[103] Measured in volatility points.

2.4.7 P&L approximation with Greek sensitivities

1. Let $\boldsymbol{\mathcal{C}}_t$ be the value of an option at time t. Define the delta, gamma, theta and vega coefficients of the option.

2. We consider an European call option with strike K. Give the value of option in the case of the Black-Scholes model. Deduce then the Greek coefficients.

3. We assume that the underlying asset is a non-dividend stock, the residual maturity of the call option is equal to one year, the current price of the stock is equal to \$100 and the interest rate is equal to 5%. We also assume that the implied volatility is constant and equal to 20%. In the table below, we give the value of the call option $\boldsymbol{\mathcal{C}}_0$ and the Greek coefficients $\boldsymbol{\Delta}_0$, $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Theta}_0$ for different values of K:

K	80	95	100	105	120
$\boldsymbol{\mathcal{C}}_0$	24.589	13.346	10.451	8.021	3.247
$\boldsymbol{\Delta}_0$	0.929	0.728	0.637	0.542	0.287
$\boldsymbol{\Gamma}_0$	0.007	0.017	0.019	0.020	0.017
$\boldsymbol{\Theta}_0$	−4.776	−6.291	−6.414	−6.277	−4.681

   (a) Explain how these values have been calculated. Comment on these numerical results.

   (b) One day later, the value of the underlying asset is \$102. Using the Black-Scholes formula, we obtain:

K	80	95	100	105	120
$\boldsymbol{\mathcal{C}}_1$	26.441	14.810	11.736	9.120	3.837

   Explain how the option premium $\boldsymbol{\mathcal{C}}_1$ is calculated. Deduce then the P&L of a long position on this option for each strike K.

   (c) For each strike price, calculate an approximation of the P&L by considering the sensitivities $\boldsymbol{\Delta}$, $\boldsymbol{\Delta} - \boldsymbol{\Gamma}$, $\boldsymbol{\Delta} - \boldsymbol{\Theta}$ and $\boldsymbol{\Delta} - \boldsymbol{\Gamma} - \boldsymbol{\Theta}$. Comment on these results.

   (d) Six months later, the value of the underlying asset is \$148. Repeat Questions 3(b) and 3(c) with these new parameters. Comment on these results.

2.4.8 Calculating the non-linear quadratic value-at-risk

1. Let $X \sim \mathcal{N}(0,1)$. Show that the even moments of X are given by the following relationship:
$$\mathbb{E}\left[X^{2n}\right] = (2n-1)\,\mathbb{E}\left[X^{2n-2}\right]$$
with $n \in \mathbb{N}$. Calculate the odd moments of X.

2. We consider a long position on a call option. The current price S_t of the underlying asset is equal to \$100, whereas the delta and the gamma of the option are respectively equal to 50% and 2%. We assume that the annual return of the asset follows a Gaussian distribution with an annual volatility equal to 32.25%.

   (a) Calculate the daily Gaussian value-at-risk using the delta approximation with a 99% confidence level.

   (b) Calculate the daily Gaussian value-at-risk by considering the delta-gamma approximation.

   (c) Deduce the daily Cornish-Fisher value-at-risk.

3. Let $X \sim \mathcal{N}(\mu, I)$ and $Y = X^\top AX$ with A a symmetric square matrix.

   (a) We recall that:

$$\begin{aligned}\mathbb{E}[Y] &= \mu^\top A\mu + \operatorname{tr}(A)\\ \mathbb{E}\left[Y^2\right] &= \mathbb{E}^2[Y] + 4\mu^\top A^2\mu + 2\operatorname{tr}\left(A^2\right)\end{aligned}$$

   Deduce the moments of $Y = X^\top AX$ when $X \sim \mathcal{N}(\mu, \Sigma)$.

   (b) We suppose that $\mu = \mathbf{0}$. We recall that:

$$\begin{aligned}\mathbb{E}\left[Y^3\right] &= \left(\operatorname{tr}(A)\right)^3 + 6\operatorname{tr}(A)\operatorname{tr}\left(A^2\right) + 8\operatorname{tr}\left(A^3\right)\\ \mathbb{E}\left[Y^4\right] &= \left(\operatorname{tr}(A)\right)^4 + 32\operatorname{tr}(A)\operatorname{tr}\left(A^3\right) + 12\left(\operatorname{tr}\left(A^2\right)\right)^2 +\\ &\quad 12\left(\operatorname{tr}(A)\right)^2\operatorname{tr}\left(A^2\right) + 48\operatorname{tr}\left(A^4\right)\end{aligned}$$

   Compute the moments, the skewness and the excess kurtosis of $Y = X^\top AX$ when $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$.

4. We consider a portfolio $w = (w_1, \ldots, w_n)$ of options. We assume that the vector of daily asset returns is distributed according to the Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma)$. We note $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ the vector of deltas and the matrix of gammas.

   (a) Calculate the daily Gaussian value-at-risk using the delta approximation. Define the analytical expression of the risk contributions.

   (b) Calculate the daily Gaussian value-at-risk by considering the delta-gamma approximation.

   (c) Calculate the daily Cornish-Fisher value-at-risk when assuming that the portfolio is delta neutral.

   (d) Calculate the daily Cornish-Fisher value-at-risk in the general case by only considering the skewness.

5. We consider a portfolio composed of 50 options in a first asset, 20 options in a second asset and 20 options in a third asset. We assume that the gamma matrix is:

$$\boldsymbol{\Gamma} = \begin{pmatrix} 4.0\% & & \\ 1.0\% & 1.0\% & \\ 0.0\% & -0.5\% & 1.0\% \end{pmatrix}$$

   The actual price of the assets is normalized and is equal to 100. The daily volatility levels of the assets are respectively equal to 1%, 1.5% and 2% whereas the correlation matrix of asset returns is:

$$\rho = \begin{pmatrix} 100\% & & \\ 50\% & 100\% & \\ 25\% & 15\% & 100\% \end{pmatrix}$$

   (a) Compare the different methods to compute the daily value-at-risk with a 99% confidence level if the portfolio is delta neutral.

   (b) Same question if we now consider that the deltas are equal to 50%, 40% and 60%. Compute the risk decomposition in the case of the delta and delta-gamma approximations. What do you notice?

2.4.9 Risk decomposition of the expected shortfall

We consider a portfolio composed of n assets. We assume that asset returns $R = (R_1, \ldots, R_n)$ are normally distributed: $R \sim \mathcal{N}(\mu, \Sigma)$. We note $L(w)$ the loss of the portfolio.

1. Find the distribution of $L(w)$.
2. Define the expected shortfall $\mathrm{ES}_\alpha(w)$. Calculate its expression in the present case.
3. Calculate the risk contribution $\mathcal{RC}_i$ of asset i. Deduce that the expected shortfall verifies the Euler allocation principle.
4. Give the expression of $\mathcal{RC}_i$ in terms of conditional loss. Retrieve the formula of $\mathcal{RC}_i$ found in Question 3. What is the interest of the conditional representation?

2.4.10 Expected shortfall of an equity portfolio

We consider an investment universe, which is composed of two stocks A and B. The current price of the two stocks is respectively equal to \$100 and \$200, their volatilities are equal to 25% and 20% whereas the cross-correlation is equal to -20%. The portfolio is long on 4 stocks A and 3 stocks B.

1. Calculate the Gaussian expected shortfall at the 97.5% confidence level for a ten-day time horizon.
2. The eight worst scenarios of daily stock returns among the last 250 historical scenarios are the following:

s	1	2	3	4	5	6	7	8
R_A	-3%	-4%	-3%	-5%	-6%	$+3\%$	$+1\%$	-1%
R_B	-4%	$+1\%$	-2%	-1%	$+2\%$	-7%	-3%	-2%

Calculate then the historical expected shortfall at the 97.5% confidence level for a ten-day time horizon.

2.4.11 Risk measure of a long/short portfolio

We consider an investment universe, which is composed of two stocks A and B. The current prices of the two stocks are respectively equal to \$50 and \$20. Their volatilities are equal to 25% and 20% whereas the cross-correlation is equal to $+12.5\%$. The portfolio is long on 2 stocks A and short on 5 stocks B.

1. Gaussian risk measure
   (a) Calculate the Gaussian value-at-risk at the 99% confidence level for a ten-day time horizon.
   (b) Calculate the Gaussian expected shortfall at the 97.5% confidence level for a ten-day time horizon.
2. Historical risk measure
   The ten worst scenarios of daily stock returns (expressed in %) among the last 250 historical scenarios are the following:

s	1	2	3	4	5	6	7	8	9	10
R_A	-0.6	-3.7	-5.8	-4.2	-3.7	0.0	-5.7	-4.3	-1.7	-4.1
R_B	5.7	2.3	-0.7	0.6	0.9	4.5	-1.4	0.0	2.3	-0.2
D	-6.3	-6.0	-5.1	-4.8	-4.6	-4.5	-4.3	-4.3	-4.0	-3.9

   where $D = R_A - R_B$ is the difference of the returns.

(a) Calculate the historical value-at-risk at the 99% confidence level for a ten-day time horizon.

(b) Calculate the historical expected shortfall at the 97.5% confidence level for a ten-day time horizon.

(c) Give an approximation of the capital charge under Basel II, Basel 2.5 and Basel III standards by considering the historical risk measure[104].

2.4.12 Kernel estimation of the expected shortfall

1. We consider a random variable X. We note $\mathcal{K}(u)$ the kernel function associated to the sample $\{x_1, \ldots, x_n\}$. Show that:

$$\begin{aligned}\mathbb{E}\left[X \cdot \mathbb{1}\left\{X \leq x\right\}\right] &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} x_i \mathcal{K}(u)\,\mathrm{d}u + \\ &\quad \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-x_i}{\boldsymbol{h}}} \boldsymbol{h}\, u\mathcal{K}(u)\,\mathrm{d}u\end{aligned}$$

2. Find the expression of the first term by considering the integrated kernel function $\mathcal{I}(u)$.

3. Show that the second term tends to zero when $\boldsymbol{h} \to 0$.

4. Deduce an approximation of the expected shortfall $\mathrm{ES}_\alpha(w; h)$.

[104] We assume that the multiplicative factor is equal to 3 (Basel II), and the '*stressed*' risk measure is 2 times the '*normal*' risk measure (Basel 2.5).

Chapter 3

Credit Risk

In this chapter, we give an overview of the credit market. It concerns loans and bonds, but also credit derivatives whose development was impressive during the 2000s. A thorough knowledge of the products is necessary to understand the regulatory framework for computing the capital requirements for credit risk. In this second section, we will therefore compare Basel I, Basel II and Basel III approaches. The case of counterparty credit risk will be treated in the next chapter, which focuses on collateral risk. Finally, the last section is dedicated to the modeling of credit risk. We will develop the statistical methods for modeling and estimating the main parameters (probability of default, loss given default and default correlations) and we will show the tools of credit risk management. Concerning credit scoring models, we refer to Chapter 15, which is fully dedicated on this topic.

3.1 The market of credit risk

3.1.1 The loan market

In this section, we present the traditional debt market of loans based on banking intermediation, as opposed to the financial market of debt securities (money market instruments, bonds and notes). We generally distinguish this credit market along two main lines: counterparties and products.

Counterparties are divided into 4 main categories: sovereign, financial, corporate and retail. Banking groups have adopted this customer-oriented approach by differentiating retail banking and corporate and investment banking (CIB) businesses. Retail banking refers to individuals. It may also include micro-sized firms and small and medium-sized enterprises (SME). CIBs concern middle market firms, corporates, financial institutions and public entities. In retail banking, the bank pursues a client segmentation, meaning that all the clients that belongs to the same segment have the same conditions in terms of financing and financial investments. This also implies that the pricing of the loan is the same for two individuals of the same segment. The issue for the bank is then to propose or not a loan offer to his client. For that, the bank uses statistical decision-making methods, which are called credit scoring models. Contrary to this binary approach (yes or no), CIBs have a personalized approach to their clients. They estimate their probability of default and changes the pricing condition of the loan on the basis of the results. A client with a low default probability will have a lower rate or credit spread than a client with a higher default probability for the same loan.

The household credit market is organized as follows: mortgage and housing debt, consumer credit and student loans. A mortgage is a debt instrument secured by the collateral of a real estate property. In the case where the borrower defaults on the loan, the lender can take possession and sell the secured property. For instance, the home buyer pledges his house to the bank in a residential mortgage. This type of credit is very frequent in

English-speaking countries, notably England and the United States. In continental Europe, home loans are generally not collateralized for a primary home. This is not always the case for buy-to-let investments and second-home loans. Consumer credit is used for equipment financing or leasing. We usually make the distinction between auto loans, credit cards, revolving credit and other loans (personal loans and sales financing). Auto loans are personal loans to purchase a car. Credit cards and revolving credit are two forms of personal lines of credit. Revolving credit facilities for individuals are very popular in the US. It can be secured, as in the case of a home equity line of credit (HELOC). Student loans are used to finance educational expenses, for instance post-graduate studies at the university. The corporate credit market is organized differently, because large corporates have access to the financial market for long-term financing. This explains that revolving credit facilities are essential to provide liquidity for the firm's day-to-day operations. The average maturity is then lower for corporates than for individuals.

Credit statistics for the private non-financial sector (households and non-financial corporations) are reported in Figures 3.1 and 3.2. These statistics include loan instruments, but also debt securities. In the case of the United States[1], we notice that the credit amount for households[2] is close to the figure for non-financial business. We also observe the significant share of consumer credit and the strong growth of student loans. Figure 3.2 illustrates the evolution of debt outstanding[3] for different countries: China, United Kingdom, Japan, United States and the Euro area. In China, the annual growth rate is larger than 20% these last five years. Even if credit for households develops much faster than credit for corporations, it only represents 24% of the total credit market of the private non-financial sector. The Chinese market contrasts with developed markets where the share of household credit is larger[4] and growth rates are almost flat since the 2008 financial crisis. The Japanese case is also very specific, because this country experienced a strong financial crisis after the bursting of a bubble in the 1990s. At that time, the Japanese market was the world's leading market followed by the United States.

3.1.2 The bond market

Contrary to loan instruments, bonds are debt securities that are traded in a financial market. The primary market concerns the issuance of bonds whereas bond trading is organized through the secondary market. The bond issuance market is dominated by two sectors: central and local governments (including public entities) and corporates. This is the principal financing source for government projects and public budget deficits. Large corporates also use extensively the bond market for investments, business expansions and external growth. The distinction government bonds/corporate bonds was crucial before the 2008 Global Financial Crisis. Indeed, it was traditionally believed that government bonds (in developed countries) were not risky because the probability of default was very low. In this case, the main risk was the interest rate risk, which is a market risk. Conversely, corporate bonds were supposed to be risky because the probability of default was higher. Besides the interest rate risk, it was important to take into account the credit risk. Bonds issued from the financial and banking sector were considered as low risk investments. Since 2008,

[1] Data are from the statistical release Z.1 "*Financial Accounts of the United States*". They are available from the website of the Federal Reserve System: `https://www.federalreserve.gov/releases/z1` or more easily with the database of the Federal Reserve Bank of St. Louis: `https://fred.stlouisfed.org`.

[2] Data for households include non-profit institutions serving households (NPISH).

[3] Data are collected by the Bank for International Settlements and are available in the website of the BIS: `https://www.bis.org/statistics`. The series are adjusted for breaks (Dembiermont *et al.*, 2013) and we use the average exchange rate from 2000 to 2014 in order to obtain credit amounts in USD.

[4] This is especially true in the UK and the US.

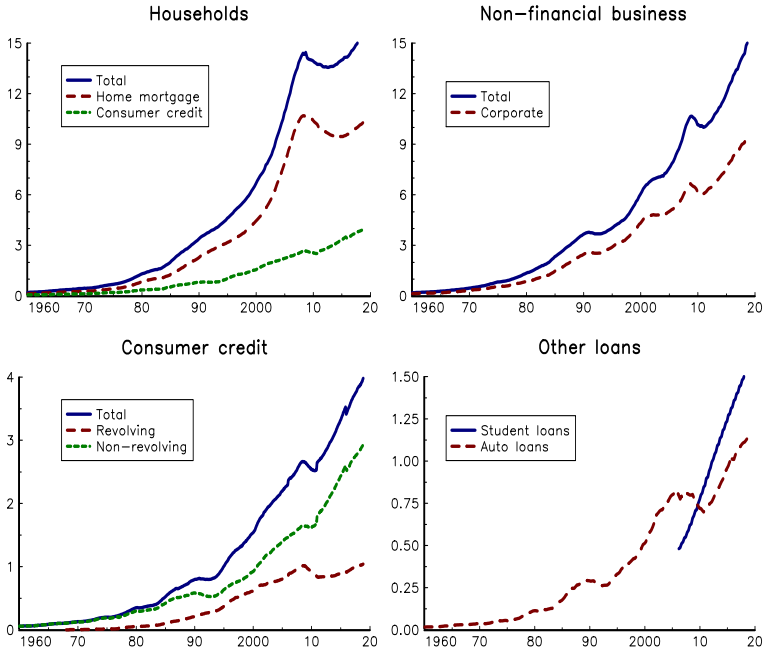


FIGURE 3.1: Credit debt outstanding in the United States (in $ tn)

Source: Board of Governors of the Federal Reserve System (2019).

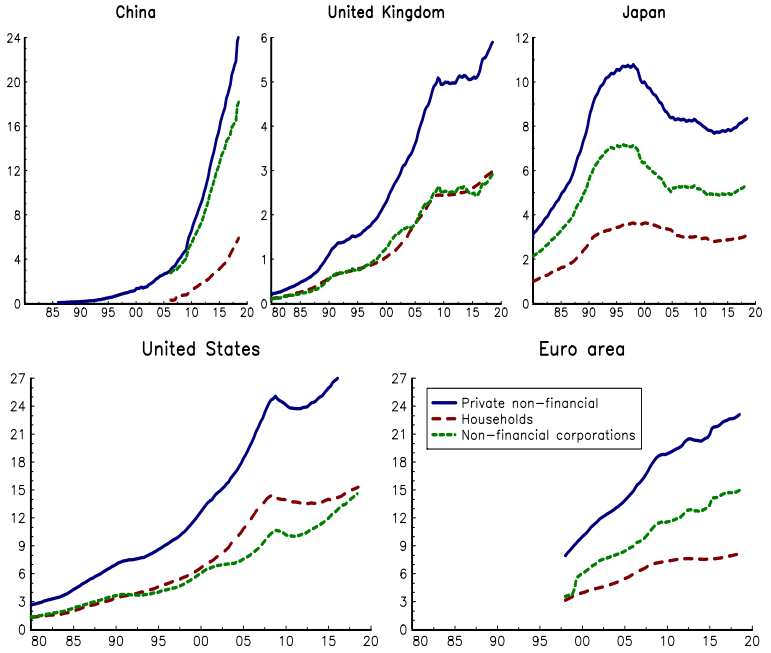


FIGURE 3.2: Credit to the private non-financial sector (in $ tn)

Source: Bank for International Settlements (2019) and author's calculations.

TABLE 3.1: Debt securities by residence of issuer (in $ bn)

		Dec. 2004	Dec. 2007	Dec. 2010	Dec. 2017
Canada	Gov.	682	841	1 149	1 264
	Fin.	283	450	384	655
	Corp.	212	248	326	477
	Total	1 180	1 544	1 863	2 400
France	Gov.	1 236	1 514	1 838	2 258
	Fin.	968	1 619	1 817	1 618
	Corp.	373	382	483	722
	Total	2 576	3 515	4 138	4 597
Germany	Gov.	1 380	1 717	2 040	1 939
	Fin.	2 296	2 766	2 283	1 550
	Corp.	133	174	168	222
	Total	3 809	4 657	4 491	3 712
Italy	Gov.	1 637	1 928	2 069	2 292
	Fin.	772	1 156	1 403	834
	Corp.	68	95	121	174
	Total	2 477	3 178	3 593	3 299
Japan	Gov.	6 336	6 315	10 173	9 477
	Fin.	2 548	2 775	3 451	2 475
	Corp.	1 012	762	980	742
	Total	9 896	9 852	14 604	12 694
Spain	Gov.	462	498	796	1 186
	Fin.	434	1 385	1 442	785
	Corp.	15	19	19	44
	Total	910	1 901	2 256	2 015
UK	Gov.	798	1 070	1 674	2 785
	Fin.	1 775	3 127	3 061	2 689
	Corp.	452	506	473	533
	Total	3 027	4 706	5 210	6 011
US	Gov.	6 459	7 487	12 072	17 592
	Fin.	12 706	17 604	15 666	15 557
	Corp.	3 004	3 348	3 951	6 137
	Total	22 371	28 695	31 960	39 504

Source: Bank for International Settlements (2019).

this difference between non-risky and risky bonds has disappeared, meaning that all issuers are risky. The 2008 GFC had also another important consequence on the bond market. It is today less liquid even for sovereign bonds. Liquidity risk is then a concern when measuring and managing the risk of a bond portfolio. This point is developed in Chapter 6.

3.1.2.1 Statistics of the bond market

In Table 3.1, we indicate the outstanding amount of debt securities by residence of issuer[5]. The total is split into three sectors: general governments (Gov.), financial corporations (Fin.) and non-financial corporations (Corp.). In most countries, debt securities issued by general governments largely dominate, except in the UK and US where debt securities

[5]The data are available in the website of the BIS: https://www.bis.org/statistics.

issued by financial corporations (banks and other financial institutions) are more important. The share of non-financial business varies considerably from one country to another. For instance, it represents less than 10% in Germany, Italy, Japan and Spain, whereas it is equal to 20% in Canada. The total amount of debt securities tends to rise, with the notable exception of Germany, Japan and Spain.

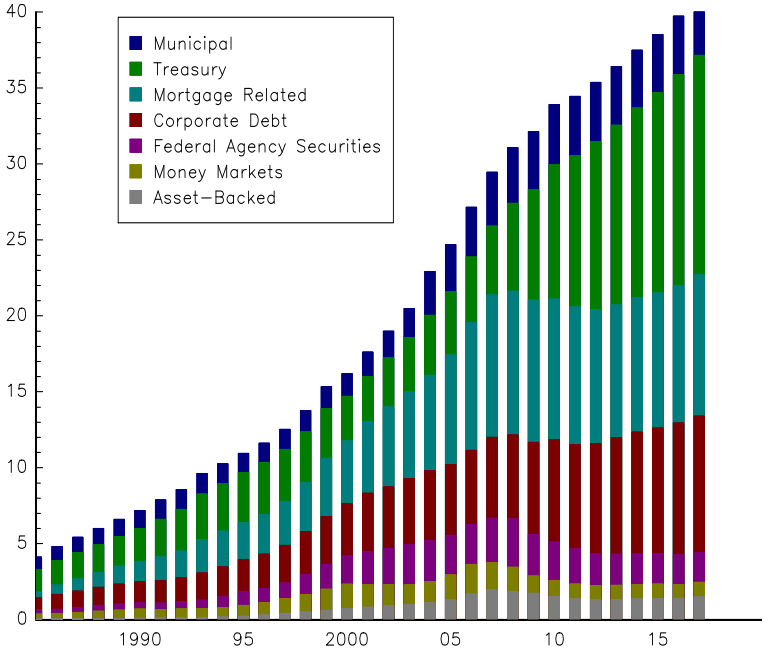


FIGURE 3.3: US bond market outstanding (in $ tn)

Source: Securities Industry and Financial Markets Association (2019a).

The analysis of the US market is particularly interesting and relevant. Using the data collected by the Securities Industry and Financial Markets Association[6] (SIFMA), we have reported in Figure 3.3 the evolution of outstanding amount for the following sectors: municipal bonds, treasury bonds, mortgage-related bonds, corporate related debt, federal agency securities, money markets and asset-backed securities. We notice an important growth during the beginning of the 2000s (see also Figure 3.4), followed by a slowdown after 2008. However, the debt outstanding continues to grow because the average maturity of new issuance increases. Another remarkable fact is the fall of the liquidity, which can be measured by the average daily volume (ADV). Figure 3.5 shows that the ADV of treasury bonds remains constant since 2000 whereas the outstanding amount has been multiplied by four during the same period. We also notice that the turnover of US bonds mainly concerns treasury and agency MBS bonds. The liquidity on the other sectors is very poor. For instance, according to SIFMA (2019a), the ADV of US corporate bonds is less than $30 bn in 2014, which is 22 times lower than the ADV for treasury bonds[7].

[6] Data are available in the website of the SIFMA: https://www.sifma.org/resources/archive/research/.

[7] However, the ratio between their outstanding amount is only 1.6.

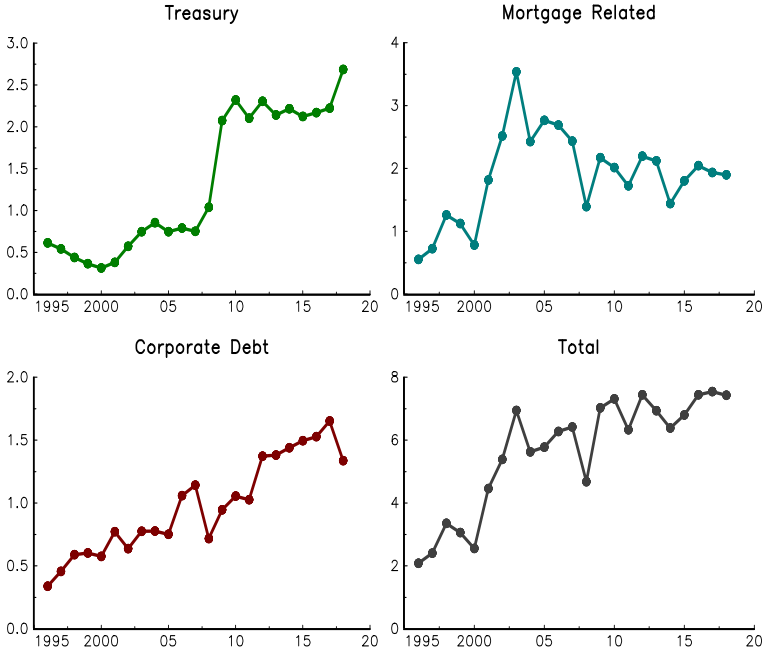


FIGURE 3.4: US bond market issuance (in $ tn)

Source: Securities Industry and Financial Markets Association (2019a).

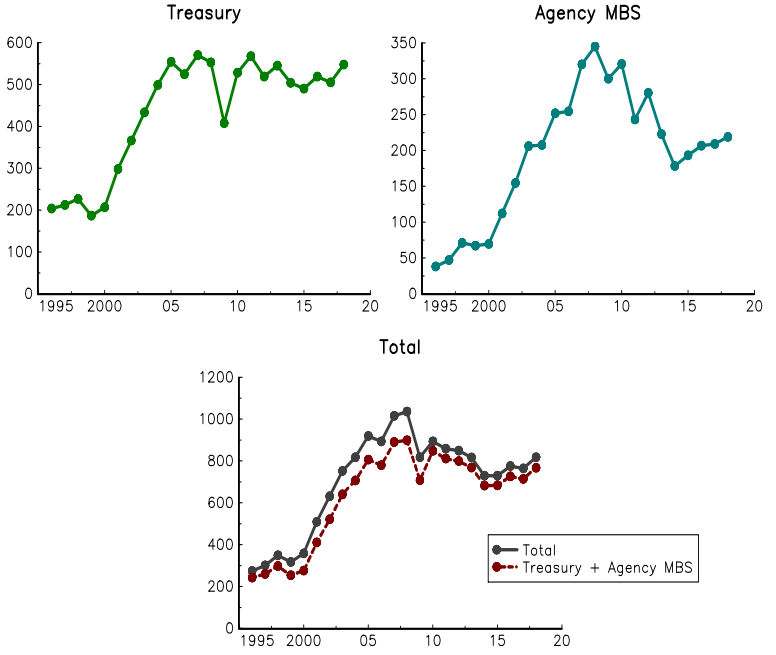


FIGURE 3.5: Average daily trading volume in US bond markets (in $ bn)

Source: Securities Industry and Financial Markets Association (2019a).

3.1.2.2 Bond pricing

We first explain how to price a bond by only considering the interest rate risk. Then, we introduce the default risk and define the concept of credit spread, which is key in credit risk modeling.

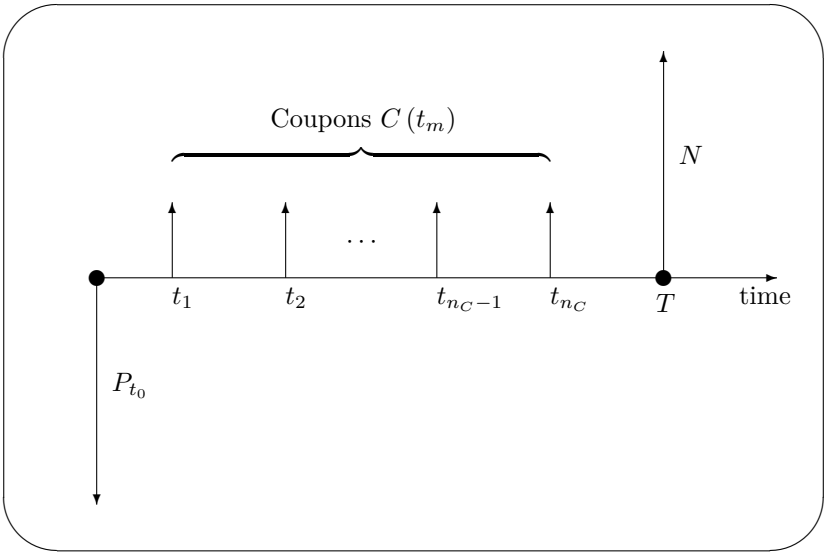


FIGURE 3.6: Cash flows of a bond with a fixed coupon rate

Without default risk We consider that the bond pays coupons $C(t_m)$ with fixing dates t_m and the notional N (or the par value) at the maturity date T. We have reported an example of a cash flows scheme in Figure 3.6. Knowing the yield curve[8], the price of the bond at the inception date t_0 is the sum of the present values of all the expected coupon payments and the par value:

$$P_{t_0} = \sum_{m=1}^{n_C} C(t_m) \cdot B_{t_0}(t_m) + N \cdot B_{t_0}(T)$$

where $B_t(t_m)$ is the discount factor at time t for the maturity date t_m. When the valuation date is not the issuance date, the previous formula remains valid if we take into account the accrued interests. In this case, the buyer of the bond has the benefit of the next coupon. The price of the bond then satisfies:

$$P_t + AC_t = \sum_{t_m \geq t} C(t_m) \cdot B_t(t_m) + N \cdot B_t(T) \tag{3.2}$$

[8] A convenient way to define the yield curve is to use a parametric model for the zero-coupon rates $R_t(T)$. The most famous model is the parsimonious functional form proposed by Nelson and Siegel (1987):

$$\begin{aligned} R_t(T) &= \theta_1 + \theta_2 \left(\frac{1 - \exp(-(T-t)/\theta_4)}{(T-t)/\theta_4} \right) + \\ &\quad \theta_3 \left(\frac{1 - \exp(-(T-t)/\theta_4)}{(T-t)/\theta_4} - \exp(-(T-t)/\theta_4) \right) \end{aligned} \tag{3.1}$$

This is a model with four parameters: θ_1 is a parameter of level, θ_2 is a parameter of rotation, θ_3 controls the shape of the curve and θ_4 permits to localize the break of the curve. We also note that the short-term and long-term interest rates $R_t(t)$ and $R_t(\infty)$ are respectively equal to $\theta_1 + \theta_2$ and θ_1.

Here, AC_t is the accrued coupon:

$$AC_t = C(t_c) \cdot \frac{t - t_c}{365}$$

and t_c is the last coupon payment date with $c = \{m : t_{m+1} > t, t_m \leq t\}$. $P_t + AC_t$ is called the '*dirty price*' whereas P_t refers to the '*clean price*'. The term structure of interest rates impacts the bond price. We generally distinguish three movements:

1. The movement of level corresponds to a parallel shift of interest rates.
2. A twist in the slope of the yield curve indicates how the spread between long and short interest rates moves.
3. A change in the curvature of the yield curve affects the convexity of the term structure.

All these movements are illustrated in Figure 3.7.

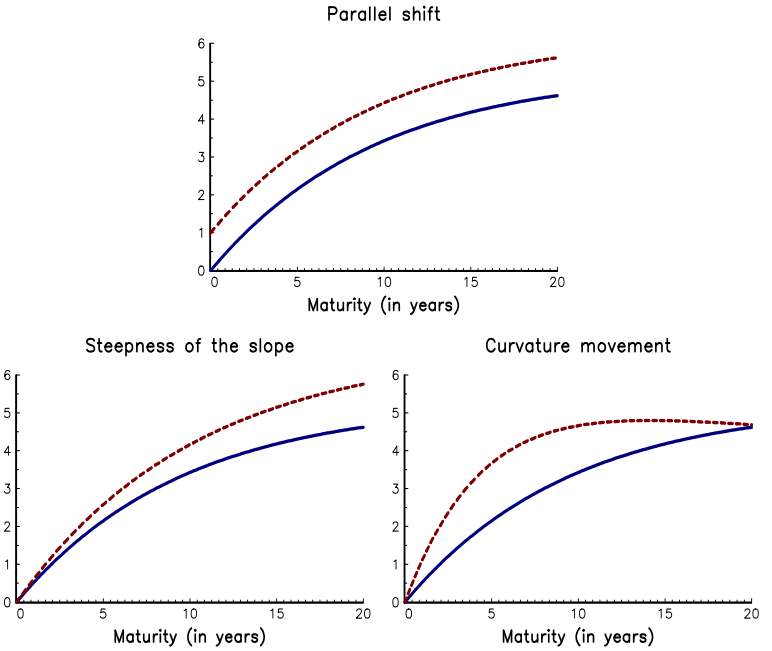


FIGURE 3.7: Movements of the yield curve

The yield to maturity y of a bond is the constant discount rate which returns its market price:

$$\sum_{t_m \geq t} C(t_m) e^{-(t_m - t)y} + Ne^{-(T-t)y} = P_t + AC_t$$

We also define the sensitivity[9] S of the bond price as the derivative of the clean price P_t with respect to the yield to maturity y:

$$\begin{aligned} S &= \frac{\partial P_t}{\partial y} \\ &= -\sum_{t_m \geq t} (t_m - t) C(t_m) e^{-(t_m - t)y} - (T - t) Ne^{-(T-t)y} \end{aligned}$$

[9]This sensitivity is also called the $-duration or DV01.

It indicates how the P&L of a long position on the bond moves when the yield to maturity changes:

$$\Pi \approx S \cdot \Delta y$$

Because $S < 0$, the bond price is a decreasing function with respect to interest rates. This implies that an increase of interest rates reduces the value of the bond portfolio.

Example 21 *We assume that the term structure of interest rates is generated by the Nelson-Siegel model with* $\theta_1 = 5\%$, $\theta_2 = -5\%$, $\theta_3 = 6\%$ *and* $\theta_4 = 10$. *We consider a bond with a constant annual coupon of* 5%. *The nominal of the bond is* \$100. *We would like to price the bond when the maturity* T *ranges from* 1 *to* 5 *years.*

TABLE 3.2: Price, yield to maturity and sensitivity of bonds

T	$R_t(T)$	$B_t(T)$	P_t	y	S
1	0.52%	99.48	104.45	0.52%	−104.45
2	0.99%	98.03	107.91	0.98%	−210.86
3	1.42%	95.83	110.50	1.39%	−316.77
4	1.80%	93.04	112.36	1.76%	−420.32
5	2.15%	89.82	113.63	2.08%	−520.16

TABLE 3.3: Impact of a parallel shift of the yield curve on the bond with five-year maturity

ΔR (in bps)	$\breve{P}_t$	ΔP_t	$\hat{P}_t$	ΔP_t	$S \times \Delta y$
−50	116.26	2.63	116.26	2.63	2.60
−30	115.20	1.57	115.20	1.57	1.56
−10	114.15	0.52	114.15	0.52	0.52
0	113.63	0.00	113.63	0.00	0.00
10	113.11	−0.52	113.11	−0.52	−0.52
30	112.08	−1.55	112.08	−1.55	−1.56
50	111.06	−2.57	111.06	−2.57	−2.60

Using the Nelson-Siegel yield curve, we report in Table 3.2 the price of the bond with maturity T (expressed in years) with a 5% annual coupon. For instance, the price of the four-year bond is calculated in the following way:

$$P_t = \frac{5}{(1+0.52\%)} + \frac{5}{(1+0.99\%)^2} + \frac{5}{(1+1.42\%)^3} + \frac{105}{(1+1.80\%)^4} = \$112.36$$

We also indicate the yield to maturity y (in %) and the corresponding sensitivity S. Let $\breve{P}_t$ (resp. $\hat{P}_t$) be the bond price by taking into account a parallel shift ΔR (in bps) directly on the zero-coupon rates (resp. on the yield to maturity). The results are given in Table 3.3 in the case of the bond with a five-year maturity[10]. We verify that the computation based on

[10] We have:

$$\breve{P}_t = \sum_{t_m \geq t} C(t_m)\, e^{-(t_m - t)(R_t(t_m) + \Delta R)} + N e^{-(T-t)(R_t(T) + \Delta R)}$$

and:

$$\hat{P}_t = \sum_{t_m \geq t} C(t_m)\, e^{-(t_m - t)(y + \Delta R)} + N e^{-(T-t)(y + \Delta R)}$$

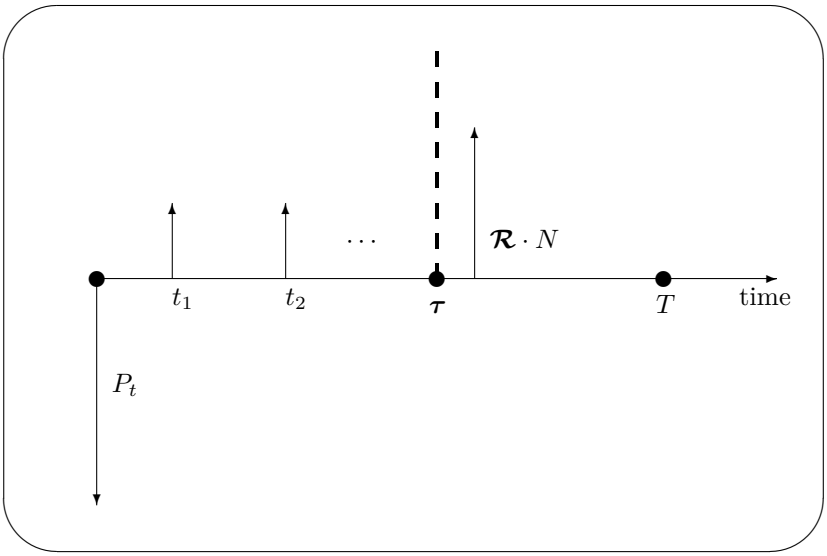


FIGURE 3.8: Cash flows of a bond with default risk

the sensitivity provides a good approximation. This method has been already used in the previous chapter on page 77 to calculate the value-at-risk of bonds.

With default risk In the previous paragraph, we assume that there is no default risk. However, if the issuer defaults at time $\boldsymbol{\tau}$ before the bond maturity T, some coupons and the notional are not paid. In this case, the buyer of the bond recovers part of the notional after the default time. An illustration is given in Figure 3.8. In terms of cash flows, we have therefore:

- the coupons $C(t_m)$ if the bond issuer does not default before the coupon date t_m:

$$\sum_{t_m \geq t} C(t_m) \cdot \mathbb{1}\{\boldsymbol{\tau} > t_m\}$$

- the notional if the bond issuer does not default before the maturity date:

$$N \cdot \mathbb{1}\{\boldsymbol{\tau} > T\}$$

- the recovery part if the bond issuer defaults before the maturity date:

$$\boldsymbol{\mathcal{R}} \cdot N \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\}$$

  where $\boldsymbol{\mathcal{R}}$ is the corresponding recovery rate.

If we assume that the recovery part is exactly paid at the default time $\boldsymbol{\tau}$, we deduce that the stochastic discounted value of the cash flow leg is:

$$\begin{aligned} SV_t &= \sum_{t_m \geq t} C(t_m) \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s} \cdot \mathbb{1}\{\boldsymbol{\tau} > t_m\} + \\ &\quad N \cdot e^{-\int_t^T r_s \, \mathrm{d}s} \cdot \mathbb{1}\{\boldsymbol{\tau} > T\} + \boldsymbol{\mathcal{R}} \cdot N \cdot e^{-\int_t^{\boldsymbol{\tau}} r_s \, \mathrm{d}s} \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\} \end{aligned}$$

The price of the bond is the expected value of the stochastic discounted value[11]: $P_t + AC_t = \mathbb{E}\left[SV_t \mid \mathcal{F}_t\right]$. If we assume that $(\mathcal{H}_1)$ the default time and the interest rates are independent and $(\mathcal{H}_2)$ the recovery rate is known and not stochastic, we obtain the following closed-form formula:

$$\begin{aligned} P_t + AC_t &= \sum_{t_m \geq t} C\left(t_m\right) B_t\left(t_m\right) \mathbf{S}_t\left(t_m\right) + NB_t\left(T\right) \mathbf{S}_t\left(T\right) + \\ &\quad \boldsymbol{\mathcal{R}} N \int_t^T B_t\left(u\right) f_t\left(u\right) \mathrm{d}u \end{aligned} \tag{3.3}$$

where $\mathbf{S}_t\left(u\right)$ is the survival function at time u and $f_t\left(u\right)$ the associated density function[12].

Remark 20 *If the issuer is not risky, we have* $\mathbf{S}_t\left(u\right) = 1$ *and* $f_t\left(u\right) = 0$. *In this case, Equation (3.3) reduces to Equation (3.2).*

Remark 21 *If we consider an exponential default time with parameter* λ – $\boldsymbol{\tau} \sim \mathcal{E}\left(\lambda\right)$, *we have* $\mathbf{S}_t\left(u\right) = e^{-\lambda(u-t)}$, $f_t\left(u\right) = \lambda e^{-\lambda(u-t)}$ *and:*

$$\begin{aligned} P_t + AC_t &= \sum_{t_m \geq t} C\left(t_m\right) B_t\left(t_m\right) e^{-\lambda(t_m - t)} + NB_t\left(T\right) e^{-\lambda(T-t)} + \\ &\quad \lambda \boldsymbol{\mathcal{R}} N \int_t^T B_t\left(u\right) e^{-\lambda(u-t)} \mathrm{d}u \end{aligned}$$

If we assume a flat yield curve – $R_t\left(u\right) = r$, *we obtain:*

$$\begin{aligned} P_t + AC_t &= \sum_{t_m \geq t} C\left(t_m\right) e^{-(r+\lambda)(t_m - t)} + Ne^{-(r+\lambda)(T-t)} + \\ &\quad \lambda \boldsymbol{\mathcal{R}} N \left(\frac{1 - e^{-(r+\lambda)(T-t)}}{r + \lambda}\right) \end{aligned}$$

Example 22 *We consider a bond with ten-year maturity. The notional is* \$100 *whereas the annual coupon rate is equal to 4.5%.*

If we consider that $r = 0$, the price of the non-risky bond is \$145. With $r = 5\%$, the price becomes \$95.19. Let us now take into account the default risk. We assume that the recovery rate $\boldsymbol{\mathcal{R}}$ is 40%. If $\lambda = 2\%$ (resp. 10%), the price of the risky bond is \$86.65 (resp. \$64.63). If the yield curve is not flat, we must use the general formula (3.3) to compute the price of the bond. In this case, the integral is evaluated with a numerical integration procedure, typically a Gauss-Legendre quadrature[13]. For instance, if we consider the yield curve defined in Example 21, the bond price is equal to \$110.13 if there is no default risk, \$99.91 if $\lambda = 2\%$ and \$73.34 if $\lambda = 10\%$.

The yield to maturity of the defaultable bond is computed exactly in the same way as without default risk. The credit spread $\boldsymbol{s}$ is then defined as the difference of the yield to maturity with default risk $\boldsymbol{y}$ and the yield to maturity without default risk $\boldsymbol{y}^\star$:

$$\boldsymbol{s} = \boldsymbol{y} - \boldsymbol{y}^\star \tag{3.4}$$

[11] It is also called the present value.

[12] We have:

$$\mathbf{S}_t\left(u\right) = \mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\tau} > u \mid \boldsymbol{\tau} > t\right\}\right] = \Pr\left\{\boldsymbol{\tau} > u \mid \boldsymbol{\tau} > t\right\}$$

The density function is then given by $f_t\left(u\right) = -\partial_u \mathbf{S}_t\left(u\right)$.

[13] See Appendix A.1.2.3 on page 1037 for a primer on numerical integration.

This spread is a credit risk measure and is an increasing function of the default risk. Reconsider the simple model with a flat yield curve and an exponential default time. If the recovery rate $\mathcal{R}$ is equal to zero, we deduce that the yield to maturity of the defaultable bond is $y = r + \lambda$. It follows that the credit spread is equal to the parameter λ of the exponential distribution. Moreover, if λ is relatively small (less than 20%), the annual probability of default is:

$$\text{PD} = \mathbf{S}_t(t+1) = 1 - e^{-\lambda} \approx \lambda$$

In this case, the credit spread is approximately equal to the annual default probability ($s \approx \text{PD}$).

If we reuse our previous example with the yield curve specified in Example 21, we obtain the results reported in Table 3.4. For instance, the yield to maturity of the bond is equal to 3.24% without default risk. If λ and $\mathcal{R}$ are set to 200 bps and 0%, the yield to maturity becomes 5.22% which implies a credit spread of 198.1 bps. If the recovery rate is higher, the credit spread decreases. Indeed, with λ equal to 200 bps, the credit spread is equal to 117.1 bps if $\mathcal{R} = 40\%$ and only 41.7 bps if $\mathcal{R} = 80\%$.

TABLE 3.4: Computation of the credit spread s

$\mathcal{R}$ (in %)	λ (in bps)	PD (in bps)	P_t (in \$)	y (in %)	s (in bps)
0	0	0.0	110.1	3.24	0.0
	10	10.0	109.2	3.34	9.9
	200	198.0	93.5	5.22	198.1
	1000	951.6	50.4	13.13	988.9
40	0	0.0	110.1	3.24	0.0
	10	10.0	109.6	3.30	6.0
	200	198.0	99.9	4.41	117.1
	1000	951.6	73.3	8.23	498.8
80	0	0.0	110.1	3.24	0.0
	10	10.0	109.9	3.26	2.2
	200	198.0	106.4	3.66	41.7
	1000	951.6	96.3	4.85	161.4

Remark 22 *In the case of loans, we do not calculate a capital requirement for market risk, only a capital requirement for credit risk. The reason is that there is no market price of the loan, because it cannot be traded in an exchange. For bonds, we calculate a capital requirement for both market and credit risks. In the case of the market risk, risk factors are the yield curve rates, but also the parameters associated to the credit risk, for instance the default probabilities and the recovery rate. In this context, market risk has a credit component. To illustrate this property, we consider the previous example and we assume that λ_t varies across time whereas the recovery rate $\mathcal{R}$ is equal to 40%. In Figure 3.9, we show the evolution of the process λ_t for the next 10 years (top panel) and the clean price*[14] *P_t (bottom/left panel). If we suppose now that the issuer defaults suddenly at time $t = 6.25$, we observe a jump in the clean price (bottom/right panel). It is obvious that the market risk takes into account the short-term evolution of the credit component (or the smooth part), but does not incorporate the jump risk (or the discontinuous part) and also the large uncertainty on the recovery price. This is why these risks are covered by credit risk capital requirements.*

[14] We assume that the yield curve remains constant.

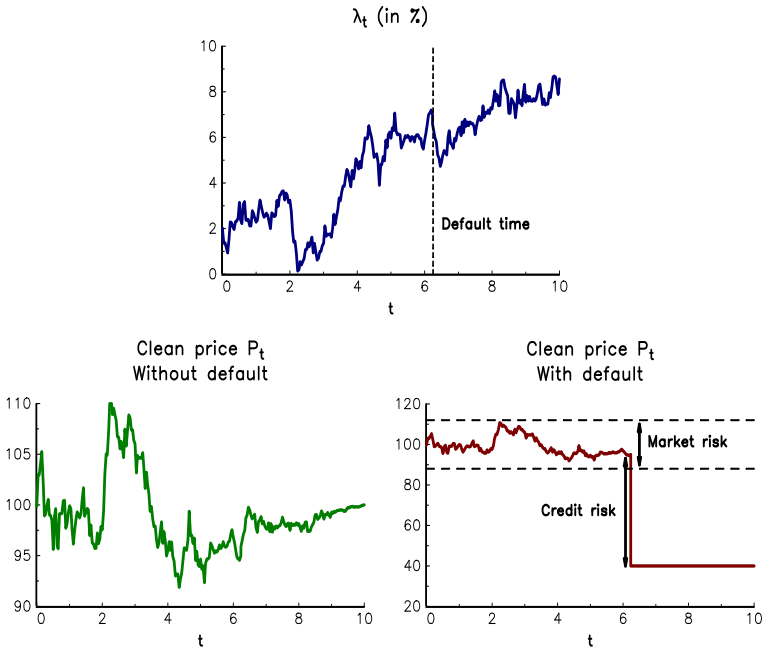


FIGURE 3.9: Difference between market and credit risks for a bond

3.1.3 Securitization and credit derivatives

Since the 1990s, banks have developed credit transfer instruments in two directions: credit securitization and credit derivatives. The term securitization refers to the process of transforming illiquid and non-tradable assets into tradable securities. Credit derivatives are financial instruments whose payoff explicitly depends on credit events like the default of an issuer. These two topics are highly connected because credit securities can be used as underlying assets of credit derivatives.

3.1.3.1 Credit securitization

According to AFME (2019), outstanding amount of securitization is close to €9 tn. Figure 3.10 shows the evolution of issuance in Europe and US since 2000. We observe that the financial crisis had a negative impact on the growth of credit securitization, especially in Europe that represents less than 20% of this market. This market is therefore dominated by the US, followed by UK, France, Spain, the Netherlands and Germany.

Credit securities are better known as asset-backed securities (ABS), even if this term is generally reserved to assets that are not mortgage, loans or corporate bonds. In its simplest form, an ABS is a bond whose coupons are derived from a collateral pool of assets. We generally make the following distinction with respect to the type of collateral assets:

- Mortgage-backed securities (MBS)
  - Residential mortgage-backed securities (RMBS)
  - Commercial mortgage-backed securities (CMBS)
- Collateralized debt obligations (CDO)
  - Collateralized loan obligations (CLO)

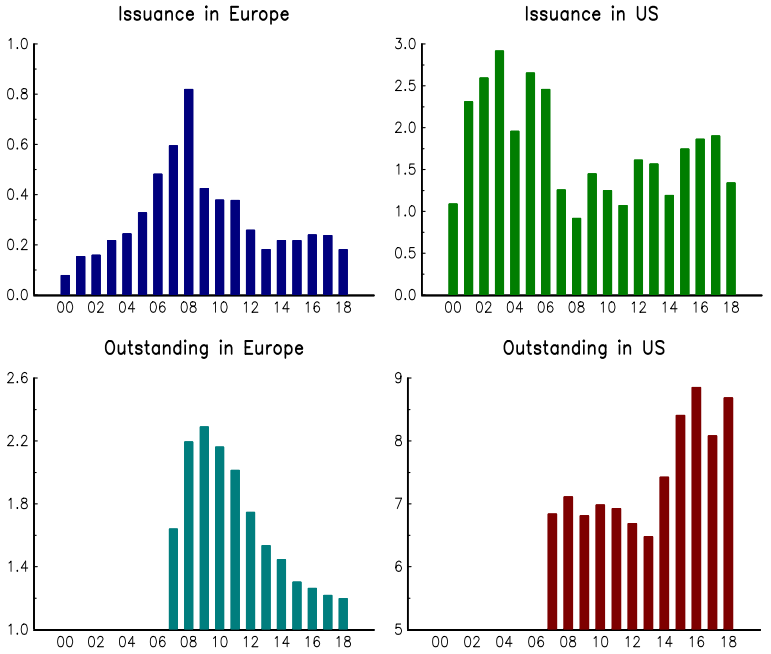


FIGURE 3.10: Securitization in Europe and US (in € tn)

Source: Association for Financial Markets in Europe (2019).

  - Collateralized bond obligations (CBO)

- Asset-backed securities (ABS)
  - Auto loans
  - Credit cards and revolving credit
  - Student loans

MBS are securities that are backed by residential and commercial mortgage loans. The most basic structure is a pass-through security, where the coupons are the same for all the investors and are proportional to the revenue of the collateral pool. Such structure is shown in Figure 3.11. The originator (e.g. a bank) sells a pool of debt to a special purpose vehicle (SPV). The SPV is an ad-hoc legal entity[15] whose sole function is to hold the loans as assets and issue the securities for investors. In the pass-through structure, the securities are all the same and the cash flows paid to investors are directly proportional to interests and principals of collateral assets. More complex structures are possible with several classes of bonds (see Figure 3.12). In this case, the cash flows differ from one type of securities to another one. The most famous example is the collateralized debt obligation, where the securities are divided into tranches. This category includes also collateralized mortgage obligations (CMO), which are both MBS and CDO. The two other categories of CDOs are CLOs, which are backed by corporate bank debt (e.g. SME loans) and CBOs, which are backed by bonds (e.g. high yield bonds). Finally, pure ABS principally concerns consumer credit such as auto loans, credit cards and student loans.

[15]It may be a subsidiary of the originator.

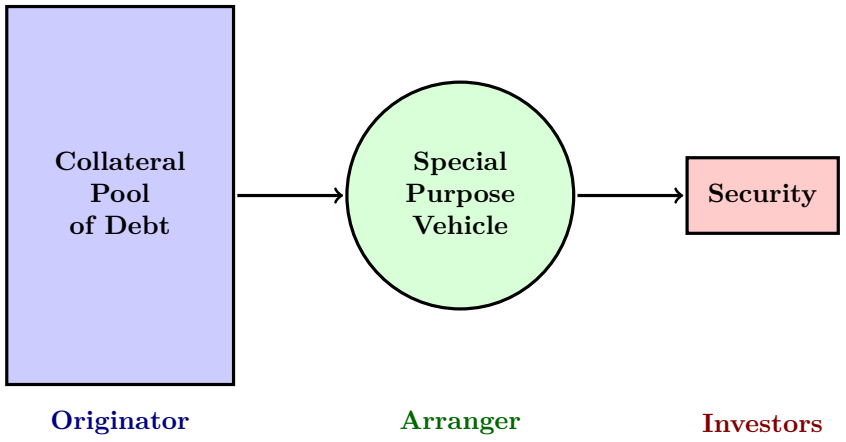


FIGURE 3.11: Structure of pass-through securities

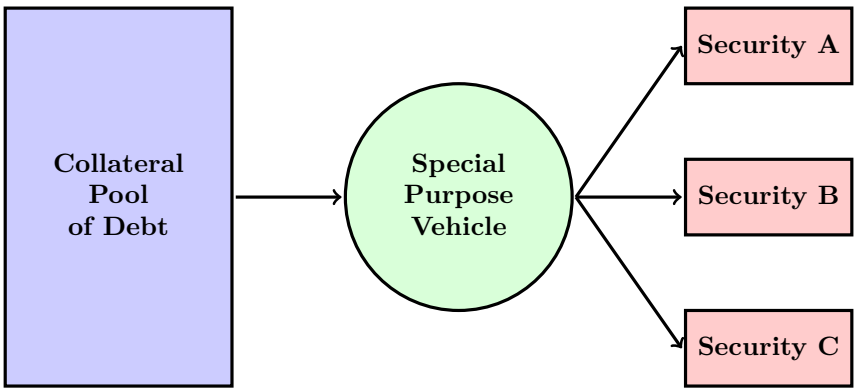


FIGURE 3.12: Structure of pay-through securities

In Table 3.5, we report some statistics about US mortgage-backed securities. SIFMA (2019b) makes the distinction between agency MBS and non-agency MBS. After the Great Depression, the US government created three public entities to promote home ownership and provide insurance of mortgage loans: the Federal National Mortgage Association (FNMA or Fannie Mae), the Federal Home Loan Mortgage Corporation (FHLMC or Freddie Mac) and the Government National Mortgage Association (GNMA or Ginnie Mae). Agency MBS refer to securities guaranteed by these three public entities and represent the main part of the US MBS market. This is especially true since the 2008 financial crisis. Indeed, non-agency MBS represent 53.5% of the issuance in 2006 and only 3.5% in 2012. Because agency MBS are principally based on home mortgage loans, the RMBS market is ten times more larger than the CMBS market. CDO and ABS markets are smaller and represent together about \$1.5 tn (see Table 3.6). The CDO market strongly suffered from the subprime crisis[16]. During the same period, the structure of the ABS market changed with an increasing proportion of ABS backed by auto loans and a fall of ABS backed by credit cards and student loans.

Remark 23 *Even if credit securities may be viewed as bonds, their pricing is not straightforward. Indeed, the measure of the default probability and the recovery depends on the*

[16]For instance, the issuance of US CDO was less than \$10 bn in 2010.

TABLE 3.5: US mortgage-backed securities

Year	Agency		Non-agency		Total
	MBS	CMO	CMBS	RMBS	(in $ bn)
			Issuance		
2002	57.5%	23.6%	2.2%	16.7%	2 515
2006	33.6%	11.0%	7.9%	47.5%	2 691
2008	84.2%	10.8%	1.2%	3.8%	1 394
2010	71.0%	24.5%	1.2%	3.3%	2 013
2012	80.1%	16.4%	2.2%	1.3%	2 195
2014	68.7%	19.2%	7.0%	5.1%	1 440
2016	76.3%	15.7%	3.8%	4.2%	2 044
2018	69.2%	16.6%	4.7%	9.5%	1 899
			Outstanding amount		
2002	59.7%	17.4%	5.6%	17.2%	5 289
2006	45.7%	14.9%	8.3%	31.0%	8 390
2008	52.4%	14.0%	8.8%	24.9%	9 467
2010	59.2%	14.6%	8.1%	18.1%	9 258
2012	64.0%	14.8%	7.2%	14.0%	8 838
2014	68.0%	13.7%	7.1%	11.2%	8 842
2016	72.4%	12.3%	5.9%	9.5%	9 023
2018	74.7%	11.3%	5.6%	8.4%	9 732

Source: Securities Industry and Financial Markets Association (2019b,c) and author's calculations.

TABLE 3.6: US asset-backed securities

Year	Auto Loans	CDO & CLO	Credit Cards	Equip-ement	Other	Student Loans	Total (in $ bn)
				Issuance			
2002	34.9%	21.0%	25.2%	2.6%	6.8%	9.5%	269
2006	13.5%	60.1%	9.3%	2.2%	4.6%	10.3%	658
2008	16.5%	37.8%	25.9%	1.3%	5.4%	13.1%	215
2010	46.9%	6.4%	5.2%	7.0%	22.3%	12.3%	126
2012	33.9%	23.1%	12.5%	7.1%	13.7%	9.8%	259
2014	25.2%	35.6%	13.1%	5.2%	17.0%	4.0%	393
2016	28.3%	36.8%	8.3%	4.6%	16.9%	5.1%	325
2018	20.8%	54.3%	6.1%	5.1%	10.1%	3.7%	517
				Outstanding amount			
2002	20.7%	28.6%	32.5%	4.1%	7.5%	6.6%	905
2006	11.8%	49.3%	17.6%	3.1%	6.0%	12.1%	1 657
2008	7.7%	53.5%	17.3%	2.4%	6.2%	13.0%	1 830
2010	7.6%	52.4%	14.4%	2.4%	7.1%	16.1%	1 508
2012	11.0%	48.7%	10.0%	3.3%	8.7%	18.4%	1 280
2014	13.2%	46.8%	10.1%	3.9%	9.8%	16.2%	1 349
2016	13.9%	48.0%	9.3%	3.7%	11.6%	13.5%	1 397
2018	13.3%	48.2%	7.4%	5.0%	16.0%	10.2%	1 677

Source: Securities Industry and Financial Markets Association (2019b,c) and author's calculations.

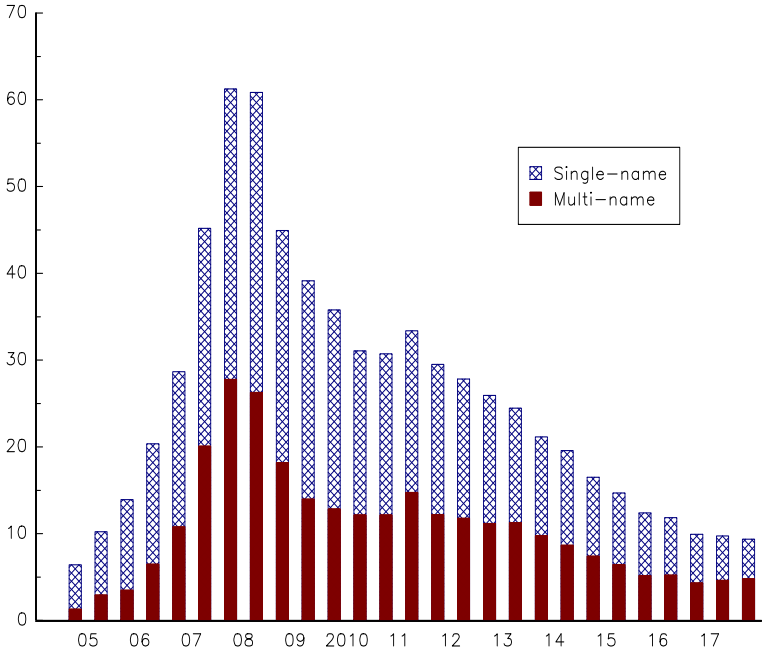


FIGURE 3.13: Outstanding amount of credit default swaps (in $ tn)

Source: Bank for International Settlements (2019).

characteristics of the collateral assets (individual default probabilities and recovery rates), but also on the correlation between these risk factors. Measuring credit risk of such securities is then a challenge. Another issue concerns design and liquidity problems faced when packaging and investing in these assets[17] *(Duffie and Rahi, 1995; DeMarzo and Duffie, 1999). This explains that credit securities suffered a lot during the 2008 financial crisis, even if some of them were not linked to subprime mortgages. In fact, securitization markets pose a potential risk to financial stability (Segoviano et al., 2013). This is a topic we will return to in Chapter 8, which deals with systemic risk.*

3.1.3.2 Credit default swap

A credit default swap (CDS) may be defined as an insurance derivative, whose goal is to transfer the credit risk from one party to another. In a standard contract, the protection buyer makes periodic payments (known as the premium leg) to the protection seller. In return, the protection seller pays a compensation (known as the default leg) to the protection buyer in the case of a credit event, which can be a bankruptcy, a failure to pay or a debt restructuring. In its most basic form, the credit event refers to an issuer (sovereign or corporate) and this corresponds to a single-name CDS. If the credit event relates to a universe of different entities, we speak about a multi-name CDS. In Figure 3.13, we report the evolution of outstanding amount of CDS since 2007. The growth of this market was very strong before 2008 with a peak close to $60 tn. The situation today is different, because the market of single-name CDS stabilized whereas the market of basket default swaps continues to fall significantly. Nevertheless, it remains an important OTC market with a total outstanding around $9 tn.

[17] The liquidity issue is treated in Chapter 6.

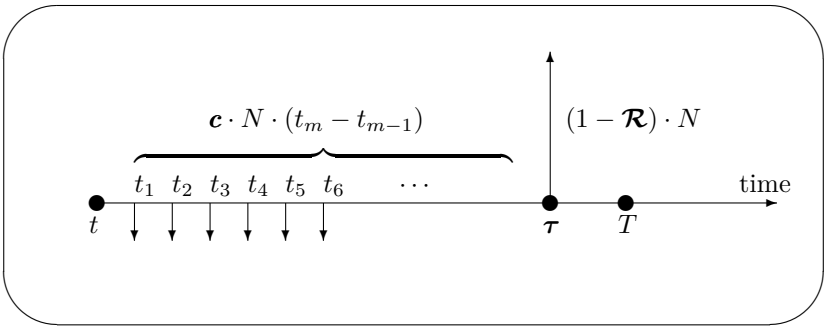


FIGURE 3.14: Cash flows of a single-name credit default swap

In Figure 3.14, we report the mechanisms of a single-name CDS. The contract is defined by a reference entity (the name), a notional principal N, a maturity or tenor T, a payment frequency, a recovery rate $\boldsymbol{\mathcal{R}}$ and a coupon rate[18] $\boldsymbol{c}$. From the inception date t to the maturity date T or the default time $\boldsymbol{\tau}$, the protection buyer pays a fixed payment, which is equal to $\boldsymbol{c} \cdot N \cdot \Delta t_m$ at the fixing date t_m with $\Delta t_m = t_m - t_{m-1}$. This means that the annual premium leg is equal to $\boldsymbol{c} \cdot N$. If there is no credit event, the protection buyer will also pay a total of $\boldsymbol{c} \cdot N \cdot (T - t)$. In case of credit event before the maturity, the protection seller will compensate the protection buyer and will pay $(1 - \boldsymbol{\mathcal{R}}) \cdot N$.

Example 23 *We consider a credit default swap, whose notional principal is* \$10 *mn, maturity is* 5 *years and payment frequency is quarterly. The credit event is the bankruptcy of the corporate entity A. We assume that the recovery rate is set to* 40% *and the coupon rate is equal to* 2%*.*

Because the payment frequency is quarterly, there are 20 fixing dates, which are 3M, 6M, 9M, 1Y, ..., 5Y. Each quarter, if the corporate A does not default, the protection buyer pays a premium, which is approximately equal to $\$10\text{mn} \times 2\% \times 0.25 = \$50\,000$. If there is no default during the next five years, the protection buyer will pay a total of $\$50\,000 \times 20 = \$1$ mn whereas the protection seller will pay nothing. Suppose now that the corporate defaults two years and four months after the CDS inception date. In this case, the protection buyer will pay \$50 000 during 9 quarters and will receive the protection leg from the protection seller at the default time. This protection leg is equal to $(1 - 40\%) \times \$10\,\text{mn} = \$6\,\text{mn}$.

To compute the mark-to-market value of a CDS, we use the reduced-form approach as in the case of bond pricing. If we assume that the premium is not paid after the default time $\boldsymbol{\tau}$, the stochastic discounted value of the premium leg is[19]:

$$SV_t(\mathcal{PL}) = \sum_{t_m \geq t} \boldsymbol{c} \cdot N \cdot (t_m - t_{m-1}) \cdot \mathbb{1}\{\boldsymbol{\tau} > t_m\} \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s}$$

[18] We will see that the coupon rate $\boldsymbol{c}$ is in fact the CDS spread $\boldsymbol{s}$ for par swaps.

[19] In order to obtain a simple formula, we do not deal with the accrued premium (see Remark 26 on page 149).

Using the standard assumptions that the default time is independent of interest rates and the recovery rate, we deduce the present value of the premium leg as follows:

$$\begin{aligned} PV_t(\mathcal{PL}) &= \mathbb{E}\left[\sum_{t_m \geq t} \boldsymbol{c} \cdot N \cdot \Delta t_m \cdot \mathbb{1}\{\boldsymbol{\tau} > t_m\} \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s} \middle| \mathcal{F}_t\right] \\ &= \sum_{t_m \geq t} \boldsymbol{c} \cdot N \cdot \Delta t_m \cdot \mathbb{E}[\mathbb{1}\{\boldsymbol{\tau} > t_m\}] \cdot \mathbb{E}\left[e^{-\int_t^{t_m} r_s \, \mathrm{d}s}\right] \\ &= \boldsymbol{c} \cdot N \cdot \sum_{t_m \geq t} \Delta t_m \mathbf{S}_t(t_m) B_t(t_m) \end{aligned}$$

where $\mathbf{S}_t(u)$ is the survival function at time u. If we assume that the default leg is exactly paid at the default time $\boldsymbol{\tau}$, the stochastic discount value of the default (or protection) leg is[20]:

$$SV_t(\mathcal{DL}) = (1 - \boldsymbol{\mathcal{R}}) \cdot N \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\} \cdot e^{-\int_t^{\boldsymbol{\tau}} r(s) \, \mathrm{d}s}$$

It follows that its present value is:

$$\begin{aligned} PV_t(\mathcal{DL}) &= \mathbb{E}\left[(1 - \boldsymbol{\mathcal{R}}) \cdot N \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\} \cdot e^{-\int_t^{\boldsymbol{\tau}} r_s \, \mathrm{d}s} \middle| \mathcal{F}_t\right] \\ &= (1 - \boldsymbol{\mathcal{R}}) \cdot N \cdot \mathbb{E}[\mathbb{1}\{\boldsymbol{\tau} \leq T\} \cdot B_t(\boldsymbol{\tau})] \\ &= (1 - \boldsymbol{\mathcal{R}}) \cdot N \cdot \int_t^T B_t(u) f_t(u) \, \mathrm{d}u \end{aligned}$$

where $f_t(u)$ is the density function associated to the survival function $\mathbf{S}_t(u)$. We deduce that the mark-to-market of the swap is[21]:

$$\begin{aligned} P_t(T) &= PV_t(\mathcal{DL}) - PV_t(\mathcal{PL}) \\ &= (1 - \boldsymbol{\mathcal{R}}) N \int_t^T B_t(u) f_t(u) \, \mathrm{d}u - \boldsymbol{c} N \sum_{t_m \geq t} \Delta t_m \mathbf{S}_t(t_m) B_t(t_m) \\ &= N\left((1 - \boldsymbol{\mathcal{R}}) \int_t^T B_t(u) f_t(u) \, \mathrm{d}u - \boldsymbol{c} \cdot \mathrm{RPV}_{01}\right) \end{aligned} \tag{3.5}$$

where $\mathrm{RPV}_{01} = \sum_{t_m \geq t} \Delta t_m \mathbf{S}_t(t_m) B_t(t_m)$ is called the risky PV01 and corresponds to the present value of 1 bp paid on the premium leg. The CDS price is then inversely related to the spread. At the inception date, the present value of the premium leg is equal to the present value of the default leg meaning that the CDS spread corresponds to the coupon rate such that $P_t^{\text{buyer}} = 0$. We obtain the following expression:

$$\boldsymbol{s} = \frac{(1 - \boldsymbol{\mathcal{R}}) \int_t^T B_t(u) f_t(u) \, \mathrm{d}u}{\sum_{t_m \geq t} \Delta t_m \mathbf{S}_t(t_m) B_t(t_m)} \tag{3.6}$$

The spread $\boldsymbol{s}$ is in fact the fair value coupon rate $\boldsymbol{c}$ in such a way that the initial value of the credit default swap is equal to zero.

[20] Here the recovery rate $\boldsymbol{\mathcal{R}}$ is assumed to be deterministic.

[21] P_t is the swap price for the protection buyer. We have then $P_t^{\text{buyer}}(T) = P_t(T)$ and $P_t^{\text{seller}}(T) = -P_t(T)$.

We notice that if there is no default risk, this implies that $\mathbf{S}_t(u) = 1$ and we get $\mathcal{S} = 0$. In the same way, the spread is also equal to zero if the recovery rate is set to 100%. If we assume that the premium leg is paid continuously, the formula (3.6) becomes:

$$\mathcal{S} = \frac{(1-\mathcal{R}) \int_t^T B_t(u) f_t(u) \, \mathrm{d}u}{\int_t^T B_t(u) \mathbf{S}_t(u) \, \mathrm{d}u}$$

If the interest rates are equal to zero ($B_t(u) = 1$) and the default times is exponential with parameter λ – $\mathbf{S}_t(u) = e^{-\lambda(u-t)}$ and $f_t(u) = \lambda e^{-\lambda(u-t)}$, we get:

$$\begin{aligned} \mathcal{S} &= \frac{(1-\mathcal{R}) \cdot \lambda \cdot \int_t^T e^{-\lambda(u-t)} \, \mathrm{d}u}{\int_t^T e^{-\lambda(u-t)} \, \mathrm{d}u} \\ &= (1-\mathcal{R}) \cdot \lambda \end{aligned}$$

If λ is relatively small, we also notice that this relationship can be written as follows:

$$\mathcal{S} \approx (1-\mathcal{R}) \cdot \mathrm{PD}$$

where PD is the one-year default probability[22]. This relationship is known as the '*credit triangle*' because it is a relationship between three variables where knowledge of any two is sufficient to calculate the third (O'Kane, 2008). It basically states that the CDS spread is approximatively equal to the one-year loss. The spread contains also the same information than the survival function and is an increasing function of the default probability. It can then be interpreted as a credit risk measure of the reference entity.

We recall that the first CDS was traded by J.P. Morgan in 1994 (Augustin *et al.*, 2014). The CDS market structure has been organized since then, especially the standardization of the CDS contract. Today, CDS agreements are governed by 2003 and 2014 ISDA credit derivatives definitions. For instance, the settlement of the CDS contract can be either physical or in cash. In the case of cash settlement, there is a monetary exchange from the protection seller to the protection buyer[23]. In the case of physical settlement, the protection buyer delivers a bond to the protection seller and receives the notional principal amount. Because the price of the defaulted bond is equal to $\mathcal{R} \cdot N$, this means that the implied mark-to-market of this operation is $N - \mathcal{R} \cdot N$ or equivalently $(1-\mathcal{R}) \cdot N$. Or course, physical settlement is only possible if the reference entity is a bond or if the credit event is based on the bond default. Whereas physical settlement was prevailing in the 1990s, most of the settlements are today in cash. Another standardization concerns the price of CDS. With the exception of very specific cases[24], CDS contracts are quoted in (fair) spread expressed in bps. In Figures 3.15 and 3.16, we show the evolution of some CDS spreads for a five-year maturity. We notice the increase of credit spreads since the 2008 financial turmoil and the

[22] We have:

$$\begin{aligned} \mathrm{PD} &= \Pr\{\tau \leq t+1 \mid \tau \leq t\} \\ &= 1 - \mathbf{S}_t(t+1) \\ &= 1 - e^{-\lambda} \\ &\simeq \lambda \end{aligned}$$

For instance, if λ is equal respectively to 1%, 5%, 10% and 20% , the one-year default probability takes the values 1.00%, 4.88%, 9.52% and 18.13%.

[23] This monetary exchange is equal to $(1-\mathcal{R}) \cdot N$.

[24] When the default probability is high (larger than 20%), CDS contracts can be quoted with an upfront meaning that the protection seller is asking an initial amount to enter into the swap. For instance, it was the case of CDS on Greece in spring 2013.

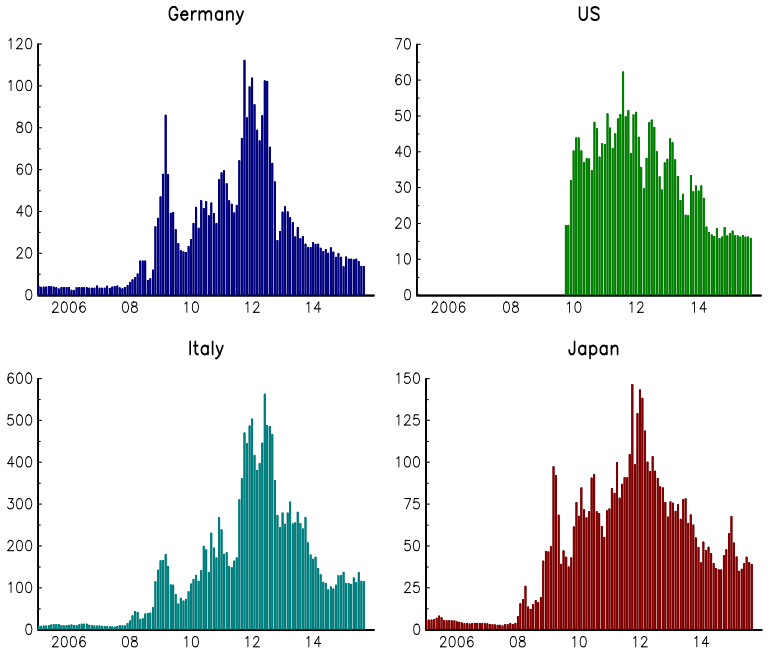


FIGURE 3.15: Evolution of some sovereign CDS spreads

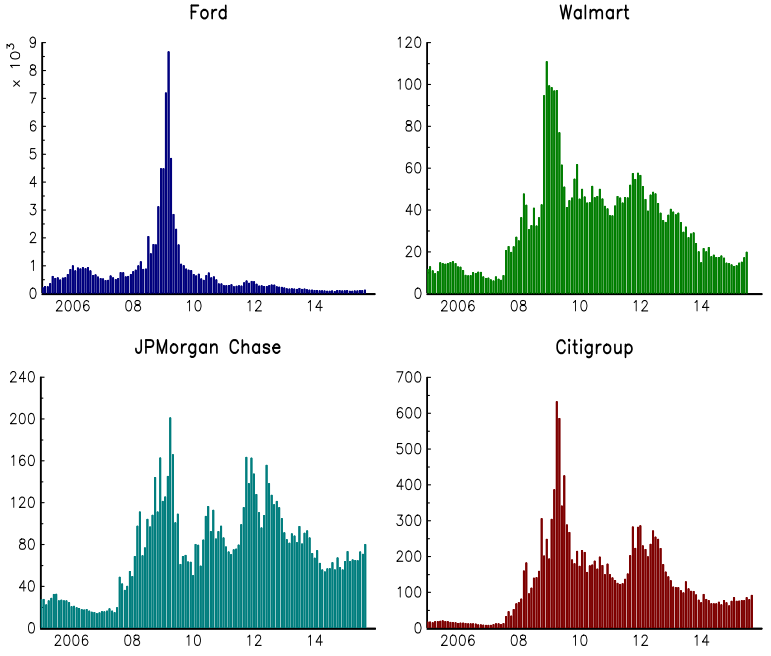


FIGURE 3.16: Evolution of some financial and corporate CDS spreads

default of Lehman Brothers bankruptcy, the sensitivity of German and Italian spreads with respect to the Eurozone crisis and also the difference in level between the different countries. Indeed, the spread is globally lower for US than for Germany or Japan. In the case of Italy, the spread is high and has reached 600 bps in 2012. We observe that the spread of some corporate entities may be lower than the spread of many developed countries (see Figure 3.16). This is the case of Walmart, whose spread is lower than 20 bps since 2014. When a company (or a country) is in great difficulty, the CDS spread explodes as in the case of Ford in February 2009. CDS spreads can be used to compare the default risk of two entities in the same sector. For instance, Figure 3.16 shows than the default risk of Citigroup is higher that this of JPMorgan Chase.

The CDS spread changes over time, but depends also on the maturity or tenor. This implies that we have a term structure of credit spreads for a given date t. This term structure is known as the credit spread curve and is noted $\mathcal{s}_t(T)$ where T is the maturity time. Figure 3.17 shows the credit curve for different entities as of 17 September 2015. We notice that the CDS spread increases with the maturity. This is the most common case for investment grade (IG) entities, whose short-term default risk is low, but long-term default risk is higher. Nevertheless, we observe some distinguishing patterns between these credit curves. For instance, the credit risk of Germany is lower than the credit risk of US if the maturity is less than five years, but it is higher in the long run. There is a difference of 4 bps between Google and Apple on average when the time-to-maturity is less than 5 years. In the case of 10Y CDS, the spread of Apple is 90.8 bps whereas it is only 45.75 bps for Google.

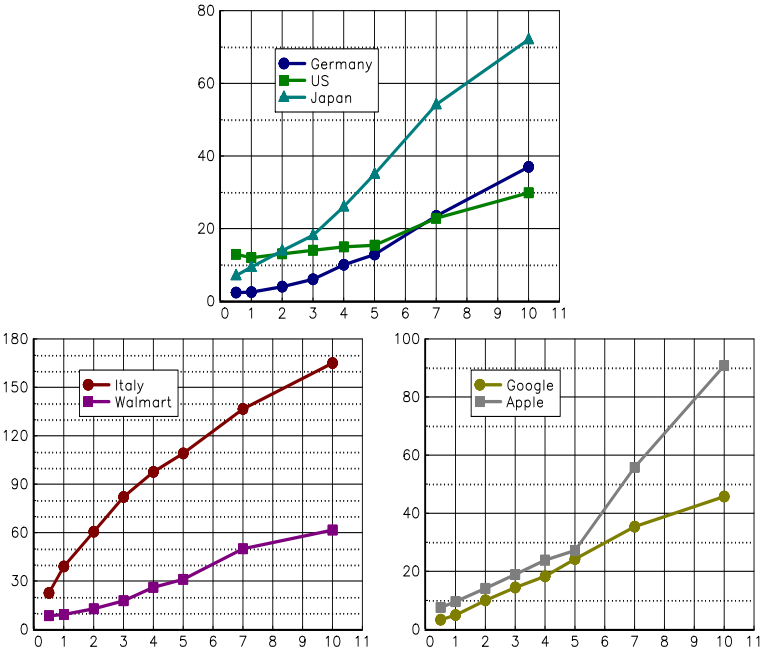


FIGURE 3.17: Example of CDS spread curves as of 17 September 2015

Remark 24 *In other cases, the credit curve may be decreasing (for some high yield corporates) or have a complex curvature (bell-shaped or U-shaped). In fact, Longstaff et al. (2005) showed that the dynamics of credit default swaps also depends on the liquidity risk. For instance, the most liquid CDS contract is generally the 5Y CDS. The liquidity on the other maturities depends on the reference entity and other characteristics such as the bond*

market liquidity. For example, the liquidity may be higher for short maturities when the credit risk of the reference entity is very high.

Initially, CDS were used to hedge the credit risk of corporate bonds by banks and insurance companies. This hedging mechanism is illustrated in Figure 3.18. We assume that the bond holder buys a protection using a CDS, whose fixing dates of the premium leg are exactly the same as the coupon dates of the bond. We also assume that the credit even is the bond default and the notional of the CDS is equal to the notional of the bond. At each fixing date t_m, the bond holder receives the coupon $C(t_m)$ of the bond and pays to the protection seller the premium $\mathcal{s} \cdot N$. This implies that the net cash flow is $C(t_m) - \mathcal{s} \cdot N$. If the default occurs, the value of the bond becomes $\mathcal{R} \cdot N$, but the protection seller pays to the bond holder the default leg $(1 - \mathcal{R}) \cdot N$. In case of default, the net cash flow is then equal to $\mathcal{R} \cdot N + (1 - \mathcal{R}) \cdot N = N$, meaning that the exposure on the defaultable bond is perfectly hedged. We deduce that the annualized return R of this hedged portfolio is the difference between the yield to maturity $\mathcal{y}$ of the bond and the annual cost $\mathcal{s}$ of the protection:

$$R = \mathcal{y} - \mathcal{s} \tag{3.7}$$

We recognize a new formulation of Equation (3.4) on page 135. In theory, R is then equal to the yield to maturity $\mathcal{y}^\star$ of the bond without credit risk.

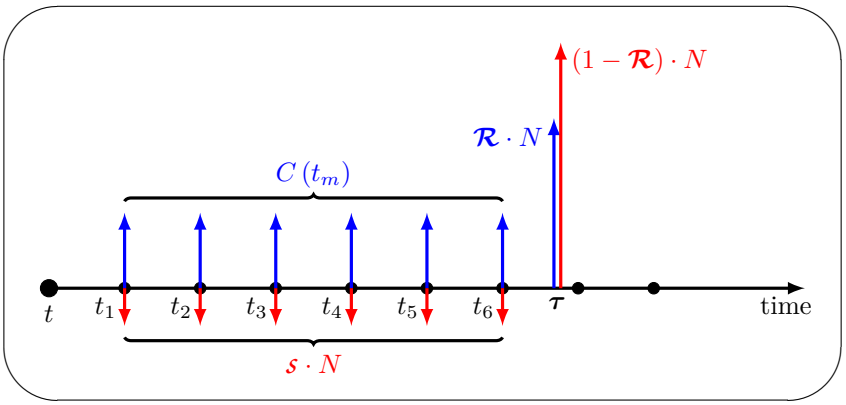


FIGURE 3.18: Hedging a defaultable bond with a credit default swap

Since the 2000s, end-users of CDS are banks and securities firms, insurance firms including pension funds, hedge funds and mutual funds. They continue to be used as hedging instruments, but they have also become financial instruments to express views about credit risk. In this case, '*long credit*' refers to the position of the protection seller who is exposed to the credit risk, whereas '*short credit*' is the position of the protection buyer who sold the credit risk of the reference entity[25]. To understand the mark-to-market of such positions, we consider the initial position at the inception date t of the CDS contract. In this case, the CDS spread $\mathcal{s}_t(T)$ verifies that the face value of the swap is equal to zero. Let us introduce the notation $P_{t,t'}(T)$, which defines the mark-to-market of a CDS position whose inception date is t, valuation date is t' and maturity date is T. We have:

$$P_{t,t}^{\text{seller}}(T) = P_{t,t}^{\text{buyer}}(T) = 0$$

[25] Said differently, a long exposure implies that the default results in a loss, whereas a short exposure implies that the default results in a gain.

At date $t' > t$, the mark-to-market price of the CDS is:

$$P_{t,t'}^{\text{buyer}}(T) = N\left((1-\boldsymbol{\mathcal{R}})\int_{t'}^{T} B_{t'}(u)\, f_{t'}(u)\, \mathrm{d}u - \boldsymbol{\mathcal{s}}_t(T)\cdot \text{RPV}_{01}\right)$$

whereas the value of the CDS spread satisfies the following relationship:

$$P_{t',t'}^{\text{buyer}}(T) = N\left((1-\boldsymbol{\mathcal{R}})\int_{t'}^{T} B_{t'}(u)\, f_{t'}(u)\, \mathrm{d}u - \boldsymbol{\mathcal{s}}_{t'}(T)\cdot \text{RPV}_{01}\right) = 0$$

We deduce that the P&L of the protection buyer is:

$$\Pi^{\text{buyer}} = P_{t,t'}^{\text{buyer}}(T) - P_{t,t}^{\text{buyer}}(T) = P_{t,t'}^{\text{buyer}}(T)$$

Using Equation (3.8), we know that $P_{t',t'}^{\text{buyer}}(T) = 0$ and we obtain:

$$\begin{aligned}
\Pi^{\text{buyer}} &= P_{t,t'}^{\text{buyer}}(T) - P_{t',t'}^{\text{buyer}}(T) \\
&= N\left((1-\boldsymbol{\mathcal{R}})\int_{t'}^{T} B_{t'}(u)\, f_{t'}(u)\, \mathrm{d}u - \boldsymbol{\mathcal{s}}_t(T)\cdot \text{RPV}_{01}\right) - \\
&\quad N\left((1-\boldsymbol{\mathcal{R}})\int_{t'}^{T} B_{t'}(u)\, f_{t'}(u)\, \mathrm{d}u - \boldsymbol{\mathcal{s}}_{t'}(T)\cdot \text{RPV}_{01}\right) \\
&= N\cdot(\boldsymbol{\mathcal{s}}_{t'}(T) - \boldsymbol{\mathcal{s}}_t(T))\cdot \text{RPV}_{01}
\end{aligned} \tag{3.8}$$

This equation highlights the role of the term RPV_{01} when calculating the P&L of the CDS position. Because $\Pi^{\text{seller}} = -\Pi^{\text{buyer}}$, we distinguish two cases:

- If $\boldsymbol{\mathcal{s}}_{t'}(T) > \boldsymbol{\mathcal{s}}_t(T)$, the protection buyer makes a profit, because this short credit exposure has benefited from the increase of the default risk.
- If $\boldsymbol{\mathcal{s}}_{t'}(T) < \boldsymbol{\mathcal{s}}_t(T)$, the protection seller makes a profit, because the default risk of the reference entity has decreased.

Suppose that we are in the first case. To realize its P&L, the protection buyer has three options (O'Kane, 2008):

1. He could unwind the CDS exposure with the protection seller if the latter agrees. This implies that the protection seller pays the mark-to-market $P_{t,t'}^{\text{buyer}}(T)$ to the protection buyer.
2. He could hedge the mark-to-market value by selling a CDS on the same reference entity and the same maturity. In this situation, he continues to pay the spread $\boldsymbol{\mathcal{s}}_t(T)$, but he now receives a premium, whose spread is equal to $\boldsymbol{\mathcal{s}}_{t'}(T)$.
3. He could reassign the CDS contract to another counterparty as illustrated in Figure 3.19. The new counterparty (the protection buyer C in our case) will then pay the coupon rate $\boldsymbol{\mathcal{s}}_t(T)$ to the protection seller. However, the spread is $\boldsymbol{\mathcal{s}}_{t'}(T)$ at time t', which is higher than $\boldsymbol{\mathcal{s}}_t(T)$. This is why the new counterparty also pays the mark-to-market $P_{t,t'}^{\text{buyer}}(T)$ to the initial protection buyer.

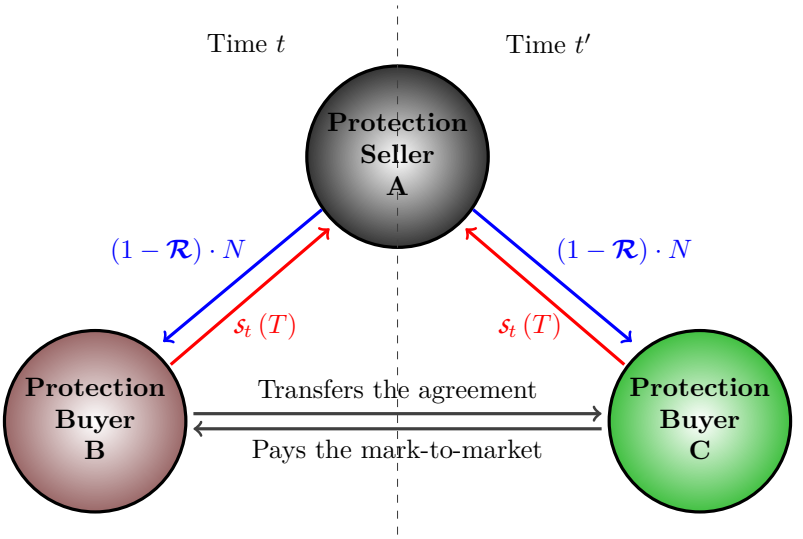


FIGURE 3.19: An example of CDS offsetting

Remark 25 *When the default risk is very high, CDS are quoted with an upfront*[26]*. In this case, the annual premium leg is equal to* $\boldsymbol{c}^\star \cdot N$ *where* $\boldsymbol{c}^\star$ *is a standard value*[27]*, and the protection buyer has to pay an upfront* UF_t *to the protection seller defined as follows:*

$$\mathrm{UF}_t = N\left((1-\boldsymbol{\mathcal{R}})\int_t^T B_t(u)\, f_t(u)\,\mathrm{d}u - \boldsymbol{c}^\star \cdot \mathrm{RPV}_{01}\right)$$

Remark 26 *Until now, we have simplified the pricing of the premium leg in order to avoid complicated calculations. Indeed, if the default occurs between two fixing dates, the protection buyer has to pay the premium accrual. For instance, if* $\boldsymbol{\tau} \in]t_{m-1}, t_m[$*, the accrued premium is equal to* $\boldsymbol{c}\cdot N\cdot(\boldsymbol{\tau}-t_{m-1})$ *or equivalently to:*

$$\mathcal{AP} = \sum_{t_m \geq t} \boldsymbol{c}\cdot N\cdot(\boldsymbol{\tau}-t_{m-1})\cdot \mathbb{1}\{t_{m-1}\leq \boldsymbol{\tau}\leq t_m\}$$

We deduce that the stochastic discount value of the accrued premium is:

$$SV_t(\mathcal{AP}) = \sum_{t_m \geq t} \boldsymbol{c}\cdot N\cdot(\boldsymbol{\tau}-t_{m-1})\cdot \mathbb{1}\{t_{m-1}\leq \boldsymbol{\tau}\leq t_m\}\cdot e^{-\int_t^{\boldsymbol{\tau}} r_s\,\mathrm{d}s}$$

It follows that:

$$PV_t(\mathcal{AP}) = \boldsymbol{c}\cdot N\cdot \sum_{t_m\geq t}\int_{t_{m-1}}^{t_m}(u-t_{m-1})\, B_t(u)\, f_t(u)\,\mathrm{d}u$$

All the previous formulas remain valid by replacing the expression of the risky PV01 by the following term:

$$\mathrm{RPV}_{01} = \sum_{t_m\geq t}\left(\Delta t_m \mathbf{S}_t(t_m)\, B_t(t_m) + \int_{t_{m-1}}^{t_m}(u-t_{m-1})\, B_t(u)\, f_t(u)\,\mathrm{d}u\right) \tag{3.9}$$

[26] It was the case several times for CDS on Greece.
[27] For distressed names, the default coupon rate $\boldsymbol{c}^\star$ is typically equal to 500 bps.

Example 24 *We assume that the yield curve is generated by the Nelson-Siegel model with the following parameters:* $\theta_1 = 5\%$, $\theta_2 = -5\%$, $\theta_3 = 6\%$ *and* $\theta_4 = 10$. *We consider several credit default swaps on the same entity with quarterly coupons and a notional of* \$1 *mn. The recovery rate* $\boldsymbol{\mathcal{R}}$ *is set to* 40% *whereas the default time* $\boldsymbol{\tau}$ *is an exponential random variable, whose parameter* λ *is equal to* 50 *bps. We consider seven maturities (6M, 1Y, 2Y, 3Y, 5Y, 7Y and 10Y) and two coupon rates (*10 *and* 100 *bps).*

To calculate the prices of these CDS, we use Equation (3.5) with $N = 10^6$, $\boldsymbol{c} = 10$ (or 100) $\times 10^{-4}$, $\Delta t_m = 1/4$, $\lambda = 50 \times 10^{-4} = 0.005$, $\boldsymbol{\mathcal{R}} = 0.40$, $\mathbf{S}_t(u) = e^{-0.005(u-t)}$, $f_t(u) = 0.005 \cdot e^{-0.005(u-t)}$ and $B_t(u) = e^{-(u-t)R_t(u)}$ where the zero-coupon rate is given by Equation (3.1). To evaluate the integral, we consider a Gauss-Legendre quadrature of 128^{th} order. By including the accrued premium[28], we obtain results reported in Table 3.7. For instance, the price of the 5Y CDS is equal to \$9 527 if $\boldsymbol{c} = 10 \times 10^{-4}$ and $-\$33\,173$ if $\boldsymbol{c} = 100 \times 10^{-4}$. In the first case, the protection buyer has to pay an upfront to the protection seller because the coupon rate is too low. In the second case, the protection buyer receives the upfront because the coupon rate is too high. We also indicate the spread $\boldsymbol{s}$ and the risky PV01. We notice that the CDS spread is almost constant. This is normal since the default rate is constant. This is why the CDS spread is approximatively equal to $(1 - 40\%) \times 50$ bps or 30 bps. The difference between the several maturities is due to the yield curve. The risky PV01 is a useful statistic to compute the mark-to-market. Suppose for instance that the two parties entered in a 7Y credit default swap of 10 bps spread two years ago. Now, the residual maturity of the swap is five years, meaning that the mark-to-market of the protection buyer is equal to:

$$\begin{aligned}\Pi^{\text{buyer}} &= 10^6 \times \left(30.08 \times 10^{-4} - 10 \times 10^{-4}\right) \times 4.744 \\ &= \$9\,526\end{aligned}$$

We retrieve the 5Y CDS price (subject to rounding error).

TABLE 3.7: Price, spread and risky PV01 of CDS contracts

T	$P_t(T)$		$\boldsymbol{s}$	RPV_{01}
	$\boldsymbol{c} = 10$	$\boldsymbol{c} = 100$		
1/2	998	−3 492	30.01	0.499
1	1 992	−6 963	30.02	0.995
2	3 956	−13 811	30.04	1.974
3	5 874	−20 488	30.05	2.929
5	9 527	−33 173	30.08	4.744
7	12 884	−44 804	30.10	6.410
10	17 314	−60 121	30.12	8.604

Example 25 *We consider a variant of Example 24 by assuming that the default time follows a Gompertz distribution:*

$$\mathbf{S}_t(u) = \exp\left(\phi\left(1 - e^{\gamma(u-t)}\right)\right)$$

The parameters ϕ *and* γ *are set to* 5% *and* 10%.

[28] This means that the risky PV01 corresponds to Equation (3.9). We also report results without taking into account the accrued premium in Table 3.8. We notice that its impact is limited.

TABLE 3.8: Price, spread and risky PV01 of CDS contracts (without the accrued premium)

T	$P_t(T)$		s	RPV_{01}
	$c = 10$	$c = 100$		
1/2	999	−3 489	30.03	0.499
1	1 993	−6 957	30.04	0.994
2	3 957	−13 799	30.06	1.973
3	5 876	−20 470	30.07	2.927
5	9 530	−33 144	30.10	4.742
7	12 888	−44 764	30.12	6.406
10	17 319	−60 067	30.14	8.598

Results are reported in Table 3.9. In this example, the spread is increasing with the maturity of the CDS. Until now, we have assumed that we know the survival function $\mathbf{S}_t(u)$ in order to calculate the CDS spread. However, in practice, the CDS spread s is a market price and $\mathbf{S}_t(u)$ has to be determined thanks to a calibration procedure. Suppose for instance that we postulate that τ is an exponential default time with parameter λ. We can calibrate the estimated value $\hat{\lambda}$ such that the theoretical price is equal to the market price. For instance, Table 3.9 shows the parameter $\hat{\lambda}$ for each CDS. We found that $\hat{\lambda}$ is equal to 51.28 bps for the six-month maturity and 82.92 bps for the ten-year maturity. We face here an issue, because the parameter $\hat{\lambda}$ is not constant, meaning that we cannot use an exponential distribution to represent the default time of the reference entity. This is why we generally consider a more flexible survival function to calibrate the default probabilities from a set of CDS spreads[29].

TABLE 3.9: Calibration of the CDS spread curve using the exponential model

T	$P_t(T)$		s	RPV_{01}	$\hat{\lambda}$
	$c = 10$	$c = 100$			
1/2	1 037	−3 454	30.77	0.499	51.28
1	2 146	−6 808	31.57	0.995	52.59
2	4 585	−13 175	33.24	1.973	55.34
3	7 316	−19 026	35.00	2.927	58.25
5	13 631	−28 972	38.80	4.734	64.54
7	21 034	−36 391	42.97	6.380	71.44
10	33 999	−42 691	49.90	8.521	82.92

3.1.3.3 Basket default swap

A basket default swap is similar to a credit default swap except that the underlying asset is a basket of reference entities rather than one single reference entity. These products are part of multi-name credit default swaps with collateralized debt obligations.

First-to-default and k^{th}-to-default credit derivatives Let us consider a credit portfolio with n reference entities, which are referenced by the index i. With a first-to-default (FtD) credit swap, the credit event corresponds to the first time that a reference entity of the

[29] This problem will be solved later in Section 3.3.3.1 on page 203.

credit portfolio defaults. We deduce that the stochastic discounted values of the premium and default legs are[30]:

$$SV_t(\mathcal{PL}) = \boldsymbol{c} \cdot N \cdot \sum_{t_m \geq t} \Delta t_m \cdot \mathbb{1}\{\boldsymbol{\tau}_{1:n} > t_m\} \cdot e^{-\int_t^{t_m} r(s)\,\mathrm{d}s}$$

and:

$$SV_t(\mathcal{DL}) = \mathfrak{X} \cdot \mathbb{1}\{\boldsymbol{\tau}_{1:n} \leq T\} \cdot e^{-\int_t^{\boldsymbol{\tau}_{1:n}} r_s\,\mathrm{d}s}$$

where $\boldsymbol{\tau}_i$ is the default time of the i^{th} reference entity, $\boldsymbol{\tau}_{1:n} = \min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ is the first default time in the portfolio and $\mathfrak{X}$ is the payout of the protection leg:

$$\begin{aligned}\mathfrak{X} &= \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{\tau}_{1:n} = \boldsymbol{\tau}_i\} \cdot (1 - \boldsymbol{\mathcal{R}}_i) \cdot N_i \\ &= (1 - \boldsymbol{\mathcal{R}}_{i^\star}) \cdot N_{i^\star}\end{aligned}$$

In this formula, $\boldsymbol{\mathcal{R}}_i$ and N_i are respectively the recovery and the notional of the i^{th} reference entity whereas the index $i^\star = \{i : \boldsymbol{\tau}_i = \boldsymbol{\tau}_{1:n}\}$ corresponds to the first reference entity that defaults. For instance, if the portfolio is composed by 10 names and the third name is the first default, the value of the protection leg will be equal to $(1 - \boldsymbol{\mathcal{R}}_3) \cdot N_3$. Using the same assumptions than previously, we deduce that the FtD spread is:

$$\boldsymbol{s}^{\text{FtD}} = \frac{\mathbb{E}\left[\mathfrak{X} \cdot \mathbb{1}\{\boldsymbol{\tau}_{1:n} \leq T\} \cdot B_t(\boldsymbol{\tau}_{1:n})\right]}{N \sum_{t_m \geq t} \Delta t_m \cdot \mathbf{S}_{1:n,t}(t_m) \cdot B_t(t_m)}$$

where $\mathbf{S}_{1:n,t}(u)$ is the survival function of $\boldsymbol{\tau}_{1:n}$. If we assume a homogenous basket (same recovery $\boldsymbol{\mathcal{R}}_i = \boldsymbol{\mathcal{R}}$ and same notional $N_i = N$), the previous formula becomes:

$$\boldsymbol{s}^{\text{FtD}} = \frac{(1 - \boldsymbol{\mathcal{R}}) \int_t^T B_t(u)\, f_{1:n,t}(u)\,\mathrm{d}u}{\sum_{t_m \geq t} \Delta t_m \mathbf{S}_{1:n,t}(t_m)\, B_t(t_m)} \tag{3.10}$$

where $f_{1:n,t}(u)$ is the survival function of $\boldsymbol{\tau}_{1:n}$.

To compute the spread[31], we use Monte Carlo simulation (or numerical integration when the number of entities is small). In fact, the survival function of $\boldsymbol{\tau}_{1:n}$ is related to the individual survival functions, but also to the dependence between the default times $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$. The spread of the FtD is then a function of default correlations[32]. If we denote by $\boldsymbol{s}_i^{\text{CDS}}$ the CDS spread of the i^{th} reference, we can show that:

$$\max\left(\boldsymbol{s}_1^{\text{CDS}}, \ldots, \boldsymbol{s}_n^{\text{CDS}}\right) \leq \boldsymbol{s}^{\text{FtD}} \leq \sum_{i=1}^{n} \boldsymbol{s}_i^{\text{CDS}} \tag{3.11}$$

When the default times are uncorrelated, the FtD is equivalent to buy the basket of all the credit defaults swaps. In the case of a perfect correlation, one default is immediately followed by the other $n - 1$ defaults, implying that the FtD is equivalent to the CDS with the worst spread. In practice, the FtD spread is therefore located between these two bounds as expressed in Equation (3.11). From the viewpoint of the protection buyer, a FtD is seen as a hedging method of the credit portfolio with a lower cost than buying the protection

[30] In order to simplify the notations, we do not take into account the accrued premium.

[31] Laurent and Gregory (2005) provide semi-explicit formulas that are useful for pricing basket default swaps.

[32] This point is developed in Section 3.3.4 on page 220 and in Chapter 11 dedicated to copula functions.

for all the credits. For example, suppose that the protection buyer would like to be hedged to the default of the automobile sector. He can buy a FtD on the basket of the largest car manufacturers in the world, e.g. Volkswagen, Toyota, Hyundai, General Motors, Fiat Chrysler and Renault. If there is only one default, the protection buyer is hedged. However, the protection buyer keeps the risk of multiple defaults, which is a worst-case scenario.

Remark 27 *The previous analysis can be extended to k^{th}-to-default swaps. In this case, the default leg is paid if the k^{th} default occurs before the maturity date. We then obtain a similar expression as Equation (3.10) by considering the order statistic $\boldsymbol{\tau}_{k:n}$ in place of $\boldsymbol{\tau}_{1:n}$.*

From a theoretical point of view, it is equivalent to buy the CDS protection for all the components of the credit basket or to buy all the k^{th}-to-default swaps. We have therefore the following relationship:

$$\sum_{i=1}^{n} \boldsymbol{s}_i^{\text{CDS}} = \sum_{i=1}^{n} \boldsymbol{s}^{i:n} \tag{3.12}$$

We see that the default correlation highly impacts the distribution of the k^{th}-to-default spreads[33].

Credit default indices Credit derivative indices[34] have been first developed by J.P. Morgan, Morgan Stanley and iBoxx between 2001 and 2003. A credit default index (or CDX) is in fact a credit default swap on a basket of reference entities. As previously, we consider a portfolio with n credit entities. The protection buyer pays a premium leg with a coupon rate $\boldsymbol{c}$. Every time a reference entity defaults, the notional is reduced by a factor, which is equal to $1/n$. At the same time, the protection buyer receives the portfolio loss between two fixing dates. The expression of the notional outstanding is then given by:

$$N_t(u) = N \cdot \left(1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{\tau}_i \leq u\}\right)$$

At the inception date, we verify that $N_t(t) = N$. After the first default, the notional outstanding is equal to $N(1 - 1/n)$. After the k^{th} default, its value is $N(1 - k/n)$. At time $u \geq t$, the cumulative loss of the credit portfolio is:

$$L_t(u) = \frac{1}{n} \sum_{i=1}^{n} N \cdot (1 - \boldsymbol{\mathcal{R}}_i) \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq u\}$$

meaning that the incremental loss between two fixing dates is:

$$\Delta L_t(t_m) = L_t(t_m) - L_t(t_{m-1})$$

We deduce that the stochastic discounted value of the premium and default legs is:

$$SV_t(\mathcal{PL}) = \boldsymbol{c} \cdot \sum_{t_m \geq t} \Delta t_m \cdot N_t(t_m) \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s}$$

and:

$$SV_t(\mathcal{DL}) = \sum_{t_m \geq t} \Delta L_t(t_m) \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s}$$

[33] See page 762 for an illustration.

[34] They are also known as synthetic credit indices, credit default swap indices or credit default indices.

We deduce that the spread of the CDX is:

$$\mathcal{s}^{\text{CDX}} = \frac{\mathbb{E}\left[\sum_{t_m \geq t} \Delta L_t(t_m) \cdot B_t(t_m)\right]}{\mathbb{E}\left[\sum_{t_m \geq t} \Delta t_m \cdot N_t(t_m) \cdot B_t(t_m)\right]} \tag{3.13}$$

Remark 28 *A CDX is then equivalent to a portfolio of CDS whose each principal notional is equal to N/n. Indeed, when a default occurs, the protection buyer receives $N/n \cdot (1 - \mathcal{R}_i)$ and stops to pay the premium leg of the defaulted reference entity. At the inception date, the annual premium of the CDX is then equal to the annual premium of the CDS portfolio:*

$$\mathcal{s}^{\text{CDX}} \cdot N = \sum_{i=1}^{n} \mathcal{s}_i^{\text{CDS}} \cdot \frac{N}{n}$$

We deduce that the spread of the CDX is an average of the credit spreads that compose the portfolio[35]*:*

$$\mathcal{s}^{\text{CDX}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{s}_i^{\text{CDS}} \tag{3.14}$$

Today, credit default indices are all managed by Markit and have been standardized. For instance, coupon payments are made on a quarterly basis (March 20, June 20, September 20, December 20) whereas indices roll every six months with an updated portfolio[36]. With respect to the original credit indices, Markit continues to produces two families:

- Markit CDX
  It focuses on North America and Emerging Markets credit default indices. The three major sub-indices are IG (investment grade), HY (high yield) and EM (emerging markets). A more comprehensive list is provided in Table 3.10. Besides these credit default indices, Markit CDX produces also four other important indices: ABX (basket of ABS), CMBX (basket of CMBS), LCDX (portfolio of 100 US secured senior loans) and MCDX (basket of 50 municipal bonds).

- Markit iTraxx
  It focuses on Europe, Japan, Asia ex-Japan and Australia (see the list in Table 3.11). Markit iTraxx also produces LevX (portfolio of 40 European secured loans), sector indices (e.g. European financials and industrials) and SovX, which corresponds to a portfolio of sovereign issuers. There are 7 SovX indices: Asia Pacific, BRIC, CEEMEA[37], G7, Latin America, Western Europe and Global Liquid IG.

In Table 3.12, we report the spread of some CDX/iTraxx indices. We note that the spread of the CDX.NA.HY index is on average four times larger than the spread of the CDX.NA.IG index. While spreads of credit default indices have generally decreased between December 2012 and December 2014, we observe a reversal in 2015. For instance, the spread of the CDX.NA.IG index is equal to 93.6 bps in September 2015 whereas it was only equal to 66.3 bps nine months ago. We observe a similar increase of 30 bps for the iTraxx Europe index. For the CDX.NA.HY index, it is more impressive with a variation of +150 bps in nine months.

[35] In fact, this is an approximation because the payment of the default leg does not exactly match between the CDX index and the CDS portfolio.
[36] See Markit (2014) for a detailed explanation of the indices' construction.
[37] Central and Eastern Europe, Middle East and Africa.

TABLE 3.10: List of Markit CDX main indices

Index name	Description	n	$\mathcal{R}$
CDX.NA.IG	Investment grade entities	125	40%
CDX.NA.IG.HVOL	High volatility IG entities	30	40%
CDX.NA.XO	Crossover entities	35	40%
CDX.NA.HY	High yield entities	100	30%
CDX.NA.HY.BB	High yield BB entities	37	30%
CDX.NA.HY.B	High yield B entities	46	30%
CDX.EM	EM sovereign issuers	14	25%
LCDX	Secured senior loans	100	70%
MCDX	Municipal bonds	50	80%

Source: Markit (2014).

TABLE 3.11: List of Markit iTraxx main indices

Index name	Description	n	$\mathcal{R}$
iTraxx Europe	European IG entities	125	40%
iTraxx Europe HiVol	European HVOL IG entities	30	40%
iTraxx Europe Crossover	European XO entities	40	40%
iTraxx Asia	Asian (ex-Japan) IG entities	50	40%
iTraxx Asia HY	Asian (ex-Japan) HY entities	20	25%
iTraxx Australia	Australian IG entities	25	40%
iTraxx Japan	Japanese IG entities	50	35%
iTraxx SovX G7	G7 governments	7	40%
iTraxx LevX	European leveraged loans	40	40%

Source: Markit (2014).

TABLE 3.12: Historical spread of CDX/iTraxx indices (in bps)

Date	CDX			iTraxx		
	NA.IG	NA.HY	EM	Europe	Japan	Asia
Dec. 2012	94.1	484.4	208.6	117.0	159.1	108.8
Dec. 2013	62.3	305.6	272.4	70.1	67.5	129.0
Dec. 2014	66.3	357.2	341.0	62.8	67.0	106.0
Sep. 2015	93.6	505.3	381.2	90.6	82.2	160.5

3.1.3.4 Collateralized debt obligations

A collateralized debt obligation (CDO) is another form of multi-name credit default swaps. It corresponds to a pay-through ABS structure[38], whose securities are bonds linked to a series of tranches. If we consider the example given in Figure 3.20, they are 4 types of bonds, whose returns depend on the loss of the corresponding tranche (equity, mezzanine, senior and super senior). Each tranche is characterized by an attachment point A and a

[38] See Figure 3.12 on page 139.

detachment point D. In our example, we have:

Tranche	Equity	Mezzanine	Senior	Super senior
A	0%	15%	25%	35%
D	15%	25%	35%	100%

The protection buyer of the tranche $[A, D]$ pays a coupon rate $\mathbf{c}^{[A,D]}$ on the nominal outstanding amount of the tranche to the protection seller. In return, he receives the protection leg, which is the loss of the tranche $[A, D]$. However, the losses satisfy a payment priority which is the following:

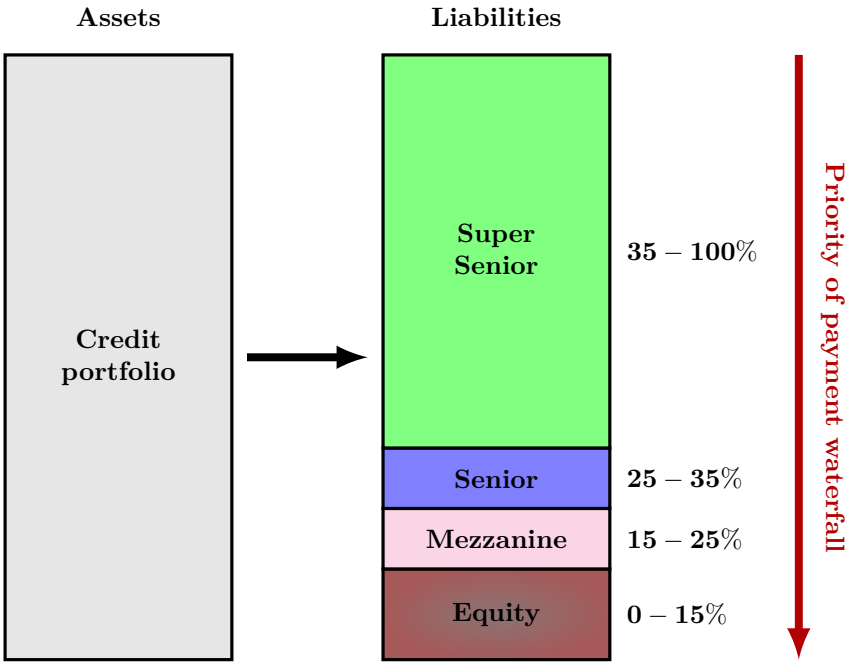


FIGURE 3.20: Structure of a collateralized debt obligation

- the equity tranche is the most risky security, meaning that the first losses hit this tranche alone until the cumulative loss reaches the detachment point;
- from the time the portfolio loss is larger than the detachment point of the equity tranche, the equity tranche no longer exists and this is the protection seller of the mezzanine tranche, who will pay the next losses to the protection buyer of the mezzanine tranche;
- the protection buyer of a tranche pays the coupon from the inception of the CDO until the death of the tranche, *i.e.*, when the cumulative loss is larger than the detachment point of the tranche; moreover, the premium payments are made on the reduced notional after each credit event of the tranche.

Each CDO tranche can then be viewed as a CDS with a time-varying notional principal to define the premium leg and a protection leg, which is paid if the portfolio loss is between the attachment and detachment points of the tranche. We can therefore interpret a CDO

as a basket default swap, where the equity, mezzanine, senior and super senior tranches correspond respectively to a first-to-default, second-to-default, third-to-default and last-to-default swaps.

Let us now see the mathematical framework to price a CDO tranche. Assuming a portfolio of n credits, the cumulative loss is equal to:

$$L_t(u) = \sum_{i=1}^{n} N_i \cdot (1 - \boldsymbol{\mathcal{R}}_i) \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq u\}$$

whereas the loss of the tranche $[A, D]$ is given by[39]:

$$\begin{aligned} L_t^{[A,D]}(u) &= (L_t(u) - A) \cdot \mathbb{1}\{A \leq L_t(u) \leq D\} + \\ &\quad (D - A) \cdot \mathbb{1}\{L_t(u) > D\} \end{aligned}$$

where A and D are the attachment and detachment points expressed in \$. The nominal outstanding amount of the tranche is therefore:

$$N_t^{[A,D]}(u) = (D - A) - L_t^{[A,D]}(u)$$

This notional principal decreases then by the loss of the tranche. At the inception of the CDO, $N_t^{[A,D]}(t)$ is equal to the tranche thickness: $(D - A)$. At the maturity date T, we have:

$$\begin{aligned} N_t^{[A,D]}(T) &= (D - A) - L_t^{[A,D]}(T) \\ &= \begin{cases} (D - A) & \text{if } L_t(T) \leq A \\ (D - L_t(T)) & \text{if } A < L_t(T) \leq D \\ 0 & \text{if } L_t(T) > D \end{cases} \end{aligned}$$

We deduce that the stochastic discounted value of the premium and default legs is:

$$SV_t(\mathcal{PL}) = \boldsymbol{c}^{[A,D]} \cdot \sum_{t_m \geq t} \Delta t_m \cdot N_t^{[A,D]}(t_m) \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s}$$

and:

$$SV_t(\mathcal{DL}) = \sum_{t_m \geq t} \Delta L_t^{[A,D]}(t_m) \cdot e^{-\int_t^{t_m} r_s \, \mathrm{d}s}$$

Therefore, the spread of the CDO tranche is[40]:

$$\boldsymbol{s}^{[A,D]} = \frac{\mathbb{E}\left[\sum_{t_m \geq t} \Delta L_t^{[A,D]}(t_m) \cdot B_t(t_m)\right]}{\mathbb{E}\left[\sum_{t_m \geq t} \Delta t_m \cdot N_t^{[A,D]}(t_m) \cdot B_t(t_m)\right]} \tag{3.15}$$

We obviously have the following inequalities:

$$\boldsymbol{s}^{\text{Equity}} > \boldsymbol{s}^{\text{Mezzanine}} > \boldsymbol{s}^{\text{Senior}} > \boldsymbol{s}^{\text{Super senior}}$$

[39] Another expression is:

$$L_t^{[A,D]}(u) = \min\left(D - A, (L_t(u) - A)^+\right)$$

[40] This formula is obtained by assuming no upfront and accrued interests.

As in the case of k^{th}-to-default swaps, the distribution of these tranche spreads highly depends on the default correlation[41]. Depending on the model and the parameters, we can therefore promote the protection buyer/seller of one specific tranche with respect to the other tranches.

When collateralized debt obligations emerged in the 1990s, they were used to transfer credit risk from the balance sheet of banks to investors (e.g. insurance companies). They were principally portfolios of loans (CLO) or asset-backed securities (ABS CDO). With these balanced-sheet CDOs, banks could recover regulatory capital in order to issue new credits. In the 2000s, a new type of CDOs was created by considering CDS portfolios as underlying assets. These synthetic CDOs are also called arbitrage CDOs, because they have used by investors to express their market views on credit.

The impressive success of CDOs with investors before the 2008 Global Financial Crisis is due to the rating mechanism of tranches. Suppose that the underlying portfolio is composed of BB rated credits. It is obvious that the senior and super senior tranches will be rated higher than BB, because the probability that these tranches will be impacted is very low. The slicing approach of CDOs enables then to create high-rated securities from medium or low-rated debts. Since the appetite of investors for AAA and AA rated bonds was very important, CDOs were solutions to meet this demand. Moreover, this lead to the development of rating methods in order to provide an attractive spread. This explains that most of AAA-rated CDO tranches promised a return higher than AAA-rated sovereign and corporate bonds. In fact, the 2008 GFC has demonstrated that many CDO tranches were more risky than expected, because the riskiness of the assets were underestimated[42].

TABLE 3.13: List of Markit credit default tranches

Index name	Tranche				
CDX.NA.IG	$0-3$	$3-7$	$7-15$	$15-100$	
CDX.NA.HY	$0-10$	$10-15$	$15-25$	$25-35$	$35-100$
LCDX	$0-5$	$5-8$	$8-12$	$12-15$	$15-100$
iTraxx Europe	$0-3$	$3-6$	$6-9$	$9-12$	$12-22$	$22-100$
iTraxx Europe XO	$0-10$	$10-15$	$15-25$	$25-35$	$35-100$
iTraxx Asia	$0-3$	$3-6$	$6-9$	$9-12$	$12-22$
iTraxx Australia	$0-3$	$3-6$	$6-9$	$9-12$	$12-22$
iTraxx Japan	$0-3$	$3-6$	$6-9$	$9-12$	$12-22$

Source: Markit (2014).

For some years now, CDOs have been created using credit default indices as the underlying portfolio. For instance, Table 3.13 provides the list of available tranches on Markit indices[43]. We notice that attachment and detachment points differ from one index to another index. The first tranche always indicates the equity tranche. For IG underlying assets, the notional corresponds to the first 3% losses of the portfolio, whereas the detachment point is higher for crossover or high yield assets. We also notice that some senior tranches are not traded (Asia, Australia and Japan). These products are mainly used in correlation trading activities and also served as benchmarks for all the other OTC credit debt obligations.

[41] See Section 3.3.4 on page 220.

[42] More details of the impact of the securitization market on the 2008 Global Financial Crisis are developed in Chapter 8 dedicated to systemic risk.

[43] They are also called credit default tranches (CDT).

3.2 Capital requirement

This section deals with regulatory aspects of credit risk. From a historical point of view, this is the first risk which has requested regulatory capital before market risk. Nevertheless, the development of credit risk management is more recent and was accelerated with the Basel II Accord. Before presenting the different approaches for calculating capital requirements, we need to define more precisely what credit risk is.

It is the risk of loss on a debt instrument resulting from the failure of the borrower to make required payments. We generally distinguish two types of credit risk. The first one is the '*default risk*', which arises when the borrower is unable to pay the principal or interests. An example is a student loan or a mortgage loan. The second type is the '*downgrading risk*', which concerns debt securities. In this case, the debt holder may face a loss, because the price of the debt security is directly related to the credit risk of the borrower. For instance, the price of the bond may go down because the credit risk of the issuer increases and even if the borrower does not default. Of course, default risk and downgrading risk are highly correlated, because it is rare that a counterparty suddenly defaults without downgrading of its credit rating.

To measure credit risk, we first need to define the default of the obligor. BCBS (2006) provides the following standard definition:

> "*A default is considered to have occurred with regard to a particular obligor when either or both of the two following events have taken place.*
>
> - *The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realizing security (if held).*
> - *The obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstandings*" (BCBS, 2006, page 100).

This definition contains both objective elements (when a payment has been missed or delayed) and subjective elements (when a loss becomes highly probable). This last case generally corresponds to an extreme situation (specific provision, distressed restructuring, etc.). The Basel definition of default covers then two types of credit: debts under litigation and doubtful debts.

Downgrading risk is more difficult to define. If the counterparty is rated by an agency, it can be measured by a single or multi-notch downgrade. However, it is not always the case in practice, because the credit quality decreases before the downgrade announcement. A second measure is to consider a market-based approach by using CDS spreads. However, we notice that the two methods concern counterparties, which are able to issue debt securities, in particular bonds. For instance, the concept of downgrading risk is difficult to apply for retail assets.

The distinction between default risk and downgrading risk has an impact on the credit risk measure. For loans and debt-like instruments that cannot be traded in a market, the time horizon for managing credit risk is the maturity of the credit. Contrary to this held-to-maturity approach, the time horizon for managing debt securities is shorter, typically one year. In this case, the big issue is not to manage the default, but the mark-to-market of the credit exposure.

3.2.1 The Basel I framework

According to Tarullo (2008), two explanatory factors were behind the Basel I Accord. The first motivation was to increase capital levels of international banks, which were very low at that time and had continuously decreased for many years. For instance, the ratio of equity capital to total assets[44] was 5.15% in 1970 and only 3.83% in 1981 for the 17 largest US banks. In 1988, this capital ratio was equal to 2.55% on average for the five largest bank in the world. The second motivation concerned the distortion risk of competition resulting from heterogeneous national capital requirements. One point that was made repeatedly, especially by US bankers, was the growth of Japanese banks. In Table 3.14, we report the ranking of the 10 world's largest banks in terms of assets ($ bn) between 2001 and 2008. While there is only one Japanese bank in the top 10 in 1981, nine Japanese banks are included in the ranking seven years later. In this context, the underlying idea of the Basel I Accord was then to increase capital requirements and harmonize national regulations for international banks.

TABLE 3.14: World's largest banks in 1981 and 1988

	1981		1988	
	Bank	Assets	Bank	Assets
1	Bank of America (US)	115.6	Dai-Ichi Kangyo (JP)	352.5
2	Citicorp (US)	112.7	Sumitomo (JP)	334.7
3	BNP (FR)	106.7	Fuji (JP)	327.8
4	Crédit Agricole (FR)	97.8	Mitsubishi (JP)	317.8
5	Crédit Lyonnais (FR)	93.7	Sanwa (JP)	307.4
6	Barclays (UK)	93.0	Industrial Bank (JP)	261.5
7	Société Générale (FR)	87.0	Norinchukin (JP)	231.7
8	Dai-Ichi Kangyo (JP)	85.5	Crédit Agricole (FR)	214.4
9	Deutsche Bank (DE)	84.5	Tokai (JP)	213.5
10	National Westminster (UK)	82.6	Mitsubishi Trust (JP)	206.0

Source: Tarullo (2008).

The Basel I Accord provides a detailed definition of bank capital C and risk-weighted assets RWA. We reiterate that tier one (T1) capital consists mainly of common stock and disclosed reserves, whereas tier two (T2) capital includes undisclosed reserves, general provisions, hybrid debt capital instruments and subordinated term debt. Risk-weighted assets are simply calculated as the product of the asset notional (the exposure at default or EAD) by a risk weight (RW). Table 3.15 shows the different values of RW with respect to the category of the asset. For off-balance sheet assets, BCBS (1988) defines credit conversion factor (CCF) for converting the amount E of a credit line or off-balance sheet asset to an exposure at default:

$$\text{EAD} = E \cdot \text{CCF}$$

The CCF values are 100% for direct credit substitutes (standby letters of credit), sale and repurchase agreements, forward asset purchases, 50% for standby facilities and credit lines with an original maturity of over one year, note issuance facilities and revolving underwriting facilities, 20% for short-term self-liquidating trade-related contingencies and 0% for standby facilities and credit lines with an original maturity of up to one year. The above framework is used to calculate the Cooke ratio, which is in fact a set of two capital ratios. The core

[44] All the statistics of this section comes from Chapters 2 and 3 of Tarullo (2008).

TABLE 3.15: Risk weights by category of on-balance sheet assets

RW	Instruments
0%	Cash Claims on central governments and central banks denominated in national currency and funded in that currency Other claims on OECD central governments and central banks Claims† collateralized by cash of OECD government securities
20%	Claims† on multilateral development banks Claims† on banks incorporated in the OECD and claims guaranteed by OECD incorporated banks Claims† on securities firms incorporated in the OECD subject to comparable supervisory and regulatory arrangements Claims† on banks incorporated in countries outside the OECD with a residual maturity of up to one year Claims† on non-domestic OECD public-sector entities Cash items in process of collection
50%	Loans fully secured by mortgage on residential property
100%	Claims on the private sector Claims on banks incorporated outside the OECD with a residual maturity of over one year Claims on central governments outside the OECD and non denominated in national currency All other assets

†or guaranteed by these entities.

Source: BCBS (1988).

capital ratio includes only tier one capital whereas the total capital ratio considers both tier one C_1 and tier two C_2 capital:

$$\begin{aligned} \text{Tier 1 ratio} &= \frac{C_1}{\text{RWA}} \geq 4\% \\ \text{Tier 2 ratio} &= \frac{C_1 + C_2}{\text{RWA}} \geq 8\% \end{aligned}$$

Example 26 *The assets of the bank are composed of* \$100 *mn of US treasury bonds,* \$20 *mn of Mexico government bonds denominated in US dollar,* \$20 *mn of Argentine debt denominated in Argentine peso,* \$500 *mn of residential mortgage,* \$500 *mn of corporate loans,* \$20 *mn of non-used standby facilities for OECD governments and* \$100 *mn of retail credit lines, which are decomposed as follows:* \$40 *mn are used and* 70% *of non-used credit lines have a maturity greater than one year.*

For each asset, we calculate RWA by choosing the right risk weight and credit conversion factor for off-balance sheet items. We obtain the results below. The risk-weighted assets of the bank are then equal to \$831 mn. We deduce that the required capital $\boldsymbol{\mathcal{K}}$ is \$33.24 mn

for tier one.

Balance Sheet	Asset	E	CCF	EAD	RW	RWA
On-	US bonds			100	0%	0
	Mexico bonds			20	100%	20
	Argentine debt			20	0%	0
	Home mortgage			500	50%	250
	Corporate loans			500	100%	500
	Credit lines			40	100%	40
Off-	Standby facilities	20	100%	20	0%	0
	Credit lines (> 1Y)	42	50%	21	100%	21
	Credit lines (≤ 1Y)	18	0%	0	100%	0
	Total					831

3.2.2 The Basel II standardized approach

The main criticism of the Cooke ratio is the lack of economic rationale with respect to risk weights. Indeed, most of the claims have a 100% risk weight and do not reflect the real credit risk of the borrower. Other reasons have been given to justify a reformulation of capital requirements for credit risk with the goal to:

- obtain a better credit risk measure by taking into account the default probability of the counterparty;
- avoid regulatory arbitrage, in particular by using credit derivatives;
- have a more coherent framework that supports credit risk mitigation.

3.2.2.1 Standardized risk weights

In Basel II, the probability of default is the key parameter to define risk weights. For the standardized approach (SA), they depend directly on external ratings whereas they are based on internal rating for the IRB approach. Table 3.16 shows the new matrix of risk weights, when we consider the Standard & Poor's rating system[45]. We notice that there are four main categories of claims[46]: sovereigns, banks, corporates and retail portfolios.

The sovereign exposure category include central governments and central banks, whereas non-central public sector entities are treated with the bank exposure category. We note that there are two options for the latter, whose choice is left to the discretion of the national supervisors[47]. Under the first option, the risk weight depends on the rating of the country where the bank is located. Under the second option, it is the rating of the bank that determines the risk weight, which is more favorable for short-term claims (three months or less). The risk weight of a corporate is calculated with respect to the rating of the entity, but uses a slightly different breakdown of ratings than the second option of the bank category. Finally, the Basel Committee uses lower levels for retail portfolios than those provided in the Basel I Accord. Indeed, residential mortgages and retail loans are now risk-weighted at 35% and 75% instead of 50% and 100% previously. Other comparisons between Basel I and Basel II (with the second option for banks) are shown in Table 3.17.

[45] NR stands for non-rated entities.

[46] The regulatory framework is more comprehensive by considering three other categories (public sector entities, multilateral development banks and securities firms), which are treated as banks. For all other assets, the standard risk weight is 100%.

[47] The second option is more frequent and was implemented in Europe, US and Japan for instance.

TABLE 3.16: Risk weights of the SA approach (Basel II)

Rating		AAA to AA−	A+ to A−	BBB+ to BBB−	BB+ to B−	CCC+ to C	NR
Sovereigns		0%	20%	50%	100%	150%	100%
Banks	1	20%	50%	100%	100%	150%	100%
	2	20%	50%	50%	100%	150%	50%
	2 ST	20%	20%	20%	50%	150%	20%
Corporates		20%	50%	BBB+ to BB− 100%		B+ to C 150%	100%
Retail					75%		
Residential mortgages					35%		
Commercial mortgages					100%		

TABLE 3.17: Comparison of risk weights between Basel I and Basel II

Entity	Rating	Maturity	Basel I	Basel II
Sovereign (OECD)	AAA		0%	0%
Sovereign (OECD)	A-		0%	20%
Sovereign	BBB		100%	50%
Bank (OECD)	BBB	2Y	20%	50%
Bank	BBB	2M	100%	20%
Corporate	AA+		100%	20%
Corporate	BBB		100%	100%

The SA approach is based on external ratings and then depends on credit rating agencies. The most famous are Standard & Poor's, Moody's and Fitch. However, they cover only large companies. This is why banks will also consider rating agencies specialized in a specific sector or a given country[48]. Of course, rating agencies must be first registered and certified by the national supervisor in order to be used by the banks. The validation process consists of two steps, which are the assessment of the six required criteria (objectivity, independence, transparency, disclosure, resources and credibility) and the mapping process between the ratings and the Basel matrix of risk weights.

Table 3.18 shows the rating systems of S&P, Moody's and Fitch, which are very similar. Examples of S&P's rating are given in Tables 3.19, 3.20 and 3.21. We note that the rating of many sovereign counterparties has been downgraded by at least one notch, except China which has now a better rating than before the 2008 GFC. For some countries, the rating in local currency is different from the rating in foreign currency, for instance Argentina, Brazil, Russia and Ukraine[49]. We observe the same evolution for banks and it is now rare to find a bank with a AAA rating. This is not the case of corporate counterparties, which present more stable ratings across time.

Remark 29 *Credit conversion factors for off-balance sheet items are similar to those defined in the original Basel Accord. For instance, any commitment that is unconditionally cancelable receives a* 0% *CCF. A CCF of* 20% *(resp.* 50%*) is applied to commitments with*

[48] For instance, banks may use Japan Credit Rating Agency Ltd for Japanese public and corporate entities, DBRS Ratings Limited for bond issuers, Cerved Rating Agency for Italian small and medium-sized enterprises, etc.

[49] An SD rating is assigned in case of selective default of the obligor.

TABLE 3.18: Credit rating system of S&P, Moody's and Fitch

	Prime Maximum Safety		High Grade High Quality			Upper Medium Grade		
S&P/Fitch	AAA		AA+	AA	AA−	A+	A	A−
Moody's	Aaa		Aa1	Aa2	Aa3	A1	A2	A3

	Lower Medium Grade			Non Investment Grade Speculative		
S&P/Fitch	BBB+	BBB	BBB−	BB+	BB	BB−
Moody's	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3

	Highly Speculative			Substantial Risk	In Poor Standing		Extremely Speculative
S&P/Fitch	B+	B	B−	CCC+	CCC	CCC−	CC
Moody's	B1	B2	B3	Caa1	Caa2	Caa3	Ca

TABLE 3.19: Examples of country's S&P rating

Country	Local currency		Foreign currency	
	Jun. 2009	Oct. 2015	Jun. 2009	Oct. 2015
Argentina	B-	CCC+	B-	SD
Brazil	BBB+	BBB-	BBB-	BB+
China	A+	AA-	A+	AA-
France	AAA	AA	AAA	AA
Italy	A+	BBB-	A+	BBB-
Japan	AA	A+	AA	A+
Russia	BBB+	BBB-	BBB	BB+
Spain	AA+	BBB+	AA+	BBB+
Ukraine	B-	CCC+	CCC+	SD
US	AAA	AA+	AA+	AA+

Source: Standard & Poor's, www.standardandpoors.com.

TABLE 3.20: Examples of bank's S&P rating

Bank	Oct. 2001	Jun. 2009	Oct. 2015
Barclays Bank PLC	AA	AA-	A-
Credit Agricole S.A.	AA	AA-	A
Deutsche Bank AG	AA	A+	BBB+
International Industrial Bank	CCC+	BB-	
JPMorgan Chase & Co.	AA-	A+	A
UBS AG	AA+	A+	A

Source: Standard & Poor's, www.standardandpoors.com.

TABLE 3.21: Examples of corporate's S&P rating

Corporate	Jul. 2009	Oct. 2015
Danone	A-	A-
Exxon Mobil Corp.	AAA	AAA
Ford Motor Co.	CCC+	BBB-
General Motors Corp.	D	BBB-
L'Oreal S.A.	NR	NR
Microsoft Corp.	AAA	AAA
Nestle S.A.	AA	AA
The Coca-Cola Co.	A+	AA
Unilever PLC	A+	A+

Source: Standard & Poor's, www.standardandpoors.com.

an original maturity up to one year (resp. greater than one year). For revolving underwriting facilities, the CCF is equal to 50% *whereas it is equal to* 100% *for other off-balance sheet items (e.g. direct credit substitutes, guarantees, sale and repurchase agreements, forward asset purchases).*

3.2.2.2 Credit risk mitigation

Credit risk mitigation (CRM) refers to the various techniques used by banks for reducing the credit risk. These methods allow to decrease the credit exposure or to increase the recovery in case of default. The most common approaches are collateralized transactions, guarantees, credit derivatives and netting agreements.

Collateralized transactions In such operations, the credit exposure of the bank is partially hedged by collateral posted by the counterparty. BCBS (2006) defines then the following eligible instruments:

1. Cash and comparable instruments;
2. Gold;
3. Debt securities which are rated AAA to BB- when issued by sovereigns or AAA to BBB- when issued by other entities or at least A-3/P-3 for short-term debt instruments;
4. Debt securities which are not rated but fulfill certain criteria (senior debt issued by banks, listed on a recognisee exchange and sufficiently liquid);
5. Equities that are included in a main index;
6. UCITS and mutual funds, whose assets are eligible instruments and which offer a daily liquidity;
7. Equities which are listed on a recognized exchange and UCITS/mutual funds which include such equities.

The bank has the choice between two approaches to take into account collateralized transactions. In the simple approach[50], the risk weight of the collateral (with a floor of

[50] Collateral instruments (7) are not eligible for this approach.

20%) is applied to the market value of the collateral C whereas the non-hedged exposure $(\text{EAD} - C)$ receives the risk weight of the counterparty:

$$\text{RWA} = (\text{EAD} - C) \cdot \text{RW} + C \cdot \max(\text{RW}_C, 20\%) \tag{3.16}$$

where EAD is the exposure at default, C is the market value of the collateral, RW is the risk weight appropriate to the exposure and RW_C is the risk weight of the collateral. The second method, called the comprehensive approach, is based on haircuts. The risk-weighted asset amount after risk mitigation is $\text{RWA} = \text{RW} \cdot \text{EAD}^\star$ whereas $\text{EAD}^\star$ is the modified exposure at default defined as follows:

$$\text{EAD}^\star = \max(0, (1 + H_E) \cdot \text{EAD} - (1 - H_C - H_{FX}) \cdot C) \tag{3.17}$$

where H_E is the haircut applied to the exposure, H_C is the haircut applied to the collateral and H_{FX} is the haircut for currency risk. Table 3.22 gives the standard supervisory values of haircuts. If the bank uses an internal model to calculate haircuts, they must be based on the value-at-risk with a 99% confidence level and an holding period which depends on the collateral type and the frequency of remargining. The standard supervisory haircuts have been calibrated by assuming daily mark-to-market, daily remargining and a 10-business day holding period.

TABLE 3.22: Standardized supervisory haircuts for collateralized transactions

Rating	Residual Maturity	Sovereigns	Others
AAA to AA−	0−1Y	0.5%	1%
	1−5Y	2%	4%
	5Y+	4%	8%
A+ to BBB−	0−1Y	1%	2%
	1−5Y	3%	6%
	5Y+	6%	12%
BB+ to BB−		15%	
Cash		0%	
Gold		15%	
Main index equities		15%	
Equities listed on a recognized exchange		25%	
FX risk		8%	

Example 27 *We consider a 10-year credit of* \$100 *mn to a corporate firm rated* A. *The credit is guaranteed by five collateral instruments: a cash deposit (*\$2 *mn), a gold deposit (*\$5 *mn), a sovereign bond rated* AA *with a 2-year residual maturity (*\$15 *mn) and repurchase transactions on Microsoft stocks (*\$20 *mn) and Wirecard*[51] *stocks (*\$20 *mn).*

Before credit risk mitigation, the risk-weighted asset amount is equal to:

$$\text{RWA} = 100 \times 50\% = \$50 \text{ mn}$$

If we consider the simple approach, the repurchase transaction on Wirecard stocks is not eligible, because it does not fall within categories (1)-(6). The risk-weighted asset amount

[51] Wirecard is a German financial company specialized in payment processing and issuing services. The stock belongs to the MSCI Small Cap Europe index.

becomes[52]:

$$\begin{aligned}\text{RWA} &= (100-2-5-15-20)\times 50\% + (2+5+15+20)\times 20\% \\ &= \$37.40 \text{ mn}\end{aligned}$$

The repurchase transaction on Wirecard stocks is eligible in the comprehensive approach, because these equity stocks are traded in Börse Frankfurt. The haircuts are 15% for gold, 2% for the sovereign bond and 15% for Microsoft stocks[53]. For Wirecard stocks, a first haircut of 25% is applied because this instrument belongs to the seventh category and a second haircut of 8% is applied because there is a foreign exchange risk. The adjusted exposure at default is then equal to:

$$\begin{aligned}\text{EAD}^\star &= (1+8\%)\times 100 - 2 - (1-15\%)\times 5 - (1-2\%)\times 15 - \\ &\quad (1-15\%)\times 20 - (1-25\%-8\%)\times 20 \\ &= \$73.65 \text{ mn}\end{aligned}$$

It follows that:

$$\text{RWA} = 73.65 \times 50\% = \$36.82 \text{ mn}$$

Guarantees and credit derivatives Banks can use these credit protection instruments if they are direct, explicit, irrevocable and unconditional. In this case, banks use the simple approach given by Equation (3.16). The case of credit default tranches is covered by rules described in the securitization framework.

Maturity mismatches A maturity mismatch occurs when the residual maturity of the hedge is less than that of the underlying asset. In this case, the bank uses the following adjustment:

$$C_A = C \cdot \frac{\min(T_G, T, 5) - 0.25}{\min(T, 5) - 0.25} \tag{3.18}$$

where T is the residual maturity of the exposure and T_G is the residual maturity of the collateral (or guarantee).

Example 28 *The bank A has granted a credit of* \$30 *mn to a corporate firm B, which is rated BB. In order to hedge the default risk, the bank A buys* \$20 *mn of a 3-year CDS protection on B to the bank C, which is rated A+.*

If the residual maturity of the credit is lower than 3 years, we obtain:

$$\text{RWA} = (30-20)\times 100\% + 20 \times 50\% = \$20 \text{ mn}$$

If the residual maturity is greater than 3 years, we first have to calculate the adjusted value of the guarantee. Assuming that the residual maturity is 4 years, we have:

$$G_A = 20 \times \frac{\min(3,4,5) - 0.25}{\min(4,5) - 0.25} = \$14.67 \text{ mn}$$

It follows that:

$$\text{RWA} = (30-14.67)\times 100\% + 14.67 \times 50\% = \$22.67 \text{ mn}$$

[52]The floor of 20% is applied to the cash, gold and sovereign bond collateral instruments. The risk weight for Microsoft stocks is 20% because the rating of Microsoft is AAA.

[53]Because Microsoft belongs to the S&P 500 index, which is a main equity index.

3.2.3 The Basel II internal ratings-based approach

The completion of the internal ratings-based (IRB) approach was a complex task, because it required many negotiations between regulators, banks and politics. Tarullo (2008) points out that the publication of the first consultative paper (CP1) in June 1999 was both "*anticlimactic and contentious*". The paper is curiously vague without a precise direction. The only tangible proposal is the use of external ratings. The second consultative paper is released in January 2001 and includes in particular the IRB approach, which has been essentially developed by US members of the Basel Committee with the support of large international banks. The press release dated 16 January 2001 indicated that the Basel Committee would finalize the New Accord by the end of 2001, for an implementation in 2004. However, it has taken much longer than originally anticipated and the final version of the New Accord was published in June 2004 and implemented from December 2006[54]. The main reason is the difficulty of calibrating the IRB approach in order to satisfy a large part of international banks. The IRB formulas of June 2004 are significantly different from the original ones and reflect compromises between the different participants without really being satisfactory.

3.2.3.1 The general framework

Contrary to the standardized approach, the IRB approach is based on internal rating systems. With such a method, the objectives of the Basel Committee were to propose a more sensitive credit risk measure and define a common basis between internal credit risk models. The IRB approach may be seen as an external credit risk model, whose parameters are provided by the bank. Therefore, it is not an internal model, but a first step to harmonize the internal risk management practices by focusing on the main risk components, which are:

- the exposure at default (EAD);
- the probability of default (PD);
- the loss given default (LGD);
- the effective maturity (M).

The exposure at default is defined as the outstanding debt at the time of default. For instance, it is equal to the principal amount for a loan. The loss given default is the expected percentage of exposure at default that is lost if the debtor defaults. At first approximation, one can consider that LGD $\simeq 1 - \boldsymbol{\mathcal{R}}$, where $\boldsymbol{\mathcal{R}}$ is the recovery rate. While EAD is expressed in \$, LGD is measured in %. For example, if EAD is equal to \$10 mn and LGD is set to 70%, the expected loss due to the default is equal to \$7 mn. The probability of default measures the default risk of the debtor. In Basel II, the time horizon of PD is set to one year. When the duration of the credit is not equal to one year, one has to specify its effective maturity M. This is the combination of the one-year default probability PD and the effective maturity M that measures the default risk of the debtor until the duration of the credit.

In this approach, the credit risk measure is the sum of individual risk contributions:

$$\mathcal{R}(w) = \sum_{i=1}^{n} \mathcal{RC}_i$$

[54] See Chapter 4 entitled "*Negotiating Basel II*" of Tarullo (2008) for a comprehensive story of the Basel II Accord.

where $\mathcal{RC}_i$ is a function of the four risk components:

$$\mathcal{RC}_i = f_{\text{IRB}}\left(\text{EAD}_i, \text{LGD}_i, \text{PD}_i, \text{M}_i\right)$$

and f_{IRB} is the IRB fomula. In fact, there are two IRB methodologies. In the foundation IRB approach (FIRB), banks use their internal estimates of PD whereas the values of the other components (EAD, LGD and M) are set by regulators. Banks that adopt the advanced IRB approach (AIRB) may calculate all the four parameters (PD, EAD, LGD and M) using their own internal models and not only the probability of default. The mechanism of the IRB approach is then the following:

- a classification of exposures (sovereigns, banks, corporates, retail portfolios, etc.);
- for each credit i, the bank estimates the probability of default PD_i;
- it uses the standard regulatory values of the other risk components (EAD_i, LGD_i and M_i) or estimates them in the case of AIRB;
- the bank calculate then the risk-weighted assets RWA_i of the credit by applying the right IRB formula f_{IRB} to the risk components.

Internal ratings are central to the IRB approach. Table 3.23 gives an example of an internal rating system, where risk increases with the number grade (1, 2, 3, etc.). Another approach is to consider alphabetical letter grades[55]. A third approach is to use an internal rating scale similar to that of S&P[56].

3.2.3.2 The credit risk model of Basel II

Decomposing the value-at-risk into risk contributions BCBS (2004a) used the Merton-Vasicek model (Merton, 1974; Vasicek, 2002) to derive the IRB formula. In this framework, the portfolio loss is equal to:

$$L = \sum_{i=1}^{n} w_i \cdot \text{LGD}_i \cdot \mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\} \tag{3.19}$$

where w_i and T_i are the exposure at default and the residual maturity of the i^{th} credit. We assume that the loss given default LGD_i is a random variable and the default time $\boldsymbol{\tau}_i$ depends on a set of risk factors X, whose probability distribution is denoted by $\mathbf{H}$. Let $p_i(X)$ be the conditional default probability. It follows that the (unconditional or long-term) default probability is:

$$\begin{aligned} p_i &= \mathbb{E}_X\left[\mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\}\right] \\ &= \mathbb{E}_X\left[p_i(X)\right] \end{aligned}$$

We also introduce the notation $D_i = \mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\}$, which is the default indicator function. Conditionally to the risk factors X, D_i is a Bernoulli random variable with probability $p_i(X)$. If we consider the standard assumption that the loss given default is independent

[55] For instance, the rating system of Crédit Agricole is: A+, A, B+, B, C+, C, C-, D+, D, D-, E+, E and E- (source: Credit Agricole, Annual Financial Report 2014, page 201).

[56] This is the case of JPMorgan Chase & Co. (source: JPMorgan Chase & Co., Annual Report 2014, page 104).

TABLE 3.23: An example of internal rating system

Rating	Degree of risk	Definition	Borrower category by self-assessment
1	No essential risk	Extremely high degree of certainty of repayment	Normal
2	Negligible risk	High degree of certainty of repayment	
3	Some risk	Sufficient certainty of repayment	
4 A 4 B 4 C	Better than average	There is certainty of repayment but substantial changes in the environment in the future may have some impact on this uncertainty	
5 A 5 B 5 C	Average	There are no problems foreseeable in the future, but a strong likelihood of impact from changes in the environment	
6 A 6 B 6 C	Tolerable	There are no problems foreseeable in the future, but the future cannot be considered entirely safe	
7	Lower than average	There are no problems at the current time but the financial position of the borrower is relatively weak	
8 A 8 B	Needs preventive management	There are problems with lending terms or fulfilment, or the borrower's business conditions are poor or unstable, or there are other factors requiring careful management	Needs attention
9	Needs serious management	There is a high likelihood of bankruptcy in the future	In danger of bankruptcy
10 I		The borrower is in serious financial straits and "effectively bankrupt"	Effectively bankruptcy
10 II		The borrower is bankrupt	Bankrupt

Source: Ieda *et al.* (2000).

from the default time and we also assume that the default times are conditionally independent[57], we obtain:

$$\begin{aligned}\mathbb{E}\left[L \mid X\right] &= \sum_{i=1}^{n} w_i \cdot \mathbb{E}\left[\mathrm{LGD}_i\right] \cdot \mathbb{E}\left[D_i \mid X\right] \\ &= \sum_{i=1}^{n} w_i \cdot \mathbb{E}\left[\mathrm{LGD}_i\right] \cdot p_i\left(X\right) \end{aligned} \tag{3.20}$$

and[58]:

$$\begin{aligned}\sigma^2\left(L \mid X\right) &= \mathbb{E}\left[L^2 \mid X\right] - \mathbb{E}^2\left[L \mid X\right] \\ &= \sum_{i=1}^{n} w_i^2 \cdot \left(\mathbb{E}\left[\mathrm{LGD}_i^2\right] \cdot \mathbb{E}\left[D_i^2 \mid X\right] - \mathbb{E}^2\left[\mathrm{LGD}_i\right] \cdot p_i^2\left(X\right)\right)\end{aligned}$$

[57]The default times are not independent, because they depend on the common risk factors X. However, conditionally to these factors, they become independent because idiosyncratic risk factors are not correlated.

[58]Because the conditional covariance between D_i and D_j is equal to zero. The derivation of this formula is given in Exercise 3.4.8 on page 255.

We have $\mathbb{E}\left[D_i^2 \mid X\right] = p_i(X)$ and $\mathbb{E}\left[\text{LGD}_i^2\right] = \sigma^2(\text{LGD}_i) + \mathbb{E}^2[\text{LGD}_i]$. We deduce that:

$$\sigma^2(L \mid X) = \sum_{i=1}^{n} w_i^2 \cdot A_i \tag{3.21}$$

where:

$$A_i = \mathbb{E}^2[\text{LGD}_i] \cdot p_i(X) \cdot (1 - p_i(X)) + \sigma^2(\text{LGD}_i) \cdot p_i(X)$$

BCBS (2004a) assumes that the portfolio is infinitely fine-grained, which means that there is no concentration risk:

$$\lim_{n \to \infty} \max \frac{w_i}{\sum_{j=1}^{n} w_j} = 0 \tag{3.22}$$

In this case, Gordy (2003) shows that the conditional distribution of L degenerates to its conditional expectation $\mathbb{E}[L \mid X]$. The intuition of this result is given by Wilde (2001a). He considers a fine-grained portfolio equivalent to the original portfolio by replacing the original credit i by m credits with the same default probability p_i, the same loss given default LGD_i but an exposure at default divided by m. Let L_m be the loss of the equivalent fine-grained portfolio. We have:

$$\begin{aligned}
\mathbb{E}[L_m \mid X] &= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \frac{w_i}{m} \right) \cdot \mathbb{E}[\text{LGD}_i] \cdot \mathbb{E}[D_i \mid X] \\
&= \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(X) \\
&= \mathbb{E}[L \mid X]
\end{aligned}$$

and:

$$\begin{aligned}
\sigma^2(L_m \mid X) &= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \frac{w_i^2}{m^2} \right) \cdot A_i \\
&= \frac{1}{m} \sum_{i=1}^{n} w_i^2 \cdot A_i \\
&= \frac{1}{m} \sigma^2(L_m \mid X)
\end{aligned}$$

When m tends to ∞, we obtain the infinitely fine-grained portfolio. We note that $\mathbb{E}[L_\infty \mid X] = \mathbb{E}[L \mid X]$ and $\sigma^2(L_\infty \mid X) = 0$. Conditionally to the risk factors X, the portfolio loss L_∞ is equal to the conditional mean $\mathbb{E}[L \mid X]$. The associated probability distribution $\mathbf{F}$ is then:

$$\begin{aligned}
\mathbf{F}(\ell) &= \Pr\{L_\infty \leq \ell\} \\
&= \Pr\{\mathbb{E}[L \mid X] \leq \ell\} \\
&= \Pr\left\{ \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(X) \leq \ell \right\}
\end{aligned}$$

Let $g(x)$ be the function $\sum_{i=1}^{n} w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(x)$. We have:

$$\mathbf{F}(\ell) = \int \cdots \int \mathbb{1}\{g(x) \leq \ell\} \, \mathrm{d}\mathbf{H}(x)$$

However, it is not possible to obtain a closed-form formula for the value-at-risk $\mathbf{F}^{-1}(\alpha)$ defined as follows:

$$\mathbf{F}^{-1}(\alpha) = \{\ell : \Pr\{g(X) \leq \ell\} = \alpha\}$$

If we consider a single risk factor and assume that $g(x)$ is an increasing function, we obtain:

$$\begin{aligned} \Pr\{g(X) \leq \ell\} = \alpha &\Leftrightarrow \Pr\left\{X \leq g^{-1}(\ell)\right\} = \alpha \\ &\Leftrightarrow \mathbf{H}\left(g^{-1}(\ell)\right) = \alpha \\ &\Leftrightarrow \ell = g\left(\mathbf{H}^{-1}(\alpha)\right) \end{aligned}$$

We finally deduce that the value-at-risk has the following expression:

$$\begin{aligned} \mathbf{F}^{-1}(\alpha) &= g\left(\mathbf{H}^{-1}(\alpha)\right) \\ &= \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\mathrm{LGD}_i] \cdot p_i\left(\mathbf{H}^{-1}(\alpha)\right) \end{aligned} \tag{3.23}$$

Equation (3.23) is appealing because the value-at-risk satisfies the Euler allocation principle. Indeed, we have:

$$\begin{aligned} \mathcal{RC}_i &= w_i \cdot \frac{\partial \mathbf{F}^{-1}(\alpha)}{\partial w_i} \\ &= w_i \cdot \mathbb{E}[\mathrm{LGD}_i] \cdot p_i\left(\mathbf{H}^{-1}(\alpha)\right) \end{aligned} \tag{3.24}$$

and:

$$\sum_{i=1}^{n} \mathcal{RC}_i = \mathbf{F}^{-1}(\alpha)$$

Remark 30 *If $g(x)$ is a decreasing function, we obtain $\Pr\left\{X \geq g^{-1}(\ell)\right\} = \alpha$ and:*

$$\mathbf{F}^{-1}(\alpha) = \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\mathrm{LGD}_i] \cdot p_i\left(\mathbf{H}^{-1}(1-\alpha)\right)$$

The risk contribution becomes:

$$\mathcal{RC}_i = w_i \cdot \mathbb{E}[\mathrm{LGD}_i] \cdot p_i\left(\mathbf{H}^{-1}(1-\alpha)\right) \tag{3.25}$$

We reiterate that Equation (3.24) has been obtained under the following assumptions:

- $\mathcal{H}_1$ the loss given default LGD_i is independent from the default time $\boldsymbol{\tau}_i$;
- $\mathcal{H}_2$ the default times $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ depend on a single risk factor X and are conditionally independent with respect to X;
- $\mathcal{H}_3$ the portfolio is infinitely fine-grained, meaning that there is no exposure concentration.

Equation (3.24) is a very important result for two main reasons. First, it implies that, under the previous assumptions, the value-at-risk of an infinitely fine-grained portfolio can be decomposed as a sum of independent risk contributions. Indeed, $\mathcal{RC}_i$ depends solely on the characteristics of the i^{th} credit (exposure at default, loss given default and probability of default). This facilitates the calculation of the value-at-risk for large portfolios. Second, the risk contribution $\mathcal{RC}_i$ is related to the expected value of the loss given default. We don't need to model the probability distribution of LGD_i, only the mean $\mathbb{E}[\mathrm{LGD}_i]$ is taken into account.

Closed-form formula of the value-at-risk In order to obtain a closed-form formula, we need a model of default times. BCBS (2004a) has selected the one-factor model of Merton (1974), which has been formalized by Vasicek (1991). Let Z_i be the normalized asset value of the entity i. In the Merton model, the default occurs when Z_i is below a given barrier B_i:

$$D_i = 1 \Leftrightarrow Z_i < B_i$$

By assuming that Z_i is Gaussian, we deduce that:

$$\begin{aligned} p_i &= \Pr\{D_i = 1\} \\ &= \Pr\{Z_i < B_i\} \\ &= \Phi(B_i) \end{aligned}$$

The value of the barrier B_i is then equal to $\Phi^{-1}(p_i)$. We assume that the asset value Z_i depends on the common risk factor X and an idiosyncratic risk factor ε_i as follows:

$$Z_i = \sqrt{\rho} X + \sqrt{1-\rho}\varepsilon_i$$

X and ε_i are two independent standard normal random variables. We note that[59]:

$$\begin{aligned} \mathbb{E}[Z_i Z_j] &= \mathbb{E}\left[\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i\right)\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_j\right)\right] \\ &= \mathbb{E}\left[\rho X^2 + (1-\rho)\varepsilon_i\varepsilon_j + X\sqrt{\rho(1-\rho)}(\varepsilon_i + \varepsilon_j)\right] \\ &= \rho \end{aligned}$$

where ρ is the constant asset correlation. We now calculate the conditional default probability:

$$\begin{aligned} p_i(X) &= \Pr\{D_i = 1 \mid X\} \\ &= \Pr\{Z_i < B_i \mid X\} \\ &= \Pr\left\{\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i < B_i\right\} \\ &= \Pr\left\{\varepsilon_i < \frac{B_i - \sqrt{\rho}X}{\sqrt{1-\rho}}\right\} \\ &= \Phi\left(\frac{B_i - \sqrt{\rho}X}{\sqrt{1-\rho}}\right) \end{aligned}$$

Using the framework of the previous paragraph, we obtain:

$$\begin{aligned} g(x) &= \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(x) \\ &= \sum_{i=1}^{n} w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}x}{\sqrt{1-\rho}}\right) \end{aligned}$$

We note that $g(x)$ is a decreasing function if $w_i \geq 0$. Using Equation (3.25) and the relationship $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$, it follows that:

$$\mathcal{RC}_i = w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{3.26}$$

[59] We have $\mathbb{E}[\varepsilon_i\varepsilon_j] = 0$ because ε_i and ε_j are two specific risk factors.

Remark 31 *We verify that p_i is the unconditional default probability. Indeed, we have:*

$$\begin{aligned}\mathbb{E}_X\left[p_i\left(X\right)\right] &= \mathbb{E}_X\left[\Phi\left(\frac{\Phi^{-1}\left(p_i\right)-\sqrt{\rho}X}{\sqrt{1-\rho}}\right)\right]\\ &= \int_{-\infty}^{\infty}\Phi\left(\frac{\Phi^{-1}\left(p_i\right)-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)\phi\left(x\right)\,\mathrm{d}x\end{aligned}$$

We recognize the integral function analyzed in Appendix A.2.2.5 on page 1063. We deduce that:

$$\begin{aligned}\mathbb{E}_X\left[p_i\left(X\right)\right] &= \Phi_2\left(\infty,\frac{\Phi^{-1}\left(p_i\right)}{\sqrt{1-\rho}}\cdot\left(\frac{1}{1-\rho}\right)^{-1/2};\frac{\sqrt{\rho}}{\sqrt{1-\rho}}\left(\frac{1}{1-\rho}\right)^{-1/2}\right)\\ &= \Phi_2\left(\infty,\Phi^{-1}\left(p_i\right);\sqrt{\rho}\right)\\ &= \Phi\left(\Phi^{-1}\left(p_i\right)\right)\\ &= p_i\end{aligned}$$

Example 29 *We consider a homogeneous portfolio with* 100 *credits. For each credit, the exposure at default, the expected LGD and the probability of default are set to* \$1 *mn,* 50% *and* 5%*.*

Let us assume that the asset correlation ρ is equal to 10%. We have reported the numerical values of $\mathbf{F}^{-1}\left(\alpha\right)$ for different values of α in Table 3.24. If we are interested in the cumulative distribution function, $\mathbf{F}\left(\ell\right)$ is equal to the numerical solution α of the equation $\mathbf{F}^{-1}\left(\alpha\right)=\ell$. Using a bisection algorithm, we find the probabilities given in Table 3.24. For instance, the probability to have a loss less than or equal to \$3 mn is equal to 70.44%. Finally, to calculate the probability density function of the portfolio loss, we use the following relationship[60]:

$$f\left(x\right)=\frac{1}{\partial_\alpha\,\mathbf{F}^{-1}\left(\mathbf{F}\left(x\right)\right)}$$

where:

$$\begin{aligned}\partial_\alpha\,\mathbf{F}^{-1}\left(\alpha\right) &= \sum_{i=1}^{n}w_i\cdot\mathbb{E}\left[\mathrm{LGD}_i\right]\cdot\sqrt{\frac{\rho}{1-\rho}}\cdot\frac{1}{\phi\left(\Phi^{-1}\left(\alpha\right)\right)}\cdot\\ &\quad\phi\left(\frac{\Phi^{-1}\left(p_i\right)+\sqrt{\rho}\Phi^{-1}\left(\alpha\right)}{\sqrt{1-\rho}}\right)\end{aligned}$$

In Figure 3.21, we compare the probability functions for two different values of the asset correlation ρ. We note that the level of ρ has a big impact on the quantile function and the shape of the density function.

TABLE 3.24: Numerical values of $f\left(\ell\right)$, $\mathbf{F}\left(\ell\right)$ and $\mathbf{F}^{-1}\left(\alpha\right)$ when ρ is equal to 10%

ℓ	(in \$ mn)	0.10	1.00	2.00	3.00	4.00	5.00
$\mathbf{F}\left(\ell\right)$	(in %)	0.03	16.86	47.98	70.44	83.80	91.26
$f\left(\ell\right)$	(in %)	1.04	31.19	27.74	17.39	9.90	5.43
α	(in %)	10.00	25.00	50.00	75.00	90.00	95.00
$\mathbf{F}^{-1}\left(\alpha\right)$	(in \$ mn)	0.77	1.25	2.07	3.28	4.78	5.90

[60] See Appendix A.2.2.3 on page 1062.

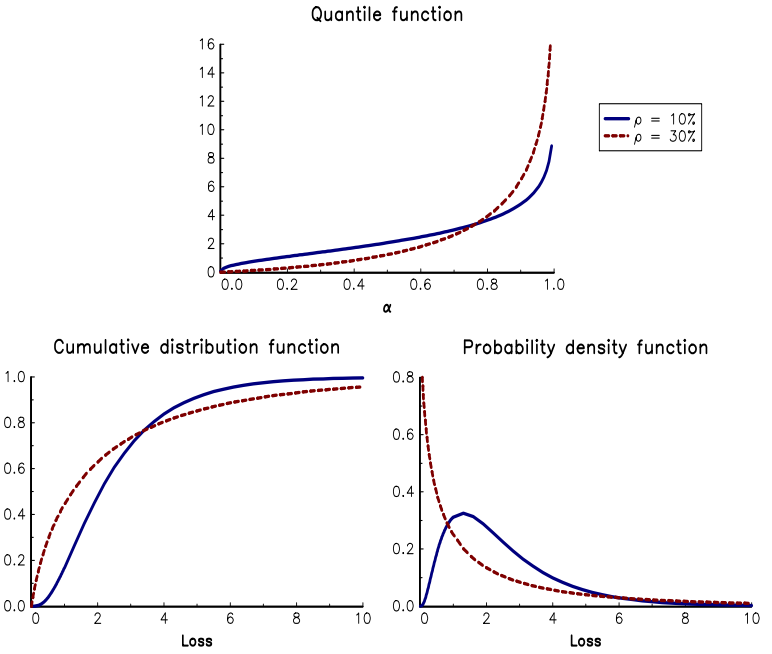


FIGURE 3.21: Probability functions of the credit portfolio loss

The risk contribution $\mathcal{RC}_i$ depends on three credit parameters (the exposure at default w_i, the expected loss given default $\mathbb{E}[\text{LGD}_i]$ and the probability of default p_i) and two model parameters (the asset correlation ρ and the confidence level α of the value-at-risk). It is obvious that $\mathcal{RC}_i$ is an increasing function of the different parameters with the exception of the correlation. We obtain:

$$\text{sign}\,\frac{\partial\,\mathcal{RC}_i}{\partial\,\rho} = \text{sign}\,\frac{1}{2\,(1-\rho)^{3/2}}\left(\Phi^{-1}(p_i) + \frac{\Phi^{-1}(\alpha)}{\sqrt{\rho}}\right)$$

We deduce that the risk contribution is not a monotone function with respect to ρ. It increases if the term $\sqrt{\rho}\Phi^{-1}(p_i)+\Phi^{-1}(\alpha)$ is positive. This implies that the risk contribution may decrease if the probability of default is very low and the confidence level is larger than 50%. The two limiting cases are $\rho = 0$ and $\rho = 1$. In the first case, the risk contribution is equal to the expected loss:

$$\mathcal{RC}_i = \mathbb{E}[L_i] = w_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i$$

In the second case, the risk contribution depends on the value of the probability of default:

$$\lim_{\rho\to 1} \mathcal{RC}_i = \begin{cases} 0 & \text{if } p_i < 1-\alpha \\ 0.5 \cdot w_i \cdot \mathbb{E}[\text{LGD}_i] & \text{if } p_i = 1-\alpha \\ w_i \cdot \mathbb{E}[\text{LGD}_i] & \text{if } p_i > 1-\alpha \end{cases}$$

The behavior of the risk contribution is illustrated in Figure 3.22 with the following base parameter values: $w_i = 100$, $\mathbb{E}[\text{LGD}_i] = 70\%$, $\rho = 20\%$ and $\alpha = 90\%$. We verify that the risk contribution is an increasing function of $\mathbb{E}[\text{LGD}_i]$ (top/left panel) and α (top/right panel). When p_i and α are set to 10% and 90%, the risk contribution increases with ρ and reaches the value 35, which corresponds to half of the nominal loss given default. When p_i and α are set to 5% and 90%, the risk contribution increases in a first time and then

decreases (bottom/left panel). The maximum is reached for the value[61] $\rho^\star = 60.70\%$. When α is equal to 99%, this behavior vanishes (bottom/right panel).

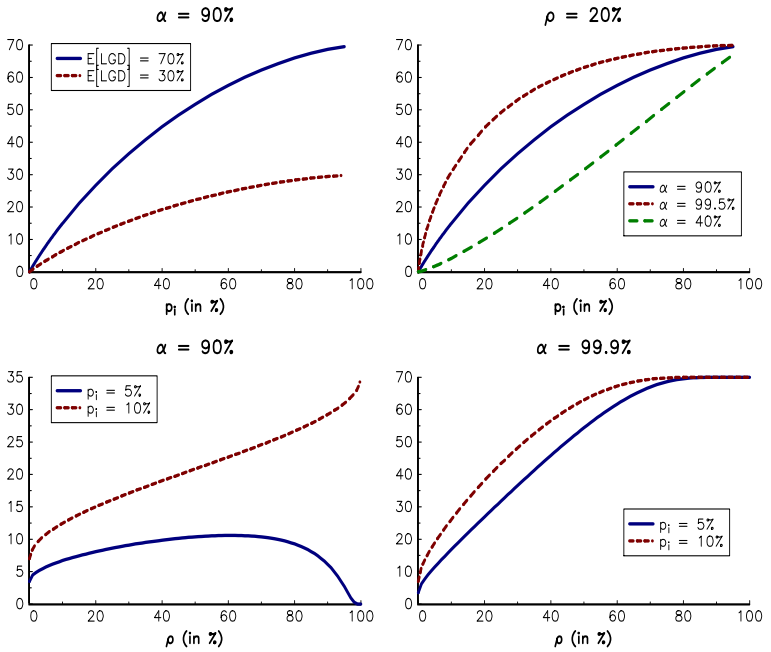


FIGURE 3.22: Relationship between the risk contribution $\mathcal{RC}_i$ and model parameters

In this model, the maturity T_i is taken into account through the probability of default. Indeed, we have $p_i = \Pr\{\boldsymbol{\tau}_i \leq T_i\}$. Let us denote PD_i the annual default probability of the obligor. If we assume that the default time is Markovian, we have the following relationship:

$$\begin{aligned} p_i &= 1 - \Pr\{\boldsymbol{\tau}_i > T_i\} \\ &= 1 - (1 - \text{PD}_i)^{T_i} \end{aligned}$$

We can then rewrite Equation (3.26) such that the risk contribution depends on the exposure at default, the expected loss given default, the annualized probability of default and the maturity, which are the 4 parameters of the IRB approach.

3.2.3.3 The IRB formulas

A long process to obtain the finalized formulas The IRB formula of the second consultative portfolio was calibrated with $\alpha = 99.5\%$, $\rho = 20\%$ and a standard maturity of three years. To measure the impact of this approach, the Basel Committee conducted a quantitative impact study (QIS) in April 2001. A QIS is an Excel workbook to be filled by the bank. It allows the Basel Committee to gauge the impact of the different proposals for capital requirements. The answers are then gathered and analyzed at the industry level. Results were published in November 2001. Overall, 138 banks from 25 countries participated in the QIS. Not all participating banks managed to calculate the capital requirements under the

[61] We have:

$$\rho^\star = \max^2\left(0, -\frac{\Phi^{-1}(\alpha)}{\Phi^{-1}(p_i)}\right) = \left(\frac{1.282}{1.645}\right)^2 = 60.70\%$$

three methods (SA, FIRB and AIRB). However, 127 banks provided complete information on the SA approach and 55 banks on the FIRB approach. Only 22 banks were able to calculate the AIRB approach for all portfolios.

TABLE 3.25: Percentage change in capital requirements under CP2 proposals

		SA	FIRB	AIRB
G10	Group 1	6%	14%	−5%
	Group 2	1%		
EU	Group 1	6%	10%	−1%
	Group 2	−1%		
Others		5%		

Source: BCBS (2001b).

In Table 3.25, we report the difference in capital requirements between CP2 proposals and Basel I. Group 1 corresponds to diversified, internationally active banks with tier 1 capital of at least €3 bn whereas Group 2 consists of smaller or more specialized banks. BCBS (2001b) concluded that "*on average, the QIS2 results indicate that the CP2 proposals for credit risk would deliver an increase in capital requirements for all groups under both the SA and FIRB approaches*". It was obvious that these figures were not satisfactory. The Basel Committee considered then several modifications in order to (1) maintain equivalence on average between current required capital and the revised SA approach and (2) provide incentives under the FIRB approach. A third motivation has emerged rapidly. According to many studies[62], Basel II may considerably increase the procyclicality of capital requirements. Indeed, capital requirements may increase in an economic meltdown, because LGD increases in bad times and credits receive lower ratings. In this case, capital requirements may move in an opposite direction than the macroeconomic cycle, leading banks to reduce their supply of credit during a crisis. In this scenario, Basel II proposals may amplify credit crises and economic downturns. All these reasons explain the long period to finalize the Basel II Accord. After two new QIS (QIS 2.5 in July 2002 and QIS 3 in May 2003) and a troubled period at the end of 2003, the new Capital Accord is finally published in June 2004. However, there was a shared feeling that it was more a compromise than a terminated task. Thus, several issues remained unresolved and two new QIS will be conducted in 2004 and 2005 before the implementation in order to confirm the calibration.

The supervisory formula If we use the notations of the Basel Committee, the risk contribution has the following expression:

$$\mathcal{RC} = \text{EAD} \cdot \text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}\left(1-(1-\text{PD})^{\text{M}}\right)+\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$$

where EAD is the exposure at default, LGD is the (expected) loss given default, PD is the (one-year) probability of default and M is the effective maturity. Because $\mathcal{RC}$ is directly the capital requirement ($\mathcal{RC} = 8\% \times \text{RWA}$), we deduce that the risk-weighted asset amount is equal to:

$$\text{RWA} = 12.50 \cdot \text{EAD} \cdot \boldsymbol{\mathcal{K}}^{\star} \tag{3.27}$$

[62] See for instance Goodhart *et al.* (2004) or Kashyap and Stein (2004).

where $\mathcal{K}^\star$ is the normalized required capital for a unit exposure:

$$\mathcal{K}^\star = \text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}\left(1-(1-\text{PD})^{\text{M}}\right)+\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{3.28}$$

In order to obtain the finalized formulas, the Basel Committee has introduced the following modifications:

- a maturity adjustment $\varphi(\text{M})$ has been added in order to separate the impact of the one-year probability of default and the effect of the maturity; the function $\varphi(\text{M})$ has then been calibrated such that Expression (3.28) becomes:

$$\mathcal{K}^\star \approx \text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}(\text{PD})+\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \cdot \varphi(\text{M}) \tag{3.29}$$

- it has used a confidence level of 99.9% instead of the 99.5% value;
- it has defined a parametric function $\rho(\text{PD})$ for the default correlation in order that low ratings are not too penalizing for capital requirements;
- it has considered the unexpected loss as the credit risk measure:

$$\text{UL}_\alpha = \text{VaR}_\alpha - \mathbb{E}[L]$$

In summary, the risk-weighted asset amount in the IRB approach is calculated using Equation (3.27) and the following normalized required capital:

$$\mathcal{K}^\star = \left(\text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}(\text{PD})+\sqrt{\rho(\text{PD})}\Phi^{-1}(99.9\%)}{\sqrt{1-\rho(\text{PD})}}\right) - \text{LGD} \cdot \text{PD}\right) \cdot \varphi(\text{M}) \tag{3.30}$$

Risk-weighted assets for corporate, sovereign, and bank exposures The three asset classes use the same formula:

$$\begin{aligned}\mathcal{K}^\star &= \left(\text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}(\text{PD})+\sqrt{\rho(\text{PD})}\Phi^{-1}(99.9\%)}{\sqrt{1-\rho(\text{PD})}}\right) - \text{LGD} \cdot \text{PD}\right) \cdot \\ &\left(\frac{1+(\text{M}-2.5)\cdot b(\text{PD})}{1-1.5\cdot b(\text{PD})}\right)\end{aligned} \tag{3.31}$$

with $b(\text{PD}) = (0.11852 - 0.05478 \cdot \ln(\text{PD}))^2$ and:

$$\rho(\text{PD}) = 12\% \times \left(\frac{1-e^{-50\times\text{PD}}}{1-e^{-50}}\right) + 24\% \times \left(1 - \frac{1-e^{-50\times\text{PD}}}{1-e^{-50}}\right) \tag{3.32}$$

We note that the maturity adjustment $\varphi(\text{M})$ vanishes when the effective maturity is one year. For a defaulted exposure, we have:

$$\mathcal{K}^\star = \max(0, \text{LGD} - \text{EL})$$

where EL is the bank's best estimate of the expected loss[63].

[63] We can assimilate it to specific provisions.

For small and medium-sized enterprises[64], a firm-size adjustment is introduced by defining a new parametric function for the default correlation:

$$\rho^{\text{SME}}(\text{PD}) = \rho(\text{PD}) - 0.04 \cdot \left(1 - \frac{(\max(S, 5) - 5)}{45}\right)$$

where S is the reported sales expressed in € mn. This adjustment has the effect to reduce the default correlation and then the risk-weighted assets. Similarly, the Basel Committee proposes specific arrangements for specialized lending and high-volatility commercial real estate (HVCRE).

In the foundation IRB approach, the bank estimates the probability of default, but uses standard values for the other parameters. In the advanced IRB approach, the bank always estimates the parameters PD and M, and may use its own estimates for the parameters EAD and LGD subject to certain minimum requirements. The risk components are defined as follows:

1. The exposure at default is the amount of the claim, without taking into account specific provisions or partial write-offs. For off-balance sheet positions, the bank uses similar credit conversion factors for the FIRB approach as for the SA approach. In the AIRB approach, the bank may use its own internal measures of CCF.

2. In the FIRB approach, the loss given default is set to 45% for senior claims and 75% for subordinated claims. In the AIRB approach, the bank may use its own estimates of LGD. However, they must be conservative and take into account adverse economic conditions. Moreover, they must include all recovery costs (litigation cost, administrative cost, etc.).

3. PD is the one-year probability of default calculated with the internal rating system. For corporate and bank exposures, a floor of 3 bps is applied.

4. The maturity is set to 2.5 years in the FIRB approach. In the advanced approach, M is the weighted average time of the cash flows, with a one-year floor and a five-year cap.

Example 30 *We consider a senior debt of* \$3 *mn on a corporate firm. The residual maturity of the debt is equal to* 2 *years. We estimate the one-year probability of default at* 5%.

To determine the capital charge, we first calculate the default correlation:

$$\begin{aligned}\rho(\text{PD}) &= 12\% \times \left(\frac{1 - e^{-50 \times 0.05}}{1 - e^{-50}}\right) + 24\% \times \left(1 - \frac{1 - e^{-50 \times 0.05}}{1 - e^{-50}}\right) \\ &= 12.985\%\end{aligned}$$

We have:

$$\begin{aligned}b(\text{PD}) &= (0.11852 - 0.05478 \times \ln(0.05))^2 \\ &= 0.0799\end{aligned}$$

It follows that the maturity adjustment is equal to:

$$\begin{aligned}\varphi(\text{M}) &= \frac{1 + (2 - 2.5) \times 0.0799}{1 - 1.5 \times 0.0799} \\ &= 1.0908\end{aligned}$$

[64] They are defined as corporate entities where the reported sales for the consolidated group of which the firm is a part is less than € 50 mn.

The normalized capital charge with a one-year maturity is:

$$\begin{aligned}\mathcal{K}^\star &= 45\% \times \Phi\left(\frac{\Phi^{-1}(5\%) + \sqrt{12.985\%}\Phi^{-1}(99.9\%)}{\sqrt{1-12.985\%}}\right) - 45\% \times 5\% \\ &= 0.1055\end{aligned}$$

When the maturity is two years, we obtain:

$$\begin{aligned}\mathcal{K}^\star &= 0.1055 \times 1.0908 \\ &= 0.1151\end{aligned}$$

We deduce the value taken by the risk weight:

$$\begin{aligned}\text{RW} &= 12.5 \times 0.1151 \\ &= 143.87\%\end{aligned}$$

It follows that the risk-weighted asset amount is equal to \$4.316 mn whereas the capital charge is \$345 287. Using the same process, we have calculated the risk weight for different values of PD, LGD and M in Table 3.26. The last two columns are for a SME claim by considering that sales are equal to €5 mn.

TABLE 3.26: IRB risk weights (in %) for corporate exposures

Maturity		M = 1		M = 2.5		M = 2.5 (SME)	
LGD		45%	75%	45%	75%	45%	75%
PD (in %)	0.10	18.7	31.1	29.7	49.4	23.3	38.8
	0.50	52.2	86.9	69.6	116.0	54.9	91.5
	1.00	73.3	122.1	92.3	153.9	72.4	120.7
	2.00	95.8	159.6	114.9	191.4	88.5	147.6
	5.00	131.9	219.8	149.9	249.8	112.3	187.1
	10.00	175.8	292.9	193.1	321.8	146.5	244.2
	20.00	223.0	371.6	238.2	397.1	188.4	314.0

Risk-weighted assets for retail exposures Claims can be included in the regulatory retail portfolio if they meet certain criteria: in particular, the exposure must be to an individual person or to a small business; it satisfies the granularity criterion, meaning that no aggregate exposure to one counterpart can exceed 0.2% of the overall regulatory retail portfolio; the aggregated exposure to one counterparty cannot exceed €1 mn. In these cases, the bank uses the following IRB formula:

$$\mathcal{K}^\star = \text{LGD} \cdot \Phi\left(\frac{\Phi^{-1}(\text{PD}) + \sqrt{\rho(\text{PD})}\Phi^{-1}(99.9\%)}{\sqrt{1-\rho(\text{PD})}}\right) - \text{LGD} \cdot \text{PD} \tag{3.33}$$

We note that this IRB formula correspond to a one-year fixed maturity. The value of the default correlation depends on the categories. For residential mortgage exposures, we have $\rho(\text{PD}) = 15\%$ whereas the default correlation $\rho(\text{PD})$ is equal to 4% for qualifying revolving retail exposures. For other retail exposures, it is defined as follows:

$$\rho(\text{PD}) = 3\% \times \left(\frac{1-e^{-35\times\text{PD}}}{1-e^{-35}}\right) + 16\% \times \left(1 - \frac{1-e^{-35\times\text{PD}}}{1-e^{-35}}\right) \tag{3.34}$$

In Table 3.27, we report the corresponding risk weights for the three categories and for two different values of LGD.

TABLE 3.27: IRB risk weights (in %) for retail exposures

LGD		Mortgage		Revolving		Other retail	
		45%	25%	45%	85%	45%	85%
PD (in %)	0.10	10.7	5.9	2.7	5.1	11.2	21.1
	0.50	35.1	19.5	10.0	19.0	32.4	61.1
	1.00	56.4	31.3	17.2	32.5	45.8	86.5
	2.00	87.9	48.9	28.9	54.6	58.0	109.5
	5.00	148.2	82.3	54.7	103.4	66.4	125.5
	10.00	204.4	113.6	83.9	158.5	75.5	142.7
	20.00	253.1	140.6	118.0	222.9	100.3	189.4

The other two pillars The first pillar of Basel II, which concerns minimum capital requirements, is completed by two new pillars. The second pillar is the supervisory review process (SRP) and is composed of two main processes: the supervisory review and evaluation process (SREP) and the internal capital adequacy assessment process (ICAAP). The SREP defines the regulatory response to the first pillar, in particular the validation processes of internal models. Nevertheless, the SREP is not limited to capital requirements. More generally, the SREP evaluates the global strategy and resilience of the bank. ICAAP addresses risks that are not captured in Pillar 1 like concentration risk or non-granular portfolios in the case of credit risk[65]. For instance, stress tests are part of Pillar 2. The goal of the second pillar is then to encourage banks to continuously improve their internal models and processes for assessing the adequacy of their capital and to ensure that supervisors have the adequate tools to control them. The third pillar, which is also called market discipline, requires banks to publish comprehensive information about their risk management process. This is particularly true since the publication in January 2015 of the revised Pillar 3 disclosure requirements. Indeed, BCBS (2015a) imposes the use of templates for quantitative disclosure with a fixed format in order to facilitate the comparison between banks.

3.2.4 The Basel III revision

For credit risk capital requirements, Basel III is close to the Basel II framework with some adjustments, which mainly concern the parameters[66]. Indeed, the SA and IRB methods continue to be the two approaches for computing the capital charge for credit risk.

3.2.4.1 The standardized approach

Risk-weighted exposures External credit ratings continue to be the backbone of the standardized approach in Basel III. Nevertheless, they are not the only tool for measuring the absolute riskiness of debtors and loans. First, the Basel Committee recognizes that external credit ratings are prohibited in some jurisdictions for computing regulatory capital. For example, this is the case of the United States, which had abandoned in 2010 the use of commercial credit ratings after the Dodd-Frank reform. Second, the Basel Committee links risk weights to the loan-to-value ratio (LTV) for some categories.

When external ratings are allowed[67], the Basel Committee defines a new table of risk weights, which is close to the Basel II table. In Table 3.28, we indicate the main categories and the risk weights associated to credit ratings. We notice that the risk weights for

[65] Since Basel III, ICAAP is completed by the internal liquidity adequacy assessment process (ILAAP).
[66] The Basel III framework for credit risk is described in BCBS (2017c).
[67] This method is called the external credit risk assessment approach (ECRA).

TABLE 3.28: Risk weights of the SA approach (ECRA, Basel III)

Rating		AAA to AA−	A+ to A−	BBB+ to BBB−	BB+ to B−	CCC+ to C	NR
Sovereigns		0%	20%	50%	100%	150%	100%
PSE	1	20%	50%	100%	100%	150%	100%
	2	20%	50%	50%	100%	150%	50%
MDB		20%	30%	50%	100%	150%	50%
Banks	2	20%	30%	50%	100%	150%	SCRA
	2 ST	20%	20%	20%	50%	150%	SCRA
	Covered	10%	20%	20%	50%	100%	(*)
Corporates		20%	50%	75%	100%	150%	100%
Retail					75%		

(*) For unrated covered bonds, the risk weight is generally half of the risk weight of the issuing bank.

sovereign exposures and non-central government public sector entities (PSE) are unchanged. The risk weights for multilateral development banks (MDB) continue to be related to the risk weights for banks. However, we notice that the first option is removed and we observe some differences for exposures to banks. First, the risk weight for the category A+/A− is reduced from 50% to 30%. Second, for unrated exposures, the standard figure of 50% is replaced by the standardized credit risk approach (SCRA). Third, the Basel Committee considers the special category of covered bonds, whose development has emerged after the 2008 Global Financial Crisis and the introduction of capital requirements for systemic risks[68]. For exposures to corporates, the Basel Committee uses the same scale than for other categories contrary to Basel II (see Table 3.16 on page 163). Finally, the risk weight for retail exposures remains unchanged.

The standardized credit risk approach (SCRA) must be used for all exposures to banks in two situations: (1) when the exposure is unrated; (2) when external credit ratings are prohibited. In this case, the bank must conduct a due diligence analysis in order to classify the exposures into three grades: A, B, and C. Grade A refers to the most solid banks, whose capital exceeds the minimum regulatory capital requirements, whereas Grade C refers to the most vulnerable banks. The risk weight is respectively equal to 40%, 75% and 150% (20%, 50% and 150% for short-term exposures).

When external credit ratings are prohibited, the risk weight of exposures to corporates is equal to 100% with two exceptions. A 65% risk weight is assigned to corporates, which can be considered investment grade (IG). For exposures to small and medium-sized enterprises, a 75% risk weight can be applied if the exposure can be classified in the retail category and 85% for the others.

The case of retail is particular because we have to distinguish real estate exposures and other retail exposures. By default, the risk weight is equal to 75% for this last category, which includes revolving credits, credit cards, consumer credit loans, auto loans, student loans, etc. For real estate exposures, the risk weights depend on the loan-to-value ratio (LTV). Suppose that someone borrows \$100 000 to purchase a house of \$150 000, the LTV ratio is 100 000/150 000 or 66.67%. This ratio is extensively used in English-speaking

[68] See Chapter 8 on page 453.

countries (e.g. the United States) to measure the risk of the loan. The idea is that the lender's haircut ($100 000 in our example) represents the lender risk. If the borrower defaults, the lender recovers the property, that will be sold. The risk is then to sell the property below the lender's haircut. The higher the LTV ratio, the riskier the loan is for the lender. In continental Europe, the risk of home property loans is measured by the ability of the borrower to repay the capital and service his debt. In this case, the risk of the loan is generally related to the income of the borrower. It is obvious that these two methods for assessing the credit risk are completely different and this explains the stress in Europe to adopt the LTV approach. In Table 3.29, we have reported the value of risk weights with respect to the LTV (expressed in %) in the case of residential real estate exposures. The Basel Committee considers two categories depending if the repayment depends on the cash flows generated by property (D) or not (ND). The risk weight ranges from 20% to 105% in Basel III, whereas it was equal to 35% in Basel II.

TABLE 3.29: Risk weights of the SA approach (ECRA, Basel III)

Residential real estate			Commercial real estate		
Cash flows	ND	D	Cash flows	ND	D
$\text{LTV} \leq 50$	20%	30%	$\text{LTV} \leq 60$	$\min(60\%, \text{RW}_C)$	70%
$50 < \text{LTV} \leq 60$	25%	35%			
$60 < \text{LTV} \leq 80$	30%	45%	$60 < \text{LTV} \leq 80$	RW_C	90%
$80 < \text{LTV} \leq 90$	40%	60%			
$90 < \text{LTV} \leq 100$	50%	75%	$\text{LTV} > 80$	RW_C	110%
$\text{LTV} > 100$	70%	105%			

The LTV ratio is also used to determine the risk weight of commercial real estate, land acquisition, development and construction exposures. Table 3.29 gives the risk weight for commercial real estate exposures. If the repayment does not depend on the cash flows generated by property (ND), we use the risk weight of the counterparty with a cap of 60%. If the repayment depends on the cash flows generated by the property (D), the risk weight ranges from 70% to 110%, whereas it was equal to 100% in Basel II. Commercial real estate exposures that do not meet specific qualitative requirements will be risk-weighted at 150%, which is also the default figure for land acquisition, development and construction exposures.

For off-balance sheet items, credit conversion factors (CCF) have been revised. They can take the values 10%, 20%, 40%, 50% and 100%. This is a more granular scale without the possibility to set the CCF to 0%. Generally speaking, the CCF values in Basel III are more conservative than in Basel II.

Credit risk mitigation The regulatory framework for credit risk mitigation techniques changes very little from Basel II to Basel III: the two methods remain the simple and comprehensive approaches; the treatment of maturity mismatches is the same; the formulas for computing the risk weighted assets are identical, etc. Minor differences concern the description of eligible financial collateral and the haircut parameters, which are given in Table 3.30. For instance, we see that the Basel Committee makes the distinction of issuers for debt securities between sovereigns, other issuers and securitization exposures. While the haircuts do not change for sovereign debt securities with respect to Basel II, the scale is more granular for the two other categories. Haircuts are also increased by 5% for gold and equity collateral instruments.

The major difference concerns the treatment of securities financing transactions (SFT) such as repo-style transactions, since the Basel Committee has developed a specific approach

TABLE 3.30: Standardized supervisory haircuts for collateralized transactions (Basel III)

Rating	Residual Maturity	Sovereigns	Others	Securitization exposures
AAA to AA−	0−1Y	0.5%	1%	2%
	1−3Y	2%	3%	8%
	3−5Y	2%	4%	8%
	5Y−10Y	4%	6%	16%
	10Y+	4%	12%	16%
A+ to BBB−	0−1Y	1%	2%	4%
	1−3Y	3%	4%	12%
	3−5Y	3%	6%	12%
	5Y−10Y	6%	12%	24%
	10Y+	6%	20%	24%
BB+ to BB−		15%		
Cash			0%	
Gold			20%	
Main index equities			20%	
Equities listed on a recognized exchange			30%	
FX risk			8%	

for calculating the modified exposure $\text{EAD}^\star$ of these instruments in the comprehensive approach (BCBS, 2017c, pages 43-47).

3.2.4.2 The internal ratings-based approach

The methodology of the IRB approach does not change with respect to Basel II, since the formulas are the same[69]. The only exception is the correlation parameter for bank exposures[70], which becomes:

$$\begin{aligned}\rho(\text{PD}) &= 1.25 \times \left(12\% \times \left(\frac{1-e^{-50\times\text{PD}}}{1-e^{-50}}\right) + 24\% \times \frac{1-\left(1-e^{-50\times\text{PD}}\right)}{1-e^{-50}}\right) \\ &= 15\% \times \left(\frac{1-e^{-50\times\text{PD}}}{1-e^{-50}}\right) + 30\% \times \left(\frac{1-\left(1-e^{-50\times\text{PD}}\right)}{1-e^{-50}}\right) \end{aligned} \tag{3.35}$$

Therefore, the correlation range for the bank category increases from $12\% - 24\%$ to $15\% - 30\%$. In fact, the main differences concern the computation of the LGD parameter, and the validation of the IRB approach, which is much more restrictive. For instance, the IRB approaches are not permitted for exposures to equities, and we cannot develop an AIRB approach for exposures to banks and exposures to corporates with annual revenues greater than €500 mn. For banks and large corporates, only the FIRB approach is available.

The Basel Committee still considers five asset classes: corporates, sovereigns, banks, retail and equities. In the FIRB approach, the bank estimates the PD parameter, while

[69] This concerns Equation (3.27) for risk-weighted assets, Equations (3.31) and (3.32) for corporate, sovereign, and bank exposures, Equations (3.33) and (3.34) for retail exposures, the maturity adjustment $b(\text{PD})$, the correlation formula $\rho^{\text{SME}}(\text{PD})$ for SME exposures, the correlation parameters for retail exposures, etc.

[70] The multiplier of 1.25 is applied for regulated financial institutions with a total asset larger than $100 bn and all unregulated financial institutions.

it uses the regulatory estimates of EAD, LGD and M[71]. In the AIRB approach, the bank estimates all the parameters, but they are subject to some input floors. For example, the minimum PD is set to 5 bps for corporate and bank exposures.

Certainly, LGD is the most challenging parameter in Basel III. In the FIRB approach, the default values are 75% for subordinated claims, 45% for senior claims on financial institutions and 40% for senior claims on corporates. When considering a collateral, the LGD parameter becomes:

$$\text{LGD}_\star = \omega \cdot \text{LGD} + (1-\omega) \cdot \text{LGD}_C$$

where LGD and LGD_C apply to the unsecured exposure and the collateralized part, and ω is the relative weight between LGD and LGD_C:

$$\omega = 1 - \frac{(1-H_C) \cdot C}{(1+H_E) \cdot \text{EAD}}$$

Here, H_E is the SA haircut for the exposure, C is the value of the collateral, and H_C is the specific haircut for the collateral. LGD_C is equal to 0% for financial collateral, 20% for receivables and real estate and 25% for other physical collateral, whereas H_C can be from 0% to 100%. In the AIRB approach, the LGD parameter may be estimated by the bank, under the constraint that it is greater than the input floor $\text{LGD}^{\text{Floor}}$. For unsecured exposures, we have $\text{LGD} \geq \text{LGD}^{\text{Floor}}$ where $\text{LGD}^{\text{Floor}} = 25\%$. For secured exposures, we have $\text{LGD}_\star \geq \text{LGD}_\star^{\text{Floor}}$ where:

$$\text{LGD}_\star^{\text{Floor}} = \omega \cdot \text{LGD}^{\text{Floor}} + (1-\omega) \cdot \text{LGD}_C^{\text{Floor}}$$

$\text{LGD}^{\text{Floor}} = 25\%$ and $\text{LGD}_C^{\text{Floor}}$ depends on the collateral type: 0% for financial collateral, 10% for receivables and real estate and 15% for other physical collateral.

Remark 32 *Since the capital requirement is based on the unexpected loss, the Basel Committee imposes that the expected loss is deduced from regulatory capital.*

3.2.5 The securitization framework

Capital calculations for securitization require developing a more complex approach than the IRB approach, because the bank is not directly exposed to the loss of the credit portfolio, but to the conditional loss of the credit portfolio. This is particularly true if we consider a CDO tranche since we cannot measure the risk of equity, mezzanine and senior tranches in the same way. In what follows, we do not study the Basel II framework, which was very complex, but presented many weaknesses during the 2008 Global Financial Crisis. We prefer to focus on the Basel III framework (BCBS, 2016e), which is implemented since January 2018.

3.2.5.1 Overview of the approaches

The securitization framework consists of three approaches:

1. Securitization internal ratings-based approach (SEC-IRBA)
2. Securitization external ratings-based approach (SEC-ERBA)
3. Securitization standardized approach (SEC-SA)

[71] We recall that M is set to 2.5 years for all exposures, except for repo-style and retail exposures where the maturity is set to 6 and 12 months.

Contrary to credit risk, the hierarchy is reversed. The SEC-IRBA must be first used and is based on the capital charge $\mathcal{K}_{\text{IRB}}$ of the underlying exposures. If the bank cannot calculate $\mathcal{K}_{\text{IRB}}$ for a given securitization exposure, because it has not access to the collateral pool of the debt[72], it has to use the SEC-ERBA. If the tranche is unrated or if external ratings are not allowed, the bank must finally use the SEC-SA. When it is not possible to use one of the three approaches, the risk weight of the securitization exposure is set to 1 250%.

This framework has been developed for three types of exposures: STC securitization, non-STC securitization and resecuritization. STC stands for simple, transparent and comparable securitizations. In July 2015, the BCBS and the Board of IOSCO have published a set of 14 criteria for identifying STC exposures. These criteria are related to the collateral pool (asset risk), the transparency (structural risk) and the governance (fiduciary and servicer risk) of the SPV. Examples of criteria are the nature of the assets, the payment status, alignment of interests, transparency to investors, etc. Resecuritization implies that some underlying assets are themselves securitization exposures. For example, a CDO-squared is a resecuritization, because the asset pool is a basket of CDO tranches.

3.2.5.2 Internal ratings-based approach (SEC-IRBA)

In order to implement SEC-IRBA, the bank must conduct a strict due diligence of the pay-through securitization exposure in order to have a comprehensive information of the underlying exposures. For each asset that composes the collateral pool, it calculates the capital charge. Then, the bank determines $\mathcal{K}_{\text{IRB}}$ as the ratio between the sum of individual capital charges and the exposure amount of the collateral pool. If the bank has not all the information, it can use the following formula:

$$\mathcal{K}_{\text{IRB}} = \omega \cdot \mathcal{K}^{\star}_{\text{IRB}} + (1-\omega) \cdot \mathcal{K}_{\text{SA}}$$

where $\mathcal{K}^{\star}_{\text{IRB}}$ is the IRB capital requirement for the IRB pool[73], $\mathcal{K}_{\text{SA}}$ is the SA capital requirement for the underlying exposures and ω is the percentage of the IRB pool. However, this formula is only valid if $\omega \geq 95\%$. Otherwise, the bank must use the SEC-SA.

We consider a tranche, where A is the attachment point and D is the detachment point. If $\mathcal{K}_{\text{IRB}} \geq D$, the Basel Committee considers that the risk is very high and RW is set to 1 250%. Otherwise, we have:

$$\begin{aligned} \text{RW} &= 12.5 \cdot \left(\frac{\max\left(\mathcal{K}_{\text{IRB}}, A\right) - A}{D - A} \right) + \\ &\quad 12.5 \cdot \left(\frac{D - \max\left(\mathcal{K}_{\text{IRB}}, A\right)}{D - A} \right) \cdot \mathcal{K}_{\text{SSFA}}\left(\mathcal{K}_{\text{IRB}}\right) \end{aligned} \tag{3.36}$$

where $\mathcal{K}_{\text{SSFA}}\left(\mathcal{K}_{\text{IRB}}\right)$ is the capital charge for one unit of securitization exposure[74]. Therefore, we obtain two cases. If $A < \mathcal{K}_{\text{IRB}} < D$, we replace $\max\left(\mathcal{K}_{\text{IRB}}, A\right)$ by $\mathcal{K}_{\text{IRB}}$ in the previous formula. It follows that the capital charge between the attachment point A and $\mathcal{K}_{\text{IRB}}$ is risk-weighted by 1 250% and the remaining part between $\mathcal{K}_{\text{IRB}}$ and the detachment point D is risk-weighted by $12.5 \cdot \mathcal{K}_{\text{SSFA}}\left(\mathcal{K}_{\text{IRB}}\right)$. This is equivalent to consider that the sub-tranche $\mathcal{K}_{\text{IRB}} - A$ has already defaulted, while the credit risk is on the sub-tranche $D - \mathcal{K}_{\text{IRB}}$. In the second case $\mathcal{K}_{\text{IRB}} < A < D$, the first term of the formula vanishes, and we retrieve the RWA formula (3.27) on page 177.

[72] The structure of pay-through securitization is shown in Figure 3.12 on page 139.

[73] It corresponds to the part of the collateral pool, for which the bank has the information on the individual underlying exposures.

[74] It corresponds to the variable $\mathcal{K}^{\star}$ in the IRB formula on page 177.

The capital charge for one unit of securitization exposure is equal to[75]:

$$\boldsymbol{\mathcal{K}}_{\text{SSFA}}\left(\boldsymbol{\mathcal{K}}_{\text{IRB}}\right) = \frac{\exp\left(cu\right) - \exp\left(cl\right)}{c\left(u-l\right)}$$

where $c = -\left(p\boldsymbol{\mathcal{K}}_{\text{IRB}}\right)^{-1}$, $u = D - \boldsymbol{\mathcal{K}}_{\text{IRB}}$, $l = \left(A - \boldsymbol{\mathcal{K}}_{\text{IRB}}\right)^{+}$ and:

$$p = \max\left(0.3; m_{\text{STC}}\left(\alpha + \frac{\beta}{N} + \gamma \cdot \boldsymbol{\mathcal{K}}_{\text{IRB}} + \delta \cdot \text{LGD} + \epsilon \cdot M_{[A;D]}\right)\right)$$

The parameter p is called the supervisory parameter and is a function of the effective number[76] of loans N, the average LGD and the effective maturity[77] $M_{[A;D]}$ of the tranche. The coefficient m_{STC} is equal to 1 for non-STC securitizations and 0.5 for STC securitizations, while the other parameters α, β, γ, δ and ϵ are given in Table 3.31. We notice that the values depend on the underlying portfolio (wholesale or retail), the granularity ($N < 25$ or $N \geq 25$) and the seniority.

TABLE 3.31: Value of the parameters α, β, γ, δ and ϵ (SEC-IRBA)

Category	Senior	Granularity	α	β	γ	δ	ϵ
Wholesale	✓	$N \geq 25$	0.00	3.56	−1.85	0.55	0.07
	✓	$N < 25$	0.11	2.61	−2.91	0.68	0.07
		$N \geq 25$	0.16	2.87	−1.03	0.21	0.07
		$N < 25$	0.22	2.35	−2.46	0.48	0.07
Retail	✓		0.00	0.00	−7.48	0.71	0.24
			0.00	0.00	−5.78	0.55	0.27

Remark 33 *The derivation of these formulas is based on the model of Gordy and Jones (2003).*

Example 31 *We consider a non-STC CDO based on wholesale assets with three tranches: equity (0%−5%), mezzanine (5%−30%) and senior (30%−100%). The remaining maturity is equal to 10 years. The analysis of the underlying portfolio shows that the effective number of loans N is equal to 30 and the average LGD is equal to 30%. We also assume that $\boldsymbol{\mathcal{K}}^{\star}_{\text{IRB}} = 18\%$, $\boldsymbol{\mathcal{K}}_{\text{SA}} = 20\%$ and $\omega = 95\%$.*

We have $\boldsymbol{\mathcal{K}}_{\text{IRB}} = 0.95 \times 18\% + 0.05 \times 20\% = 18.1\%$. Since $\boldsymbol{\mathcal{K}}_{\text{IRB}} > D_{\text{equity}}$, we deduce that $\text{RW}_{\text{equity}} = 1\,250\%$. For the mezzanine tranche, we have $1 + 0.8 \times (M-1) = 8.2$ years, meaning that the 5-year cap is applied. Using Table 3.31 (fourth row), we deduce that $\alpha = 0.16$, $\beta = 2.87$, $\gamma = -1.03$, $\delta = 0.21$ and $\epsilon = 0.07$. It follows that:

$$\begin{aligned} p &= \max\left(0.30; 0.16 + \frac{2.87}{30} - 1.03 \times 18.1\% + 0.21 \times 30\% + 0.07 \times 5\right) \\ &= 48.22\% \end{aligned}$$

[75] SSFA means simplified supervisory formula approach.

[76] The effective number is equal to the inverse of the Herfindahl index H where $H = \sum_{i=1}^{n} w_i^2$ and w_i is the weight of the i^{th} asset. In our case, we have $w_i = \text{EAD}_i \,/\, \sum_{j=1}^{n} \text{EAD}_j$, implying that:

$$N = \frac{\left(\sum_{i=1}^{n} \text{EAD}_i\right)^2}{\sum_{i=1}^{n} \text{EAD}_i^2}$$

[77] Like for the IRB approach, $M_{[A;D]}$ is the effective maturity with a one-year floor and five-year cap. The effective maturity can be calculated as the weighted-average maturity of the cash-flows of the tranche or $1 + 0.8 \cdot (M-1)$ where M is the legal maturity of the tranche.

Since we have $c = -11.46$, $u = 11.90\%$ and $l = 0\%$, we obtain $\boldsymbol{\mathcal{K}}_{\text{SSFA}}(\boldsymbol{\mathcal{K}}_{\text{IRB}}) = 54.59\%$. Finally, Equation (3.36) gives $\text{RW}_{\text{mezzanine}} = 979.79\%$. If we perform the same analysis for the senior tranche[78], we obtain $\text{RW}_{\text{senior}} = 10.84\%$.

3.2.5.3 External ratings-based approach (SEC-ERBA)

Under the ERBA, we have:

$$\text{RWA} = \text{EAD} \cdot \text{RW}$$

where EAD is the securitization exposure amount and RW is the risk weight that depends on the external rating[79] and four other parameters: the STC criterion, the seniority of the tranche, the maturity and the thickness of the tranche. In the case of short-term ratings, the risk weights are given below:

Rating	A-1/P-1	A-2/P-2	A-3/P-3	Other
STC	10%	30%	60%	1 250%
non-STC	15%	50%	100%	1 250%

For long term ratings, the risk weight goes from 15% for AAA-grade to 1 250% (Table 2, BCBS 2016e, page 27). An example of risk weights for non-STC securitizations is given below:

Rating	Senior		Non-senior	
	1Y	5Y	1Y	5Y
AAA	15%	20%	15%	70%
AA	25%	40%	30%	120%
A	50%	65%	80%	180%
BBB	90%	105%	220%	310%
BB	160%	180%	620%	760%
B	310%	340%	1 050%	1 050%
CCC	460%	505%	1 250%	1 250%
Below CCC-	1 250%	1 250%	1 250%	1 250%

These risk weights are then adjusted for taking into account the effective maturity $M_{[A;D]}$ and the thickness $D - A$ of the tranche. The maturity adjustment corresponds to a linear interpolation between one and five years. The thickness adjustment must be done for non-senior tranches by multiplying the risk weight by the factor $1 - \min(D - A; 0.5)$.

Example 32 *We consider Example 31 and we assume that the mezzanine and senior tranches are rated BB and AAA.*

Using the table above, we deduce that the non-adjusted risk weights are equal to 1 250% for the equity tranche, 760% for the mezzanine tranche and 20% for the senior tranche. There is no maturity adjustment because $M_{[A;D]}$ is equal to five years. Finally, we obtain $\text{RW}_{\text{equity}} = 1\,250\% \times (1 - \min(5\%, 50\%)) = 1187.5\%$, $\text{RW}_{\text{mezzanine}} = 760\% \times (1 - \min(25\%, 50\%)) = 570\%$ and $\text{RW}_{\text{senior}} = 20\%$.

3.2.5.4 Standardized approach (SEC-SA)

The SA is very close to the IRBA since it uses Equation (3.36) by replacing $\boldsymbol{\mathcal{K}}_{\text{IRB}}$ by $\boldsymbol{\mathcal{K}}_{\text{A}}$ and the supervisory parameter p by the default values 0.5 and 1 for STC and non-STC securitizations. To calculate $\boldsymbol{\mathcal{K}}_{\text{A}}$, we first determine $\boldsymbol{\mathcal{K}}_{\text{SA}}$ which is the ratio between the

[78] In this case, the parameters are $\alpha = 0$, $\beta = 3.56$, $\gamma = -1.85$, $\delta = 0.55$ and $\epsilon = 0.07$ (second row in Table 3.31). We have $p = \max(30\%; 29.88\%) = 30\%$, $c = -18.42$, $u = 81.90\%$, $l = 11.90\%$, and $\boldsymbol{\mathcal{K}}_{\text{SSFA}}(\boldsymbol{\mathcal{K}}_{\text{IRB}}) = 0.87\%$.

[79] By definition, this approach is only available for tranches that are rated.

weighted average capital charge of the underlying portfolio computed with the SA approach and the exposure amount of the underlying portfolio. Then, we have:

$$\mathcal{K}_{\text{A}} = (1 - \varpi) \cdot \mathcal{K}_{\text{SA}} + \varpi \cdot 50\%$$

where ϖ is the percentage of underlying exposures that are 90 days or more past due.

Remark 34 *The SEC-SA is the only approach allowed for calculating the capital requirement of resecuritization exposures. In this case, ϖ is set to zero and the supervisory parameter p is equal to* 1.5.

If we consider Example 31 on page 187 and assume that $\varpi = 0$, we obtain $\text{RW}_{\text{equity}} = 1\,250\%$, $\text{RW}_{\text{mezzanine}} = 1\,143\%$ and $\text{RW}_{\text{senior}} = 210.08\%$.

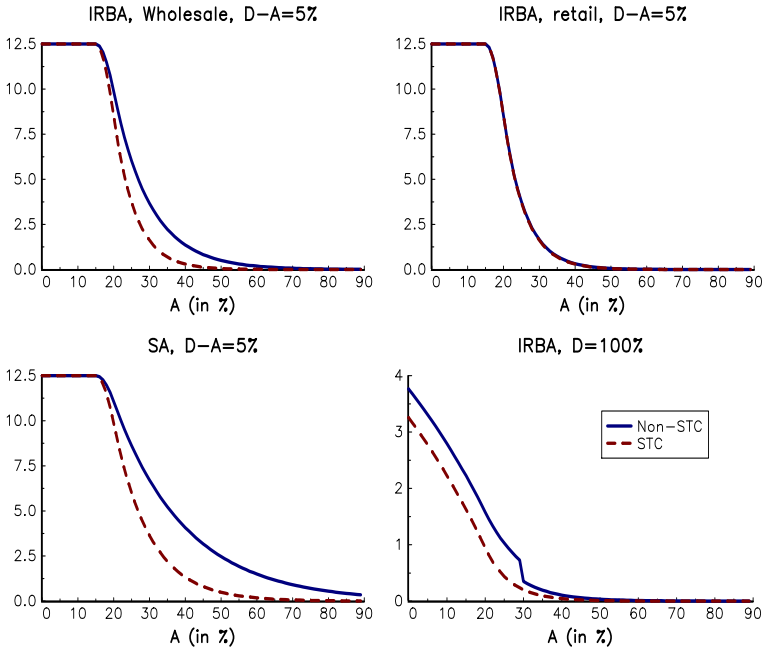


FIGURE 3.23: Risk weight of securitization exposures

Example 33 *We consider a CDO tranche, whose attachment and detachment points are A and D. We assume that $\mathcal{K}_{\text{IRB}} = \mathcal{K}_{\text{A}} = 20\%$, $N = 30$, $\text{LGD} = 50\%$ and $\varpi = 0$.*

In Figure 3.23, we have represented the evolution of the risk weight RW of the tranche $[A, D]$ for different values of A and D. For the first third panels, the thickness of the tranche is equal to 5%, while the detachment point is set to 100% for the fourth panel. In each panel, we consider two cases: non-STC and STC. If we compare the first and second panels, we notice the impact of the asset category (wholesale vs retail) on the risk weight. The third panel shows that the SA approach penalizes more non-STC securitization exposures. Since the detachment point is equal to 100%, the fourth panel corresponds to a senior tranche for high values of the attachment point A and a non-senior tranche when the attachment point A is low. In this example, we assume that the tranche becomes non-senior when $A < 30\%$. We observe a small cliff effect for non-STC securitization exposures.

3.3 Credit risk modeling

We now address the problem of parameter specification. This mainly concerns the exposure at default, the loss given default and the probability of default because the effective maturity is well defined. This section also analyzes default correlations and non granular portfolios when the bank develops its own credit model for calculating economic capital and satisfying Pillar 2 requirements.

3.3.1 Exposure at default

According to BCBS (2017c), the exposure at default "*for an on-balance sheet or off-balance sheet item is defined as the expected gross exposure of the facility upon default of the obligor*". Generally, the computation of EAD for on-balance sheet assets is not an issue. For example, EAD corresponds to the gross notional in the case of a loan or a credit. In fact, the big issue concerns off-balance sheet items, such as revolving lines of credit, credit cards or home equity lines of credit (HELOC). At the default time $\boldsymbol{\tau}$, we have (Taplin *et al.*, 2007):

$$\mathrm{EAD}\left(\boldsymbol{\tau} \mid t\right) = B\left(t\right) + \mathrm{CCF} \cdot \left(L\left(t\right) - B\left(t\right)\right) \tag{3.37}$$

where $B\left(t\right)$ is the outstanding balance (or current drawn) at time t, $L\left(t\right)$ is the current undrawn limit of the credit facility[80] and CCF is the credit conversion factor. This means that the exposure at default for off-balance sheet items has two components: the current drawn, which is a non-random component and the future drawn, which is a random component.

From Equation (3.37), we deduce that:

$$\mathrm{CCF} = \frac{\mathrm{EAD}\left(\boldsymbol{\tau} \mid t\right) - B\left(t\right)}{L\left(t\right) - B\left(t\right)} \tag{3.38}$$

At first sight, it looks easy to estimate the credit conversion factor. Let us consider the off-balance sheet item i that has defaulted. We have:

$$\mathrm{CCF}_i\left(\boldsymbol{\tau}_i - t\right) = \frac{B_i\left(\boldsymbol{\tau}_i\right) - B_i\left(t\right)}{L_i\left(t\right) - B_i\left(t\right)}$$

At time $\boldsymbol{\tau}_i$, we observe the default of Asset i and the corresponding exposure at default, which is equal to the outstanding balance $B_i\left(\boldsymbol{\tau}_i\right)$. Then, we have to choose a date $t < \boldsymbol{\tau}_i$ to observe $B_i\left(t\right)$ and $L_i\left(t\right)$ in order to calculate the CCF. We notice that it is sensitive to the time period $\boldsymbol{\tau}_i - t$, but banks generally use a one-year time period. Therefore, we can calculate the mean or the quantile α of a sample $\{\mathrm{CCF}_1, \ldots, \mathrm{CCF}_n\}$ for a given homogenous category of off-balance sheet items. Like the supervisory CCF values, the estimated CCF is a figure between 0% and 100%.

In practice, it is difficult to estimate CCF values for five reasons:

1. As explained by Qi (2009), there is a '*race to default*' between borrowers and lenders. Indeed, "*as borrowers approach default, their financial conditions deteriorate and they may use the current undrawn as a source of funding, whereas lenders may cut back credit lines to reduce potential losses*" (Qi, 2009, page 4).

2. $L_i\left(t\right)$ depends on the current time t, meaning that it could evolve over time.

[80] The current undrawn $L\left(t\right) - B\left(t\right)$ is the amount that the debtor is able to draw upon in addition to the current drawn $B\left(t\right)$.

3. The computation of the CCF is sensitive to the denominator $L_i(t) - B_i(t)$, which can be small. When $L_i(t) \approx B_i(t)$, the CCF ratio is unstable.

4. We have made the assumption that $\text{CCF}_i(\boldsymbol{\tau}_i - t) \in [0,1]$, implying that $B_i(\boldsymbol{\tau}_i) \geq B_i(t)$ and $B_i(\boldsymbol{\tau}_i) \leq L_i(t)$. This is not always true. We can imagine that the outstanding balance decreases between the current time and the default time ($\text{CCF}_i(\boldsymbol{\tau}_i - t) < 0$) or the outstanding balance at the default time is greater than the limit $L_i(t)$. Jacobs, Jr. (2010) reports extreme variation larger than $\pm 3\,000\%$ when computing raw CCF values!

5. The credit conversion factor is generally an increasing function of the default probability of the borrower.

Because of the previous issues, the observed CCF is floored at 0% and capped at 100%. Tong *et al.* (2016) report the distribution of the credit conversion factor of credit cards from a UK bank[81], and notice that the observations are mainly concentred on the two extreme points 0% and 100% after truncation. Another measure for modeling the exposure at default is to consider the facility utilization change factor (Yang and Tkachenko, 2012):

$$\text{UCF} = \frac{B_i(\boldsymbol{\tau}_i) - B_i(t)}{L_i(t)}$$

It corresponds to the credit conversion factor, where the current undrawn amount $L_i(t) - B_i(t)$ is replaced by the current authorized limit $L_i(t)$. It has the advantage to be more stable, in particular around the singularity $L_i(t) = B_i(t)$.

The econometrics of CCF is fairly basic. As said previously, it consists in estimating the mean or the quantile α of a sample $\{\text{CCF}_1, \ldots, \text{CCF}_n\}$. For that, we can use the cohort method or the time horizon approach (Witzany, 2011). In the cohort method, we divide the study period into fixed intervals (6 or 12 months). For each asset, we identify if it has defaulted during the interval, and then we set t to the starting date of the interval. In the time horizon approach, t is equal to the default time $\boldsymbol{\tau}_i$ minus a fixed horizon (e.g. one, three or 12 months). Sometimes, it can be useful to include some explanatory variables. In this case, the standard model is the Tobit linear regression, which is presented on page 708, because data are censored and the predicted value of CCF must lie in the interval $[0,1]$.

3.3.2 Loss given default

3.3.2.1 Definition

The recovery rate $\boldsymbol{\mathcal{R}}$ is the percentage of the notional on the defaulted debt that can be recovered. In the Basel framework, the recovery rate is not explicitly used, and the concept of loss given default is preferred for measuring the credit portfolio loss. The two metrics are expressed as a percentage of the face value, and we have:

$$\text{LGD} \geq 1 - \boldsymbol{\mathcal{R}}$$

Let us consider a bank that is lending \$100 mn to a corporate firm. We assume that the firm defaults at one time and the bank recovers \$60 mn. We deduce that the recovery rate is equal to:

$$\boldsymbol{\mathcal{R}} = \frac{60}{100} = 60\%$$

[81] See Figure 1 on page 912 in Tong *et al.* (2016).

In order to recover \$60 mn, the bank has incurred some operational and litigation costs, whose amount is \$5 mn. In this case, the bank has lost \$40 mn plus \$5 mn, implying that the loss given default is equal to:

$$\text{LGD} = \frac{40+5}{100} = 45\%$$

In fact, this example shows that $\boldsymbol{\mathcal{R}}$ and LGD are related in the following way:

$$\text{LGD} = 1 - \boldsymbol{\mathcal{R}} + c$$

where c is the litigation cost. We now understand why the loss given default is the right measure when computing the portfolio loss.

Schuermann (2004) identifies three approaches for calculating the loss given default:

1. Market LGD
2. Implied LGD
3. Workout LGD

The market LGD is deduced from the bond price just after the default[82]. It is easy to calculate and available for large corporates and banks. The implied LGD is calculated from a theoretical pricing model of bonds or CDS. The underlying idea is to estimate the implied loss given default, which is priced by the market. As for the first method, this metric is easy to calculate, but it depends on the model assumptions. The last approach is the workout or ultimate LGD. Indeed, the loss given default has three components: the direct loss of principal, the loss of carrying non-performing loans and the workout operational and legal costs. The workout LGD is the right measure when considering the IRB approach. Nevertheless, Schuermann (2004) notices that between two and three years are needed on average to obtain the recovery.

In what follows, we present two approaches for modeling LGD. The first approach considers that LGD is a random variable, whose probability distribution has to be estimated:

$$\text{LGD} \sim \mathbf{F}(x) \tag{3.39}$$

However, we recall that the loss given default in the Basel IRB formulas does not correspond to the random variable, but to its expectation $\mathbb{E}[\text{LGD}]$. Therefore, the second approach consists in estimating the conditional expectation:

$$\begin{aligned} \mathbb{E}[\text{LGD}] &= \mathbb{E}[\text{LGD} \mid X_1 = x_1, \ldots, X_m = x_m] \\ &= g(x_1, \ldots, x_m) \end{aligned} \tag{3.40}$$

where $(X_1, \ldots, X_m)$ are the risk factors that determine the loss given default.

Remark 35 *We notice that* $\boldsymbol{\mathcal{R}} \in [0,1]$*, but* $\text{LGD} \geq 0$*. Indeed, we can imagine that the litigation cost can be high compared to the recovery part of the debt. In this case, we can have* $c > \boldsymbol{\mathcal{R}}$*, implying that* $\text{LGD} > 100\%$*. For instance, if* $\boldsymbol{\mathcal{R}} = 20\%$ *and* $c = 30\%$*, we obtain* $\text{LGD} = 110\%$*. This situation is not fanciful, because* $\boldsymbol{\mathcal{R}}$ *and* c *are not known at the default time. The bank will then begin to engage costs without knowing the recovery amount. For example, one typical situation is* $\boldsymbol{\mathcal{R}} = 0\%$ *and* $c > 0$*, when the bank discovers that there is no possible recovery, but has already incurs some litigation costs. Even if* LGD *can be larger than* 100%*, we assume that* $\text{LGD} \in [0,1]$ *because these situations are unusual.*

[82] This measure is also called '*trading price recovery*'.

3.3.2.2 Stochastic modeling

Using a parametric distribution In this case, we generally use the beta distribution $\mathcal{B}(\alpha, \beta)$, which is described on page 1053. Its density function is given by:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathfrak{B}(\alpha, \beta)}$$

where $\mathfrak{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt$. The mean and the variance are:

$$\mu(X) = \mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$$

and:

$$\sigma^2(X) = \text{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

When α and β are greater than 1, the distribution has one mode $x_{\text{mode}} = (\alpha-1)/(\alpha+\beta-2)$. This probability distribution is very flexible and allows to obtain various shapes that are given in Figure 3.24:

- if $\alpha = 1$ and $\beta = 1$, we obtain the uniform distribution; if $\alpha \to \infty$ and $\beta \to \infty$, we obtain the Dirac distribution at the point $x = 0.5$; if one parameter goes to zero, we obtain a Bernoulli distribution;
- if $\alpha = \beta$, the distribution is symmetric around $x = 0.5$; we have a bell curve when the two parameters α and β are higher than 1, and a **U**-shape curve when the two parameters α and β are lower than 1;
- if $\alpha > \beta$, the skewness is negative and the distribution is left-skewed, if $\alpha < \beta$, the skewness is positive and the distribution is right-skewed.

Given the estimated mean $\hat{\mu}_{\text{LGD}}$ and standard deviation $\hat{\sigma}_{\text{LGD}}$ of a sample of losses given default, we can calibrate the parameters α and β using the method of moments[83]:

$$\hat{\alpha} = \frac{\hat{\mu}_{\text{LGD}}^2(1-\hat{\mu}_{\text{LGD}})}{\hat{\sigma}_{\text{LGD}}^2} - \hat{\mu}_{\text{LGD}} \tag{3.41}$$

and:

$$\hat{\beta} = \frac{\hat{\mu}_{\text{LGD}}(1-\hat{\mu}_{\text{LGD}})^2}{\hat{\sigma}_{\text{LGD}}^2} - (1-\hat{\mu}_{\text{LGD}}) \tag{3.42}$$

The other approach is to use the method of maximum likelihood, which is described in Section 10.1.2 on page 614.

Example 34 *We consider the following sample of losses given default:* 68%, 90%, 22%, 45%, 17%, 25%, 89%, 65%, 75%, 56%, 87%, 92% *and* 46%.

We obtain $\hat{\mu}_{\text{LGD}} = 59.77\%$ and $\hat{\sigma}_{\text{LGD}} = 27.02\%$. Using the method of moments, the estimated parameters are $\hat{\alpha}_{\text{MM}} = 1.37$ and $\hat{\beta}_{\text{MM}} = 0.92$, whereas we have $\hat{\alpha}_{\text{ML}} = 1.84$ and $\hat{\beta}_{\text{ML}} = 1.25$ for the method of maximum likelihood. We notice that the two calibrated probability distributions have different shapes (see Figure 3.25).

[83] See Section 10.1.3.1 on page 628.

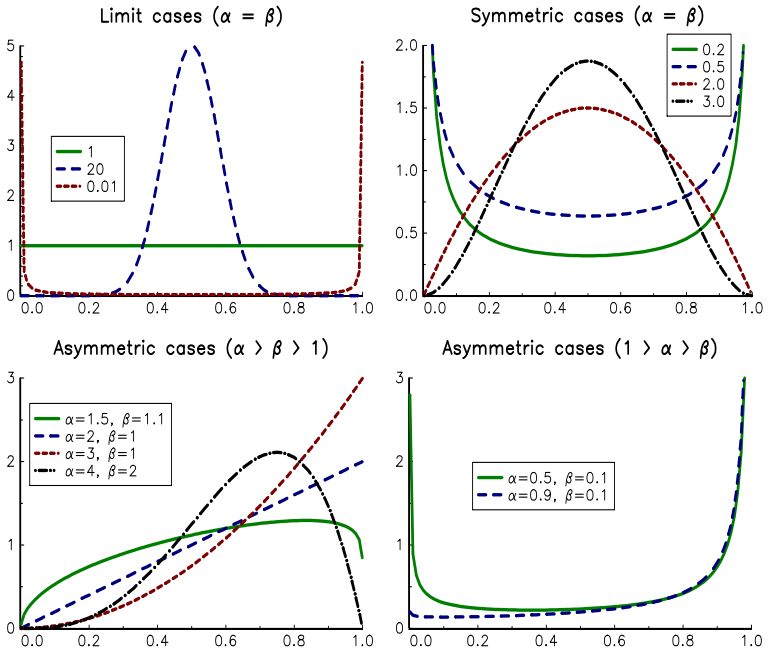


FIGURE 3.24: Probability density function of the beta distribution $\mathcal{B}(\alpha, \beta)$

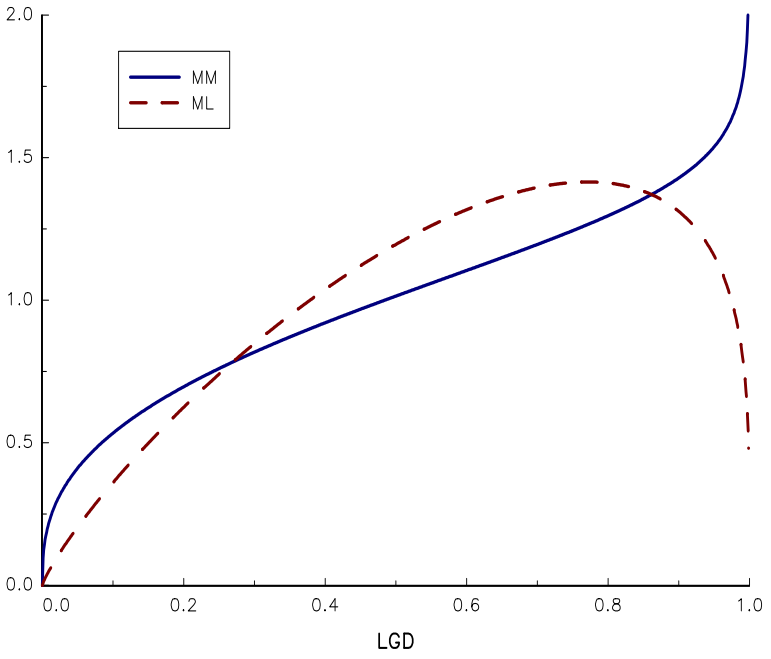


FIGURE 3.25: Calibration of the beta distribution

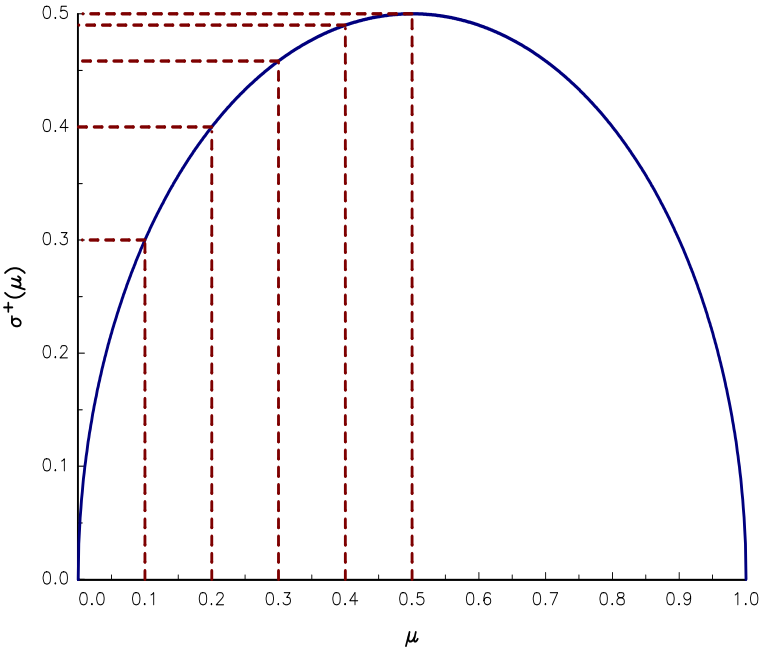


FIGURE 3.26: Maximum standard deviation $\sigma^+(\mu)$

Remark 36 *We can calibrate the beta distribution as long as we respect some constraints on* $\hat{\mu}_{\text{LGD}}$ *and* $\hat{\sigma}_{\text{LGD}}$*. Using Equations (3.41) and (3.42), we deduce that:*

$$\hat{\sigma}_{\text{LGD}} < \sqrt{\hat{\mu}_{\text{LGD}}\left(1-\hat{\mu}_{\text{LGD}}\right)}$$

because $\hat{\alpha}$ *and* $\hat{\beta}$ *must be positive. This condition is not well restrictive. Indeed, if we consider a general random variable* X *on* $[0,1]$*, we have* $\mathbb{E}\left[X^2\right] \leq \mathbb{E}\left[X\right]$*, implying that:*

$$\sigma(X) \leq \sigma^+(\mu) = \sqrt{\mu(1-\mu)}$$

where $\mu = \mathbb{E}[X]$*. Therefore, only the limit case cannot be reached by the beta distribution*[84]*. However, we notice that the standard deviation cannot be arbitrary fixed to a high level. For example, Figure 3.26 shows that there is no random variable on* $[0,1]$ *such that* $\mu = 10\%$ *and* $\sigma > 30\%$*,* $\mu = 20\%$ *and* $\sigma > 40\%$*,* $\mu = 50\%$ *and* $\sigma > 50\%$*, etc.*

In Figure 3.27, we have reported the calibrated beta distribution using the method of moments for several values of μ_{LGD} and $\sigma_{\text{LGD}} = 30\%$. We obtain **U**-shaped probability distributions. In order to obtain a concave (or bell-shaped) distribution, the standard deviation σ_{LGD} must be lower (see Figure 3.28).

Remark 37 *The previous figures may leave us believing that the standard deviation must be very low in order to obtain a concave beta probability density function. In fact, this is not a restriction due to the beta distribution, since it is due to the support* $[0,1]$ *of the random variable. Indeed, we can show that the standard deviation is bounded*[85] *by* $\sqrt{1/12} \simeq 28.86\%$ *when the probability distribution has one mode on* $[0,1]$*.*

[84] The limit case corresponds to the Bernoulli distribution $\mathcal{B}(p)$ where $p = \mu$.
[85] The bound is the standard deviation of the uniform distribution $\mathcal{U}_{[0,1]}$.

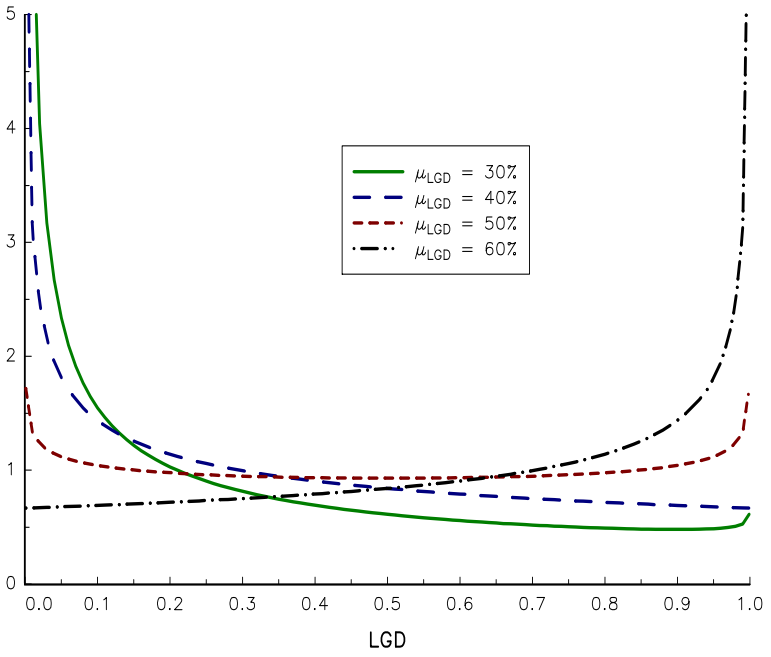


FIGURE 3.27: Calibration of the beta distribution when $\sigma_{\text{LGD}} = 30\%$

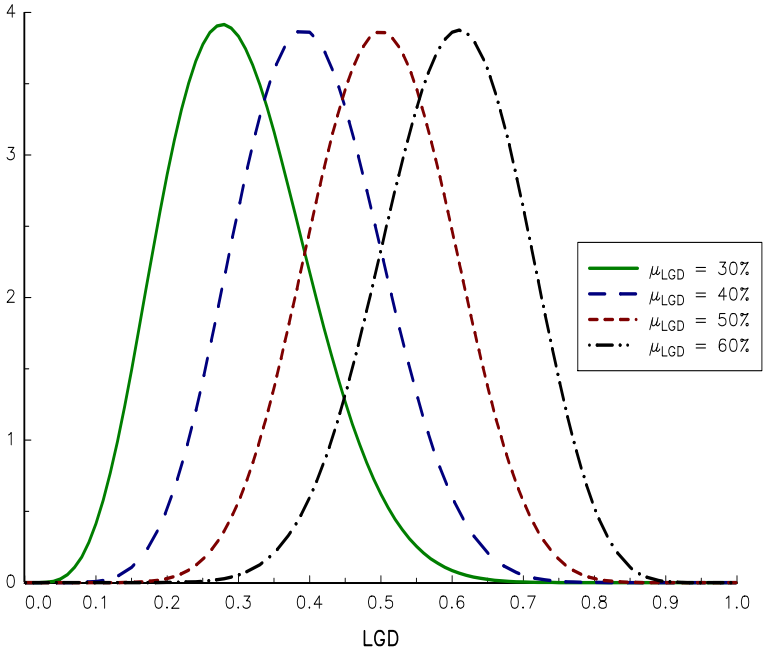


FIGURE 3.28: Calibration of the beta distribution when $\sigma_{\text{LGD}} = 10\%$

As noted by Altman and Kalotay (2014), the beta distribution is not always appropriate for modeling loss given default even if it is widespread used by the industry. Indeed, we observe that losses given default tend to be bimodal, meaning that the recovery rate is quite high or quite low (Loterman *et al.*, 2012). This is why Altman and Kalotay (2014) propose to model the loss given default as a Gaussian mixture model. They first apply the transformation $y_i = \Phi^{-1}(\text{LGD}_i)$ to the sample, then calibrate[86] the 4-component mixture model on the transformed data $(y_1, \ldots, y_n)$ and finally perform the inverse transform for estimating the parametric distribution. They show that the estimated distribution fits relatively well the non-parametric distribution estimated with the kernel method.

Using a non-parametric distribution The beta distribution is either bell-shaped or **U**-shaped. In this last case, the limit is the Bernoulli distribution:

LGD	0%	100%
Probability	$(1 - \mu_{\text{LGD}})$	μ_{LGD}

This model is not necessarily absurd, since it means that the recovery can be very high or very low. Figure 2 in Bellotti and Crook (2012) represents the histogram of recovery rates of 55 000 defaulted credit card accounts from 1999 to 2005 in the UK. The two extreme cases ($\boldsymbol{\mathcal{R}} = 0\%$ and $\boldsymbol{\mathcal{R}} = 100\%$) are the most frequent cases. Therefore, it is interesting to consider the empirical distribution instead of an estimated distribution. In this case, we generally consider risk classes, e.g. $0\% - 5\%, 5\% - 10\%, 10\% - 20\%, \ldots, 80\% - 90\%, 90\% - 100\%$.

Example 35 *We consider the following empirical distribution of* LGD*:*

LGD *(in %)*	0	10	20	25	30	40	50	60	70	75	80	90	100
$\hat{p}$ *(in %)*	1	2	10	25	10	2	0	2	10	25	10	2	1

This example illustrates the shortcoming of the beta modeling when we have a bimodal LGD distribution. In Figure 3.29, we have reported the empirical distribution, and the corresponding (rescaled) calibrated beta distribution. We notice that it is very far from the empirical distribution.

Remark 38 *Instead of using the empirical distribution by risk classes, we can also consider the kernel approach, which is described on page 637.*

Example 36 *We consider a credit portfolio of* 10 *loans, whose loss is equal to:*

$$L = \sum_{i=1}^{10} \text{EaD}_i \cdot \text{LGD}_i \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq T_i\}$$

where the maturity T_i *is equal to 5 years, the exposure at default* EaD_i *is equal to* \$1 000 *and the default time* $\boldsymbol{\tau}_i$ *is exponential with the following intensity parameter* λ_i*:*

i	1	2	3	4	5	6	7	8	9	10
λ_i *(in bps)*	10	10	25	25	50	100	250	500	500	1 000

The loss given default LGD_i *is given by the empirical distribution, which is described in Example 35.*

[86] The estimation of Gaussian mixture models is presented on page 624.

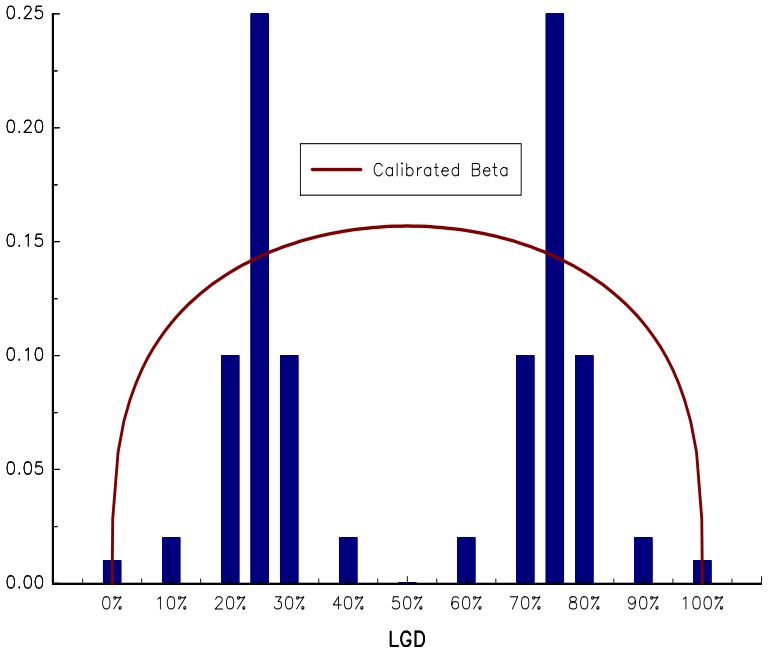


FIGURE 3.29: Calibration of a bimodal LGD distribution

In Figure 3.30, we have calculated the distribution of the portfolio loss with the Monte Carlo method. We compare the loss distribution when we consider the empirical distribution and the calibrated beta distribution for the loss given default. We also report the loss distribution when we replace the random variable LGD_i by its expected value $\mathbb{E}[\text{LGD}_i] = 50\%$. We observe that the shape of L highly depends on the LGD model. For example, we observe a more pronounced fat tail with the calibrated beta distribution. This implies that the LGD model has a big impact for calculating the value-at-risk. For instance, we have reported the loss distribution using the beta model for different values of $(\mu_{\text{LGD}}, \sigma_{\text{LGD}})$ in Figure 3.31. We conclude that the modeling of LGD must not be overlooked. In many cases, the model errors have more impact when they concern the loss given default than the probability of default.

Remark 39 *The expression of the portfolio loss is:*

$$L = \sum_{i=1}^{n} \text{EAD}_i \cdot \text{LGD}_i \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq T_i\}$$

If the portfolio is fined grained, we have:

$$\mathbb{E}[L \mid X] = \sum_{i=1}^{n} \text{EAD}_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(X)$$

We deduce that the distribution of the portfolio loss is equal to:

$$\Pr\{L \leq \ell\} = \int \cdots \int \mathbb{1}\left\{\sum_{i=1}^{n} \text{EAD}_i \cdot \mathbb{E}[\text{LGD}_i] \cdot p_i(x) \leq \ell\right\} \mathrm{d}\mathbf{H}(x)$$

This loss distribution does not depend on the random variables LGD_i*, but on their expected values* $\mathbb{E}[\text{LGD}_i]$*. This implies that it is not necessary to model the loss given default, but*

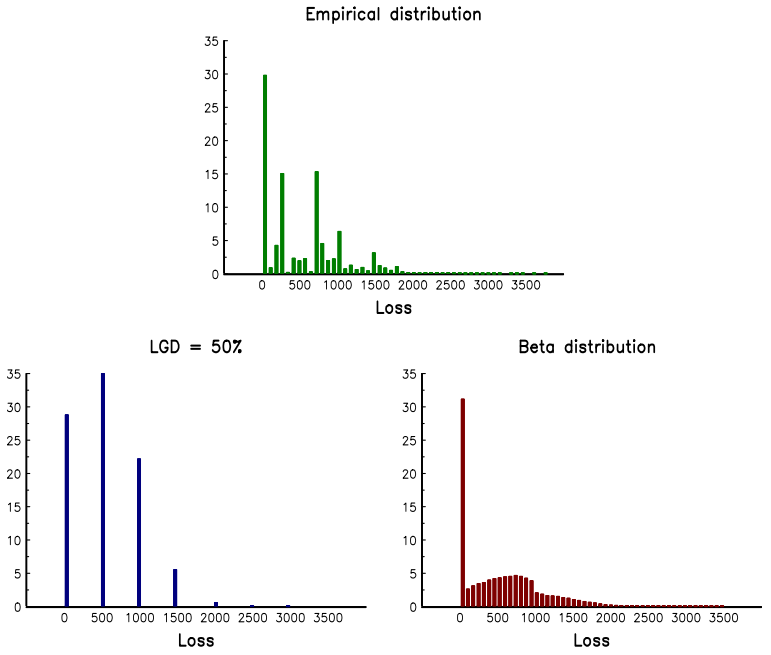


FIGURE 3.30: Loss frequency in % of the three LGD models

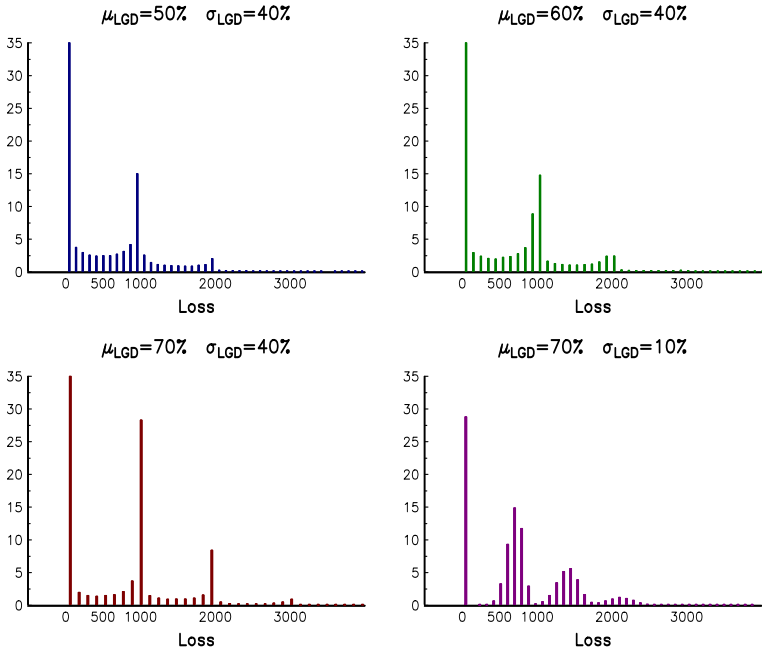


FIGURE 3.31: Loss frequency in % for different values of μ_{LGD} and σ_{LGD}

only the mean. Therefore, we can replace the previous expression of the portfolio loss by:

$$L = \sum_{i=1}^{n} \text{EAD}_i \cdot \mathbb{E}\left[\text{LGD}_i\right] \cdot \mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\}$$

3.3.2.3 Economic modeling

There are many factors that influence the recovery process. In the case of corporate debt, we distinguish between specific and economic factors. For instance, specific factors are the relative seniority of the debt or the guarantees. Senior debt must be repaid before subordinated or junior debt is repaid. If the debt is collateralized, this affects the loss given default. Economic factors are essentially the business cycle and the industry. In the third version of Moody's LossCalc, Dwyer and Korablev (2009) consider seven factors that are grouped in three major categories:

1. factors external to the issuer: geography, industry, credit cyle stage;
2. factors specific to the issuer: distance-to-default, probability of default (or leverage for private firms);
3. factors specific to the debt issuance: debt type, relative standing in capital structure, collateral.

Curiously, Dwyer and Korablev (2009) explain that "*some regions have been characterized as creditor-friendly, while others are considered more creditor-unfriendly*". For instance, recovery rates are lower in the UK and Europe than in the rest of the world. However, the most important factors are the seniority followed by the industry, as it is illustrated by the Moody's statistics on ultimate recoveries. From 1987 to 2017, the average corporate debt recovery rate is equal to 80.4% for loans, 62.3% for senior secured bonds, 47.9% for senior unsecured bonds and 28.0% for subordinated bonds (Moody's, 2018). It is interesting to notice that the recovery rate and the probability of default are negatively correlated. Indeed, Dwyer and Korablev (2009) take the example of two corporate firms A and B, and they assume that $\text{PD}_B \gg \text{PD}_A$. In this case, we may think that the assets of A relative to its liabilities is larger than the ratio of B. Therefore, we must observe a positive relationship between the loss given default and the probability of default.

Remark 40 *The factors depend of the asset class. For instance, we will consider more microeconomic variables when modeling the loss given default for mortgage loans (Tong et al., 2013).*

Once the factors are identified, we must estimate the LGD model:

$$\text{LGD} = f\left(X_1, \ldots, X_m\right)$$

where $X_1, \ldots, X_m$ are the m factors, and f is a non-linear function. Generally, we consider a transformation of LGD in order to obtain a more tractable variable. We can apply a logit transform $Y = \ln\left(\text{LGD}\right) - \ln\left(1 - \text{LGD}\right)$, a probit transform $Y = \Phi^{-1}\left(\text{LGD}\right)$ or a beta transformation (Bellotti and Crook, 2012). In this case, we can use the different statistical tools given in Chapters 10 and 15 to model the random variable Y. The most popular models are the logistic regression, regression trees and neural networks (Bastos, 2010). However, according to EBA (2017), multivariate regression remains the most widely used methods, despite the strong development of machine learning techniques, that are presented on page 943.

Remark 41 *We do not develop here the econometric approach, because it is extensively presented in Chapter 15 dedicated to the credit scoring. Indeed, statistical models of* LGD *use the same methods than statistical models of* PD*. We also refer to Chapter 14 dedicated to stress testing methods when we would like to calculate stressed LGD parameters.*

3.3.3 Probability of default

3.3.3.1 Survival function

The survival function is the main tool to characterize the probability of default. It is also known as reduced-form modeling.

Definition and main properties Let $\boldsymbol{\tau}$ be a default (or survival) time. The survival function[87] is defined as follows:

$$\begin{aligned}\mathbf{S}(t) &= \Pr\{\boldsymbol{\tau} > t\} \\ &= 1 - \mathbf{F}(t)\end{aligned}$$

where $\mathbf{F}$ is the cumulative distribution function. We deduce that the probability density function is related to the survival function in the following manner:

$$f(t) = -\frac{\partial \mathbf{S}(t)}{\partial t} \tag{3.43}$$

In survival analysis, the key concept is the hazard function $\lambda(t)$, which is the instantaneous default rate given that the default has not occurred before t:

$$\lambda(t) = \lim_{\mathrm{d}t \to 0^+} \frac{\Pr\{t \leq \boldsymbol{\tau} \leq t + \mathrm{d}t \mid \boldsymbol{\tau} \geq t\}}{\mathrm{d}t}$$

We deduce that:

$$\begin{aligned}\lambda(t) &= \lim_{\mathrm{d}t \to 0^+} \frac{\Pr\{t \leq \boldsymbol{\tau} \leq t + \mathrm{d}t\}}{\mathrm{d}t} \cdot \frac{1}{\Pr\{\boldsymbol{\tau} \geq t\}} \\ &= \frac{f(t)}{\mathbf{S}(t)}\end{aligned}$$

Using Equation (3.43), another expression of the hazard function is:

$$\begin{aligned}\lambda(t) &= -\frac{\partial_t \mathbf{S}(t)}{\mathbf{S}(t)} \\ &= -\frac{\partial \ln \mathbf{S}(t)}{\partial t}\end{aligned}$$

The survival function can then be rewritten with respect to the hazard function and we have:

$$\mathbf{S}(t) = e^{-\int_0^t \lambda(s)\,\mathrm{d}s} \tag{3.44}$$

In Table 3.32, we have reported the most common hazard and survival functions. They can be extended by adding explanatory variables in order to obtain proportional hazard models (Cox, 1972). In this case, the expression of the hazard function is $\lambda(t) = \lambda_0(t) \exp\left(\beta^\top x\right)$ where $\lambda_0(t)$ is the baseline hazard rate and x is the vector of explanatory variables, which are not dependent on time.

[87] Previously, we have noted the survival function as $\mathbf{S}_{t_0}(t)$. Here, we assume that the current time t_0 is 0.

TABLE 3.32: Common survival functions

Model	$\mathbf{S}(t)$	$\lambda(t)$
Exponential	$\exp(-\lambda t)$	λ
Weibull	$\exp(-\lambda t^{\gamma})$	$\lambda\gamma t^{\gamma-1}$
Log-normal	$1-\Phi(\gamma\ln(\lambda t))$	$\gamma t^{-1}\phi(\gamma\ln(\lambda t))/(1-\Phi(\gamma\ln(\lambda t)))$
Log-logistic	$1/\left(1+\lambda t^{\frac{1}{\gamma}}\right)$	$\lambda\gamma^{-1}t^{\frac{1}{\gamma}}/\left(t+\lambda t^{1+\frac{1}{\gamma}}\right)$
Gompertz	$\exp(\lambda(1-e^{\gamma t}))$	$\lambda\gamma\exp(\gamma t)$

The exponential model holds a special place in default time models. It can be justified by the following problem in physics:

> "*Assume that a system consists of n identical components which are connected in series. This means that the system fails as soon as one of the components fails. One can assume that the components function independently. Assume further that the random time interval until the failure of the system is one n^{th} of the time interval of component failure*" (Galambos, 1982).

We have $\Pr\{\min(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n)\leq t\}=\Pr\{\boldsymbol{\tau}_i\leq n\cdot t\}$. The problem is then equivalent to solve the functional equation $\mathbf{S}(t)=\mathbf{S}^n(t/n)$ with $\mathbf{S}(t)=\Pr\{\boldsymbol{\tau}_1>t\}$. We can show that the unique solution for $n\geq 1$ is the exponential distribution. Following Galambos and Kotz (1978), its other main properties are:

1. the mean residual life $\mathbb{E}[\boldsymbol{\tau}\mid\boldsymbol{\tau}\geq t]$ is constant;
2. it satisfies the famous lack of memory property:

$$\Pr\{\boldsymbol{\tau}\geq t+u\mid\boldsymbol{\tau}\geq t\}=\Pr\{\boldsymbol{\tau}\geq u\}$$

   or equivalently $\mathbf{S}(t+u)=\mathbf{S}(t)\,\mathbf{S}(u)$;
3. the probability distribution of $n\cdot\boldsymbol{\tau}_{1:n}$ is the same as probability distribution of $\boldsymbol{\tau}_i$.

Piecewise exponential model In credit risk models, the standard probability distribution to define default times is a generalization of the exponential model by considering piecewise constant hazard rates:

$$\begin{aligned}\lambda(t) &= \sum_{m=1}^{M}\lambda_m\cdot\mathbb{1}\left\{t_{m-1}^{\star}<t\leq t_m^{\star}\right\}\\ &= \lambda_m \qquad \text{if } t\in\left]t_{m-1}^{\star},t_m^{\star}\right]\end{aligned}$$

where $t_m^{\star}$ are the knots of the function[88]. For $t\in\left]t_{m-1}^{\star},t_m^{\star}\right]$, the expression of the survival function becomes:

$$\begin{aligned}\mathbf{S}(t) &= \exp\left(-\sum_{k=1}^{m-1}\lambda_k\left(t_k^{\star}-t_{k-1}^{\star}\right)-\lambda_m\left(t-t_{m-1}^{\star}\right)\right)\\ &= \mathbf{S}\left(t_{m-1}^{\star}\right)e^{-\lambda_m\left(t-t_{m-1}^{\star}\right)}\end{aligned}$$

[88] We have $t_0^{\star}=0$ and $t_{M+1}^{\star}=\infty$.

It follows that the density function is equal to[89]:

$$f(t) = \lambda_m \exp\left(-\sum_{k=1}^{m-1} \lambda_k \left(t_k^\star - t_{k-1}^\star\right) - \lambda_m \left(t - t_{m-1}^\star\right)\right)$$

In Figure 3.32, we have reported the hazard, survival and density functions for three set of parameters $\{(t_m^\star, \lambda_m), m = 1, \ldots, M\}$:

$$\begin{aligned} &\{(1, 1\%), (2, 1.5\%), (3, 2\%), (4, 2.5\%), (\infty, 3\%)\} && \text{for } \lambda_1(t) \\ &\{(1, 10\%), (2, 7\%), (5, 5\%), (7, 4.5\%), (\infty, 6\%)\} && \text{for } \lambda_2(t) \end{aligned}$$

and $\lambda_3(t) = 4\%$. We note the special shape of the density function, which is not smooth at the knots.

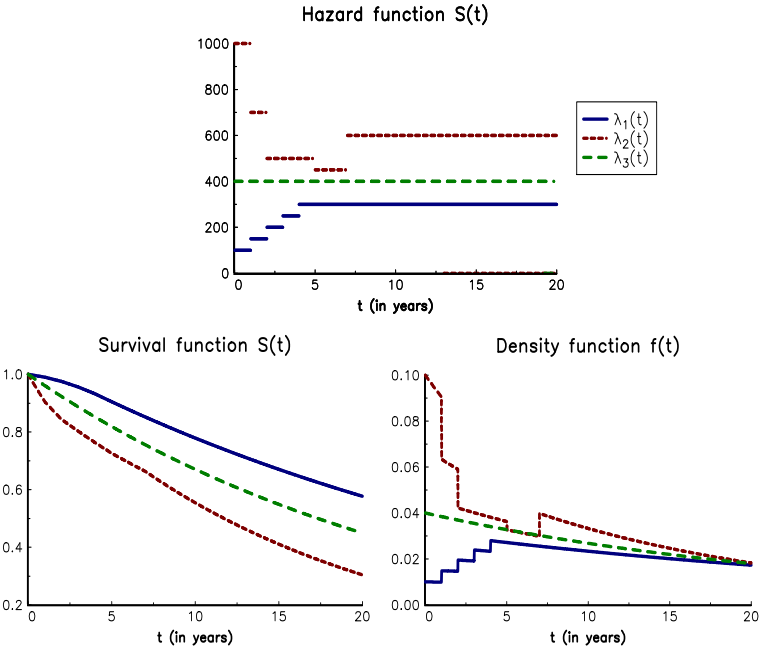


FIGURE 3.32: Example of the piecewise exponential model

Estimation To estimate the parameters of the survival function, we can use the cohort approach. Under this method, we estimate the empirical survival function by counting the number of entities for a given population that do not default over the period Δt:

$$\hat{\mathbf{S}}(\Delta t) = 1 - \frac{\sum_{i=1}^n \mathbb{1}\{t < \boldsymbol{\tau}_i \leq t + \Delta t\}}{n}$$

where n is the number of entities that compose the population. We can then fit the survival function by using for instance the least squares method.

[89] We verify that:

$$\frac{f(t)}{\mathbf{S}(t)} = \lambda_m \qquad \text{if } t \in \left]t_{m-1}^\star, t_m^\star\right]$$

Example 37 *We consider a population of* 1 000 *corporate firms. The number of defaults* $n_D(\Delta t)$ *over the period* Δt *is given in the table below:*

Δt *(in months)*	3	6	9	12	15	18	21	22
$n_D(\Delta t)$	2	5	9	12	16	20	25	29

We obtain $\hat{\mathbf{S}}(0.25) = 0.998$, $\hat{\mathbf{S}}(0.50) = 0.995$, $\hat{\mathbf{S}}(0.75) = 0.991$, $\hat{\mathbf{S}}(1.00) = 0.988$, $\hat{\mathbf{S}}(1.25) = 0.984$, $\hat{\mathbf{S}}(1.50) = 0.980$, $\hat{\mathbf{S}}(1.75) = 0.975$ and $\hat{\mathbf{S}}(2.00) = 0.971$. For the exponential model, the least squares estimator $\hat{\lambda}$ is equal to 1.375%. In the case of the Gompertz survival function, we obtain $\hat{\lambda} = 2.718\%$ and $\hat{\gamma} = 0.370$. If we consider the piecewise exponential model, whose knots correspond to the different periods Δt, we have $\hat{\lambda}_1 = 0.796\%$, $\hat{\lambda}_2 = 1.206\%$, $\hat{\lambda}_3 = 1.611\%$, $\hat{\lambda}_4 = 1.216\%$, $\hat{\lambda}_5 = 1.617\%$, $\hat{\lambda}_6 = 1.640\%$, $\hat{\lambda}_7 = 2.044\%$ and $\hat{\lambda}_8 = 1.642\%$. To compare these three calibrations, we report the corresponding hazard functions in Figure 3.33. We deduce that the one-year default probability[90] is respectively equal to 1.366%, 1.211% and 1.200%.

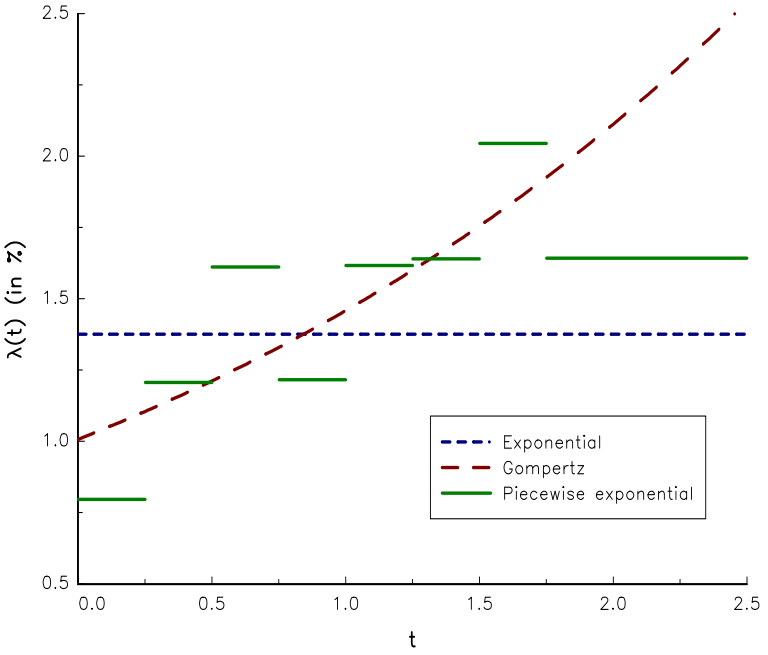


FIGURE 3.33: Estimated hazard function

In the piecewise exponential model, we can specify an arbitrary number of knots. In the previous example, we use the same number of knots than the number of observations to calibrate. In such case, we can calibrate the parameters using the following iterative process:

1. We first estimate the parameter λ_1 for the earliest maturity Δt_1.
2. Assuming that $\left(\hat{\lambda}_1, \ldots, \hat{\lambda}_{i-1}\right)$ have been estimated, we calculate $\hat{\lambda}_i$ for the next maturity Δt_i.
3. We iterate step 2 until the last maturity Δt_m.

[90] We have $\text{PD} = 1 - \mathbf{S}(1)$.

This algorithm works well if the knots $t_m^\star$ exactly match the maturities. It is known as the bootstrap method and is very popular to estimate the survival function from market prices. Let $\{\mathcal{s}(T_1), \ldots, \mathcal{s}(T_M)\}$ be a set of CDS spreads for a given name. Assuming that $T_1 < T_2 < \ldots < T_M$, we consider the piecewise exponential model with $t_m^\star = T_m$. We first estimate $\hat{\lambda}_1$ such that the theoretical spread is equal to $\mathcal{s}(T_1)$. We then calibrate the hazard function in order to retrieve the spread $\mathcal{s}(T_2)$ of the second maturity. This means to consider that $\lambda(t)$ is known and equal to $\hat{\lambda}_1$ until time T_1 whereas $\lambda(t)$ is unknown from T_1 to T_2:

$$\lambda(t) = \begin{cases} \hat{\lambda}_1 & \text{if } t \in]0, T_1] \\ \lambda_2 & \text{if } t \in]T_1, T_2] \end{cases}$$

Estimating $\hat{\lambda}_2$ is therefore straightforward because it is equivalent to solve one equation with one variable. We proceed in a similar way for the other maturities.

Example 38 *We assume that the term structure of interest rates is generated by the Nelson-Siegel model with* $\theta_1 = 5\%$, $\theta_2 = -5\%$, $\theta_3 = 6\%$ *and* $\theta_4 = 10$. *We consider three credit curves, whose CDS spreads expressed in bps are given in the following table:*

Maturity (in years)	#1	#2	#3
1	50	50	350
3	60	60	370
5	70	90	390
7	80	115	385
10	90	125	370

The recovery rate $\mathcal{R}$ *is set to* 40%.

TABLE 3.33: Calibrated piecewise exponential model from CDS prices

Maturity (in years)	#1	#2	#3
1	83.3	83.3	582.9
3	110.1	110.1	637.5
5	140.3	235.0	702.0
7	182.1	289.6	589.4
10	194.1	241.9	498.5

Using the bootstrap method, we obtain results in Table 3.33. We notice that the piecewise exponential model coincide for the credit curves #1 and #2 for $t < 3$ years. This is normal because the CDS spreads of the two credit curves are equal when the maturity is less or equal than 3 years. The third credit curve illustrates that the bootstrap method is highly sensitive to small differences. Indeed, the calibrated intensity parameter varies from 499 to 702 bps while the CDS spreads varies from 350 to 390 bps. Finally, the survival function associated to these 3 bootstrap calibrations are shown in Figure 3.34.

Remark 42 *Other methods for estimating the probability of default are presented in Chapter 19 dedicated to credit scoring models.*

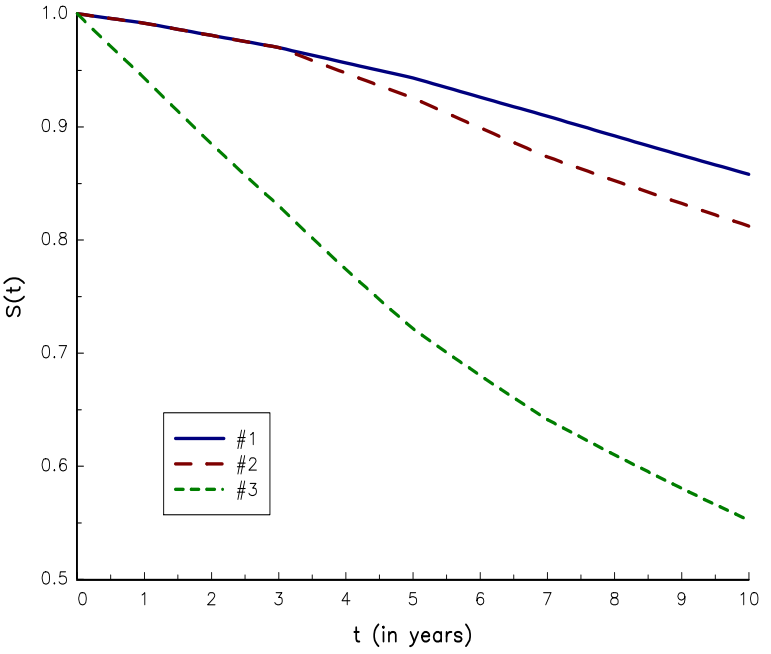


FIGURE 3.34: Calibrated survival function from CDS prices

3.3.3.2 Transition probability matrix

When dealing with risk classes, it is convenient to model a transition probability matrix. For instance, this approach is used for modeling credit rating migration.

Discrete-time modeling We consider a time-homogeneous Markov chain $\mathfrak{R}$, whose transition matrix is $P = (p_{i,j})$. We note $\mathcal{S} = \{1, 2, \ldots, K\}$ the state space of the chain and $p_{i,j}$ is the probability that the entity migrates from rating i to rating j. The matrix P satisfies the following properties:

- $\forall i, j \in \mathcal{S}$, $p_{i,j} \geq 0$;
- $\forall i \in \mathcal{S}$, $\sum_{j=1}^{K} p_{i,j} = 1$.

In credit risk, we generally assume that K is the absorbing state (or the default state), implying that any entity which has reached this state remains in this state. In this case, we have $p_{K,K} = 1$. Let $\mathfrak{R}(t)$ be the value of the state at time t. We define $p(s, i; t, j)$ as the probability that the entity reaches the state j at time t given that it has reached the state i at time s. We have:

$$\begin{aligned} p(s, i; t, j) &= \Pr\{\mathfrak{R}(t) = j \mid \mathfrak{R}(s) = i\} \\ &= p_{i,j}^{(t-s)} \end{aligned}$$

This probability only depends on the duration between s and t because of the Markov property. Therefore, we can restrict the analysis by calculating the n-step transition probability:

$$p_{i,j}^{(n)} = \Pr\{\mathfrak{R}(t+n) = j \mid \mathfrak{R}(t) = i\}$$

and the associated n-step transition matrix $P^{(n)} = \left(p_{i,j}^{(n)}\right)$. For $n = 2$, we obtain:

$$\begin{aligned}
p_{i,j}^{(2)} &= \Pr\{\mathfrak{R}(t+2) = j \mid \mathfrak{R}(t) = i\} \\
&= \sum_{k=1}^{K} \Pr\{\mathfrak{R}(t+2) = j, \mathfrak{R}(t+1) = k \mid \mathfrak{R}(t) = i\} \\
&= \sum_{k=1}^{K} \Pr\{\mathfrak{R}(t+2) = j \mid \mathfrak{R}(t+1) = k\} \cdot \Pr\{\mathfrak{R}(t+1) = k \mid \mathfrak{R}(t) = i\} \\
&= \sum_{k=1}^{K} p_{i,k} \cdot p_{k,j}
\end{aligned}$$

In a similar way, we obtain:

$$p_{i,j}^{(n+m)} = \sum_{k=1}^{K} p_{i,k}^{(n)} \cdot p_{k,j}^{(m)} \qquad \forall n, m > 0 \tag{3.45}$$

This equation is called the Chapman-Kolmogorov equation. In matrix form, we have:

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}$$

with the convention $P^{(0)} = I$. In particular, we have:

$$\begin{aligned}
P^{(n)} &= P^{(n-1)} \cdot P^{(1)} \\
&= P^{(n-2)} \cdot P^{(1)} \cdot P^{(1)} \\
&= \prod_{t=1}^{n} P^{(1)} \\
&= P^{n}
\end{aligned}$$

We deduce that:

$$p(t, i; t+n, j) = p_{i,j}^{(n)} = \mathbf{e}_i^\top P^n \mathbf{e}_j \tag{3.46}$$

When we apply this framework to credit risk, $\mathfrak{R}(t)$ denotes the rating (or the risk class) of the firm at time t, $p_{i,j}$ is the one-period transition probability from rating i to rating j and $p_{i,K}$ is the one-period default probability of rating i. In Table 3.34, we report the S&P one-year transition probability matrix for corporate bonds estimated by Kavvathas (2001). We read the figures as follows[91]: a firm rated AAA has a one-year probability of 92.82% to remain AAA; its probability to become AA is 6.50%; a firm rated CCC defaults one year later with a probability equal to 23.50%; etc. In Tables 3.35 and 3.36, we have reported the two-year and five-year transition probability matrices. We detail below the calculation of $p_{\mathsf{AAA},\mathsf{AAA}}^{(2)}$:

$$\begin{aligned}
p_{\mathsf{AAA},\mathsf{AAA}}^{(2)} &= p_{\mathsf{AAA},\mathsf{AAA}} \times p_{\mathsf{AAA},\mathsf{AAA}} + p_{\mathsf{AAA},\mathsf{AA}} \times p_{\mathsf{AA},\mathsf{AAA}} + p_{\mathsf{AAA},\mathsf{A}} \times p_{\mathsf{A},\mathsf{AAA}} + \\
&\quad p_{\mathsf{AAA},\mathsf{BBB}} \times p_{\mathsf{BBB},\mathsf{AAA}} + p_{\mathsf{AAA},\mathsf{BB}} \times p_{\mathsf{BB},\mathsf{AAA}} + p_{\mathsf{AAA},\mathsf{B}} \times p_{\mathsf{B},\mathsf{AAA}} + \\
&\quad p_{\mathsf{AAA},\mathsf{CCC}} \times p_{\mathsf{CCC},\mathsf{AAA}} \\
&= 0.9283^2 + 0.0650 \times 0.0063 + 0.0056 \times 0.0008 + \\
&\quad 0.0006 \times 0.0005 + 0.0006 \times 0.0004 \\
&= 86.1970\%
\end{aligned}$$

TABLE 3.34: Example of credit migration matrix (in %)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	92.82	6.50	0.56	0.06	0.06	0.00	0.00	0.00
AA	0.63	91.87	6.64	0.65	0.06	0.11	0.04	0.00
A	0.08	2.26	91.66	5.11	0.61	0.23	0.01	0.04
BBB	0.05	0.27	5.84	87.74	4.74	0.98	0.16	0.22
BB	0.04	0.11	0.64	7.85	81.14	8.27	0.89	1.06
B	0.00	0.11	0.30	0.42	6.75	83.07	3.86	5.49
CCC	0.19	0.00	0.38	0.75	2.44	12.03	60.71	23.50
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Source: Kavvathas (2001).

We note $\pi_i^{(n)}$ the probability of the state i at time n:

$$\pi_i^{(n)} = \Pr\left\{\mathfrak{R}(n) = i\right\}$$

and $\pi^{(n)} = \left(\pi_1^{(n)}, \ldots, \pi_K^{(n)}\right)$ the probability distribution. By construction, we have:

$$\pi^{(n+1)} = P^\top \pi^{(n)}$$

The Markov chain $\mathfrak{R}$ admits a stationary distribution $\pi^\star$ if[92]:

$$\pi^\star = P^\top \pi^\star$$

In this case, $\pi_i^\star$ is the limiting probability of state i:

$$\lim_{n\to\infty} p_{k,i}^{(n)} = \pi_i^\star$$

We can interpret $\pi_i^\star$ as the average duration spent by the chain $\mathfrak{R}$ in the state i. Let $\mathcal{T}_i$ be the return period[93] of state i:

$$\mathcal{T}_i = \inf\left\{n : \mathfrak{R}(n) = i \mid \mathfrak{R}(0) = i\right\}$$

The average return period is then equal to:

$$\mathbb{E}\left[\mathcal{T}_i\right] = \frac{1}{\pi_i^\star}$$

For credit migration matrices, there is no stationary distribution because the long-term rating $\mathfrak{R}(\infty)$ is the absorbing state as noted by Jafry and Schuermann:

> "*Given sufficient time, all firms will eventually sink to the default state. This behavior is clearly a mathematical artifact, stemming from the idealized linear, time invariant assumptions inherent in the simple Markov model. In reality the economy (and hence the migration matrix) will change on time-scales far shorter than required to reach the idealized default steady-state proscribed by an assumed constant migration matrix*" (Jafry and Schuermann, 2004, page 2609).

[91] The rows represent the initial rating whereas the columns indicate the final rating.
[92] Not all Markov chains behave in this way, meaning that $\pi^\star$ does not necessarily exist.
[93] This concept plays an important role when designing stress scenarios (see Chapter 18).

TABLE 3.35: Two-year transition probability matrix P^2 (in %)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	86.20	12.02	1.47	0.18	0.11	0.01	0.00	0.00
AA	1.17	84.59	12.23	1.51	0.18	0.22	0.07	0.02
A	0.16	4.17	84.47	9.23	1.31	0.51	0.04	0.11
BBB	0.10	0.63	10.53	77.66	8.11	2.10	0.32	0.56
BB	0.08	0.24	1.60	13.33	66.79	13.77	1.59	2.60
B	0.01	0.21	0.61	1.29	11.20	70.03	5.61	11.03
CCC	0.29	0.04	0.68	1.37	4.31	17.51	37.34	38.45
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

TABLE 3.36: Five-year transition probability matrix P^5 (in %)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	69.23	23.85	5.49	0.96	0.31	0.12	0.02	0.03
AA	2.35	66.96	24.14	4.76	0.86	0.62	0.13	0.19
A	0.43	8.26	68.17	17.34	3.53	1.55	0.18	0.55
BBB	0.24	1.96	19.69	56.62	13.19	5.32	0.75	2.22
BB	0.17	0.73	5.17	21.23	40.72	20.53	2.71	8.74
B	0.07	0.47	1.73	4.67	16.53	44.95	5.91	25.68
CCC	0.38	0.24	1.37	2.92	7.13	18.51	9.92	59.53
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

We note that the survival function $\mathbf{S}_i(t)$ of a firm whose initial rating is the state i is given by:

$$\begin{aligned}\mathbf{S}_i(t) &= 1 - \Pr\{\mathfrak{R}(t) = K \mid \mathfrak{R}(0) = i\} \\ &= 1 - \mathbf{e}_i^\top P^t \mathbf{e}_K \end{aligned} \tag{3.47}$$

In the piecewise exponential model, we recall that the survival function has the following expression:

$$\mathbf{S}(t) = \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t - t_{m-1}^\star\right)}$$

for $t \in \left]t_{m-1}^\star, t_m^\star\right]$. We deduce that $\mathbf{S}(t_m^\star) = \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t_m^\star - t_{m-1}^\star\right)}$, implying that:

$$\ln \mathbf{S}(t_m^\star) = \ln \mathbf{S}\left(t_{m-1}^\star\right) - \lambda_m\left(t_m^\star - t_{m-1}^\star\right)$$

and:

$$\lambda_m = \frac{\ln \mathbf{S}\left(t_{m-1}^\star\right) - \ln \mathbf{S}(t_m^\star)}{t_m^\star - t_{m-1}^\star}$$

It is then straightforward to estimate the piecewise hazard function:

- the knots of the piecewise function are the years $m \in \mathbb{N}^*$;
- for each initial rating i, the hazard function $\lambda_i(t)$ is defined as:

$$\lambda_i(t) = \lambda_{i,m} \qquad \text{if } t \in]m-1, m]$$

where:

$$\begin{aligned}\lambda_{i,m} &= \frac{\ln \mathbf{S}_i(m-1) - \ln \mathbf{S}_i(m)}{m - (m-1)} \\ &= \ln\left(\frac{1 - \mathbf{e}_i^\top P^{m-1} \mathbf{e}_K}{1 - \mathbf{e}_i^\top P^m \mathbf{e}_K}\right)\end{aligned}$$

and $P^0 = I$.

If we consider the credit migration matrix given in Table 3.34 and estimate the piecewise exponential model, we obtain the hazard function[94] $\lambda_i(t)$ shown in Figure 3.35. For good initial ratings, hazard rates are low for short maturities and increase with time. For bad initial ratings, we obtain the opposite effect, because the firm can only improve its rating if it did not default. We observe that the hazard function of all the ratings converges to the same level, which is equal to 102.63 bps. This indicates the long-term hazard rate of the Markov chain, meaning that 1.02% of firms default every year on average.

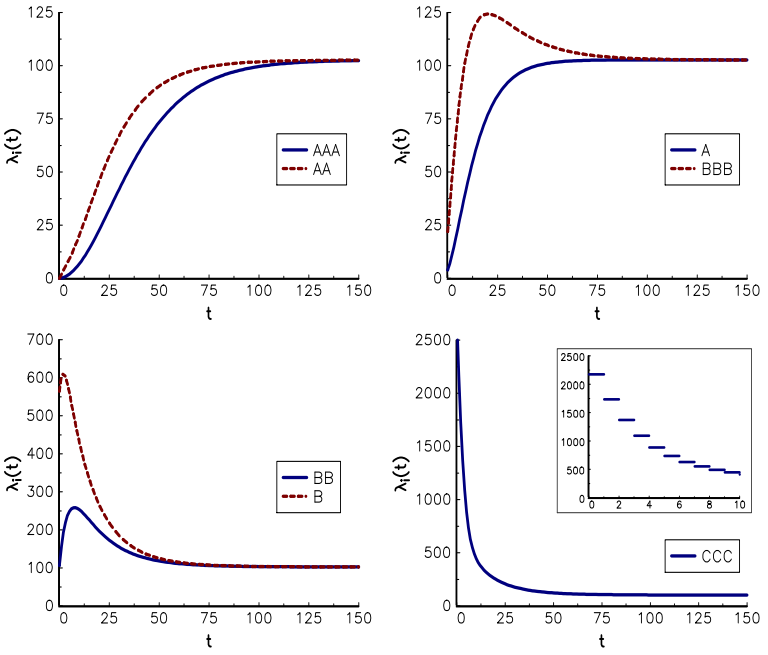


FIGURE 3.35: Estimated hazard function $\lambda_i(t)$ from the credit migration matrix

Continuous-time modeling We now consider the case $t \in \mathbb{R}_+$. We note $P(s;t)$ the transition matrix defined as follows:

$$\begin{aligned} P_{i,j}(s;t) &= p(s,i;t,j) \\ &= \Pr\{\mathfrak{R}(t) = j \mid \mathfrak{R}(s) = i\} \end{aligned}$$

Assuming that the Markov chain is time-homogenous, we have $P(t) = P(0;t)$. Jarrow *et al.* (1997) introduce the generator matrix $\Lambda = (\lambda_{i,j})$ where $\lambda_{i,j} \geq 0$ for all $i \neq j$ and:

$$\lambda_{i,i} = -\sum_{j \neq i}^{K} \lambda_{i,j}$$

In this case, the transition matrix satisfies the following relationship:

$$P(t) = \exp(t\Lambda) \tag{3.48}$$

[94] Contrary to what the graph suggests, $\lambda_i(t)$ is a piecewise constant function (see details of the curve in the fifth panel for very short maturities).

where $\exp(A)$ is the matrix exponential of A. Let us give a probabilistic interpretation of Λ. If we assume that the probability of jumping from rating i to rating j in a short time period Δt is proportional to Δt, we have:

$$p(t, i; t + \Delta t, j) = \lambda_{i,j} \Delta t$$

The matrix form of this equation is $P(t; t + \Delta t) = \Lambda \, \Delta t$. We deduce that:

$$\begin{aligned} P(t + \Delta t) &= P(t) \, P(t; t + \Delta t) \\ &= P(t) \, \Lambda \, \Delta t \end{aligned}$$

and:

$$\mathrm{d}P(t) = P(t) \, \Lambda \, \mathrm{d}t$$

Because we have $\exp(\mathbf{0}) = I$, we obtain the solution $P(t) = \exp(t\Lambda)$. We then interpret $\lambda_{i,j}$ as the instantaneous transition rate of jumping from rating i to rating j.

Remark 43 *In Appendix A.1.1.3, we present the matrix exponential function and its mathematical properties. In particular, we have* $e^{A+B} = e^A e^B$ *and* $e^{A(s+t)} = e^{As} e^{At}$ *where* A *and* B *are two square matrices such that* $AB = BA$ *and* s *and* t *are two real numbers.*

Example 39 *We consider a rating system with three states: A (good rating), B (bad rating) and D (default). The Markov generator is equal to:*

$$\Lambda = \begin{pmatrix} -0.30 & 0.20 & 0.10 \\ 0.15 & -0.40 & 0.25 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}$$

The one-year transition probability matrix is equal to:

$$P(1) = e^{\Lambda} = \begin{pmatrix} 75.16\% & 14.17\% & 10.67\% \\ 10.63\% & 68.07\% & 21.30\% \\ 0.00\% & 0.00\% & 100.00\% \end{pmatrix}$$

For the two-year maturity, we get:

$$P(2) = e^{2\Lambda} = \begin{pmatrix} 58.00\% & 20.30\% & 21.71\% \\ 15.22\% & 47.85\% & 36.93\% \\ 0.00\% & 0.00\% & 100.00\% \end{pmatrix}$$

We verify that $P(2) = P(1)^2$. This derives from the property of the matrix exponential:

$$P(t) = e^{t\Lambda} = \left(e^{\Lambda}\right)^t = P(1)^t$$

The continuous-time framework allows to calculate transition matrices for non-integer maturities, which do not correspond to full years. For instance, the one-month transition probability matrix of the previous example is equal to:

$$P(2) = e^{\frac{1}{12}\Lambda} = \begin{pmatrix} 97.54\% & 1.62\% & 0.84\% \\ 1.21\% & 96.73\% & 2.05\% \\ 0.00\% & 0.00\% & 100.00\% \end{pmatrix}$$

One of the issues with the continuous-time framework is to estimate the Markov generator Λ. One solution consists in using the empirical transition matrix $\hat{P}(t)$, which have

been calculated for a given time horizon t. In this case, the estimate $\hat{\Lambda}$ must satisfy the relationship $\hat{P}(t) = \exp\left(t\hat{\Lambda}\right)$. We deduce that:

$$\hat{\Lambda} = \frac{1}{t}\ln\left(\hat{P}(t)\right)$$

where $\ln A$ is the matrix logarithm of A. However, the matrix $\hat{\Lambda}$ cannot verify the Markov conditions $\hat{\lambda}_{i,j} \geq 0$ for all $i \neq j$ and $\sum_{j=1}^{K} \lambda_{i,j} = 0$. For instance, if we consider the previous S&P transition matrix, we obtain the generator $\hat{\Lambda}$ given in Table 3.37. We notice that six off-diagonal elements of the matrix are negative[95]. This implies that we can obtain transition probabilities which are negative for short maturities. In this case, Israel *et al.* (2001) propose two estimators to obtain a valid generator:

1. the first approach consists in adding the negative values back into the diagonal values:

$$\begin{cases} \bar{\lambda}_{i,j} = \max\left(\hat{\lambda}_{i,j}, 0\right) & i \neq j \\ \bar{\lambda}_{i,i} = \hat{\lambda}_{i,i} + \sum_{j \neq i} \min\left(\hat{\lambda}_{i,j}, 0\right) \end{cases}$$

2. in the second method, we carry forward the negative values on the matrix entries which have the correct sign:

$$\begin{cases} G_i = \left|\hat{\lambda}_{i,i}\right| + \sum_{j \neq i} \max\left(\hat{\lambda}_{i,j}, 0\right) \\ B_i = \sum_{j \neq i} \max\left(-\hat{\lambda}_{i,j}, 0\right) \\ \tilde{\lambda}_{i,j} = \begin{cases} 0 & \text{if } i \neq j \text{ and } \hat{\lambda}_{i,j} < 0 \\ \hat{\lambda}_{i,j} - B_i \left|\hat{\lambda}_{i,j}\right| / G_i & \text{if } G_i > 0 \\ \hat{\lambda}_{i,j} & \text{if } G_i = 0 \end{cases} \end{cases}$$

Using the estimator $\hat{\Lambda}$ and the two previous algorithms, we obtain the valid generators given in Tables 3.39 and 3.40. We find that $\left\|\hat{P} - \exp\left(\bar{\Lambda}\right)\right\|_1 = 11.02 \times 10^{-4}$ and $\left\|\hat{P} - \exp\left(\tilde{\Lambda}\right)\right\|_1 = 10.95 \times 10^{-4}$, meaning that the Markov generator $\tilde{\Lambda}$ is the estimator that minimizes the distance to $\hat{P}$. We can then calculate the transition probability matrix for all maturities, and not only for calendar years. For instance, we report the 207-day transition probability matrix $P\left(\frac{207}{365}\right) = \exp\left(\frac{207}{365}\tilde{\Lambda}\right)$ in Table 3.41.

Remark 44 *The continuous-time framework is more flexible when modeling credit risk. For instance, the expression of the survival function becomes:*

$$\mathbf{S}_i(t) = \Pr\{\mathfrak{R}(t) = K \mid \mathfrak{R}(0) = i\} = 1 - \mathbf{e}_i^\top \exp(t\Lambda)\,\mathbf{e}_K$$

We can therefore calculate the probability density function in an easier way:

$$f_i(t) = -\partial_t \mathbf{S}_i(t) = \mathbf{e}_i^\top \Lambda \exp(t\Lambda)\,\mathbf{e}_K$$

For illustration purposes, we represent the probability density function of S&P ratings estimated with the valid generator $\tilde{\Lambda}$ in Figure 3.36.

[95] We have also calculated the estimator described in Israel *et al.* (2001):

$$\breve{\Lambda} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\left(\hat{P} - I\right)^n}{n}$$

We do not obtain the same matrix as for the estimator $\hat{\Lambda}$, but there are also six negative off-diagonal elements (see Table 3.38).

TABLE 3.37: Markov generator $\hat{\Lambda}$ (in bps)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−747.49	703.67	35.21	3.04	6.56	−0.79	−0.22	0.02
AA	67.94	−859.31	722.46	51.60	2.57	10.95	4.92	−1.13
A	7.69	245.59	−898.16	567.70	53.96	20.65	−0.22	2.80
BBB	5.07	21.53	650.21	−1352.28	557.64	85.56	16.08	16.19
BB	4.22	10.22	41.74	930.55	−2159.67	999.62	97.35	75.96
B	−0.84	11.83	30.11	8.71	818.31	−1936.82	539.18	529.52
CCC	25.11	−2.89	44.11	84.87	272.05	1678.69	−5043.00	2941.06
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 3.38: Markov generator $\breve{\Lambda}$ (in bps)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−745.85	699.11	38.57	2.80	6.27	−0.70	−0.16	−0.05
AA	67.54	−855.70	716.56	54.37	2.81	10.81	4.62	−1.01
A	7.77	243.62	−891.46	560.45	56.33	20.70	0.07	2.53
BBB	5.06	22.68	641.55	−1335.03	542.46	91.05	16.09	16.15
BB	4.18	10.12	48.00	903.40	−2111.65	965.71	98.28	81.96
B	−0.56	11.61	29.31	19.39	789.99	−1887.69	491.46	546.49
CCC	23.33	−1.94	42.22	81.25	272.44	1530.66	−4725.22	2777.25
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 3.39: Markov generator $\bar{\Lambda}$ (in bps)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−748.50	703.67	35.21	3.04	6.56	0.00	0.00	0.02
AA	67.94	−860.44	722.46	51.60	2.57	10.95	4.92	0.00
A	7.69	245.59	−898.38	567.70	53.96	20.65	0.00	2.80
BBB	5.07	21.53	650.21	−1352.28	557.64	85.56	16.08	16.19
BB	4.22	10.22	41.74	930.55	−2159.67	999.62	97.35	75.96
B	0.00	11.83	30.11	8.71	818.31	−1937.66	539.18	529.52
CCC	25.11	0.00	44.11	84.87	272.05	1678.69	−5045.89	2941.06
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 3.40: Markov generator $\tilde{\Lambda}$ (in bps)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−747.99	703.19	35.19	3.04	6.55	0.00	0.00	0.02
AA	67.90	−859.88	721.98	51.57	2.57	10.94	4.92	0.00
A	7.69	245.56	−898.27	567.63	53.95	20.65	0.00	2.80
BBB	5.07	21.53	650.21	−1352.28	557.64	85.56	16.08	16.19
BB	4.22	10.22	41.74	930.55	−2159.67	999.62	97.35	75.96
B	0.00	11.83	30.10	8.71	818.14	−1937.24	539.06	529.40
CCC	25.10	0.00	44.10	84.84	271.97	1678.21	−5044.45	2940.22
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 3.41: 207-day transition probability matrix (in %)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	95.85	3.81	0.27	0.03	0.04	0.00	0.00	0.00
AA	0.37	95.28	3.90	0.34	0.03	0.06	0.02	0.00
A	0.04	1.33	95.12	3.03	0.33	0.12	0.00	0.02
BBB	0.03	0.14	3.47	92.75	2.88	0.53	0.09	0.11
BB	0.02	0.06	0.31	4.79	88.67	5.09	0.53	0.53
B	0.00	0.06	0.17	0.16	4.16	89.84	2.52	3.08
CCC	0.12	0.01	0.23	0.45	1.45	7.86	75.24	14.64
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

3.3.3.3 Structural models

The previous approaches are purely statistical and are called reduced-form models. We now consider economic models for modeling default times. These approaches are based on accounting and market data and are called structural models.

The Merton model The structural approach of credit risk has been formalized by Merton (1974). In this framework, the bond holders will liquidate the corporate firm if the asset value $A(t)$ goes below a threshold B related to the total amount of debt. The underlying idea is that bond holders monitor the asset value and compare $A(t)$ to the default barrier B.

Merton (1974) assumes that the dynamics of the assets $A(t)$ follows a geometric Brownian motion:

$$\mathrm{d}A(t) = \mu_A A(t)\,\mathrm{d}t + \sigma_A A(t)\,\mathrm{d}W(t)$$

where $A(0) = A_0$. The default occurs if the asset value $A(t)$ falls under the threshold B:

$$\boldsymbol{\tau} := \inf\{t : A(t) \leq B\}$$

In this case, the bond holders receive $A(T)$, and lose $B - A(T)$. The payoff of bond holders is then equal to:

$$D = B - \max(B - A(T), 0)$$

where D is the debt value of maturity T. The holding of a risky bond can be interpreted as a trading strategy where we have bought a zero-coupon and financed the cost by selling a put on $A(t)$ with an exercise price B and a maturity T. From the viewpoint of the equity holders, the payoff is equal to $\max(A(T) - D, 0)$. The holding of an equity share can be interpreted as a trading strategy where we have bought a call option with a strike equal to the debt value D. It follows that the current value E_0 of the equity is:

$$\begin{aligned} E_0 &= e^{-rT} \cdot \mathbb{E}[\max(A(T) - D, 0)] \\ &= A_0\Phi(d_1) - e^{-rT}D\Phi(d_2) \end{aligned}$$

where:

$$d_1 = \frac{\ln A_0 - \ln D + rT}{\sigma_A\sqrt{T}} + \frac{1}{2}\sigma_A\sqrt{T}$$

and $d_2 = d_1 - \sigma_A\sqrt{T}$. We notice that the equity value depends on the current asset value A_0, the leverage ratio $L = D/A_0$, the asset volatility σ_A and the time of repayment T.

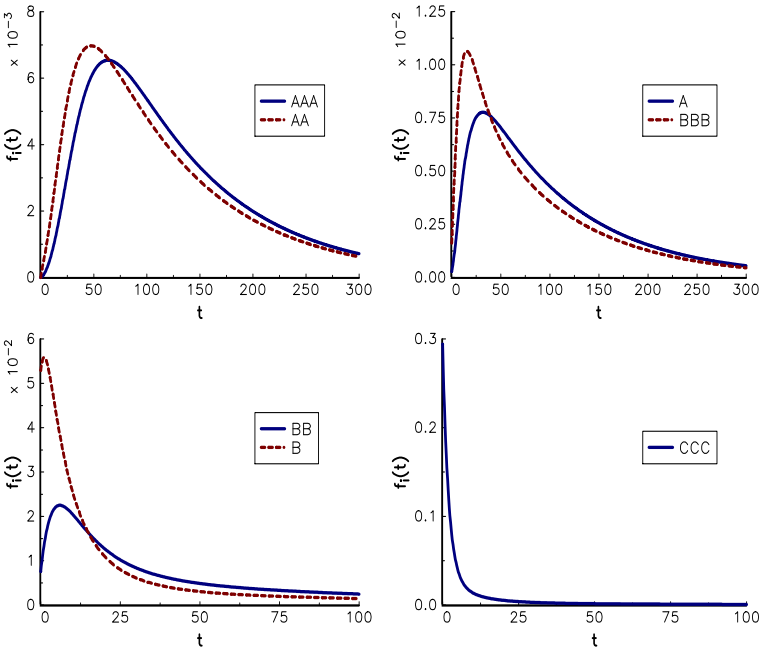


FIGURE 3.36: Probability density function $f_i(t)$ of S&P ratings

The KMV implementation In the nineties, the Merton model has been implemented by KMV[96] with a lot of success. The underlying idea of the KMV implementation is to estimate the default probability of a firm. One of the difficulties is to estimate the asset volatility σ_A. However, Jones *et al.* (1984) show that it is related to the equity volatility σ_E. Indeed, we have $E(t) = \mathcal{C}(t, A(t))$, implying that:

$$\begin{aligned} \mathrm{d}E(t) &= \partial_t \mathcal{C}(t, A(t))\,\mathrm{d}t + \mu_A A(t)\,\partial_A \mathcal{C}(t, A(t))\,\mathrm{d}t + \\ &\quad \frac{1}{2}\sigma_A^2 A^2(t)\,\partial_A^2 \mathcal{C}(t, A(t))\,\mathrm{d}t + \sigma_A A(t)\,\partial_A \mathcal{C}(t, A(t))\,\mathrm{d}W(t) \end{aligned}$$

Since the stochastic term is also equal to $\sigma_E E(t)\,\mathrm{d}W(t)$, we obtain the following equality at time $t = 0$:

$$\sigma_E E_0 = \sigma_A A_0 \Phi(d_1)$$

Therefore, Crosbie and Bohn (2002) deduce the following system of equations:

$$\begin{cases} A_0\Phi(d_1) - e^{-rT} D\Phi(d_2) - E_0 = 0 \\ \sigma_E E_0 - \sigma_A A_0 \Phi(d_1) = 0 \end{cases} \tag{3.49}$$

Once we have estimated A_0 and σ_A, we can calculate the survival function:

$$\begin{aligned} \mathbf{S}(t) &= \Pr\{A(t) \geq D \mid A(0) = A_0\} \\ &= \Phi\left(\frac{\ln A_0 - \ln D + \mu_A t}{\sigma_A \sqrt{t}} + \frac{1}{2}\sigma_A \sqrt{t}\right) \end{aligned}$$

and deduce the probability of default $\mathbf{F}(t) = 1 - \mathbf{S}(t)$ and the distance to default $\mathrm{DD}(t) = \Phi^{-1}(\mathbf{S}(t))$.

[96] KMV was a company dedicated to credit risk modeling, and was founded by Stephen Kealhofer, John McQuown and Oldrich Vasícek. In 2002, they sold KMV to Moody's.

Example 40 *Crosbie and Bohn (2002) assume that the market capitalization E_0 of the firm is \$3 bn, its debt liability D is \$10 bn, the corresponding equity volatility σ_E is equal to 40%, the maturity T is one year and the expected return μ_A is set to 7%.*

Using an interest rate $r = 5\%$ and solving Equation (3.49), we find that the asset value A_0 is equal to \$12.512 bn and the implied asset volatility σ_A is equal to 9.609%. Therefore, we can calculate the distance-to-default $\mathrm{DD}(1) = 3.012$ and the one-year probability $\mathrm{PD}(1) = 12.96$ bps. In Figure 3.37, we report the probability of default for different time horizons. We also show the impact of the equity volatility σ_E and the expected return μ_A, which can be interpreted as a return-on-equity ratio (ROE). We verify that the probability of default is an increasing function of the volatility risk and a decreasing function of the profitability.

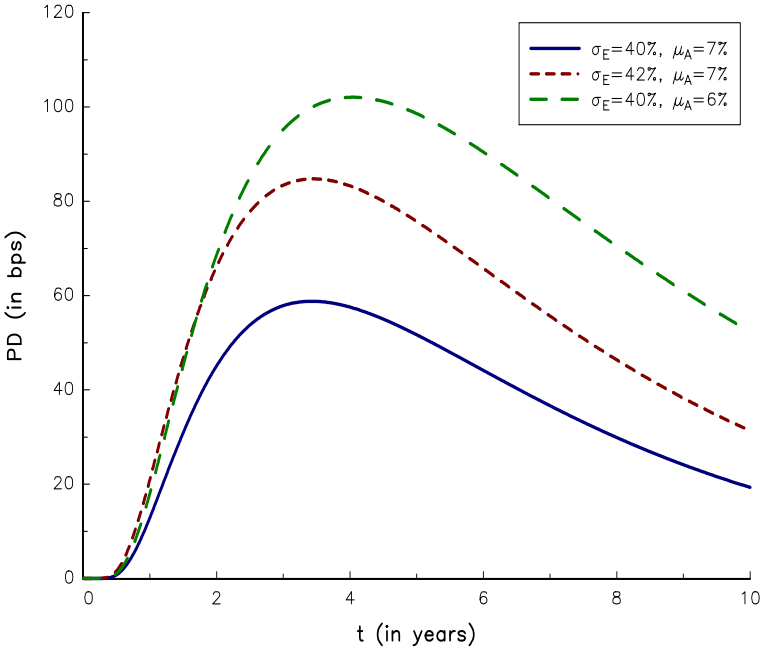


FIGURE 3.37: Probability of default in the KMV model

Remark 45 *The KMV model is more complex than the presentation above. In particular, the key variable is not the probability of default, but the distance-to-default (see Figure 3.38). Once this measure is calculated, it is converted into an expected default frequency (EDF) by considering an empirical distribution of* PD *conditionally to the distance-to-default. For instance,* $\mathrm{DD}(1) = 4$ *is equivalent to* $\mathrm{PD}(1) = 100$ *bps (Crosbie and Bohn, 2002).*

The CreditGrades implementation The CreditGrades approach is an extension of the Merton model, uses the framework of Black and Cox (1976) and has been developed by Finkelstein *et al.* (2002). They assume that the asset-per-share value $A(t)$ is a geometric Brownian motion without drift:

$$\mathrm{d}A(t) = \sigma_A A(t)\,\mathrm{d}W(t)$$

whereas the default barrier B is defined as the recovery value of bond holders. B is equal to the product $\boldsymbol{\mathcal{R}} \cdot D$, where $\boldsymbol{\mathcal{R}} \in [0,1]$ is the recovery rate and D is the debt-per-share value.

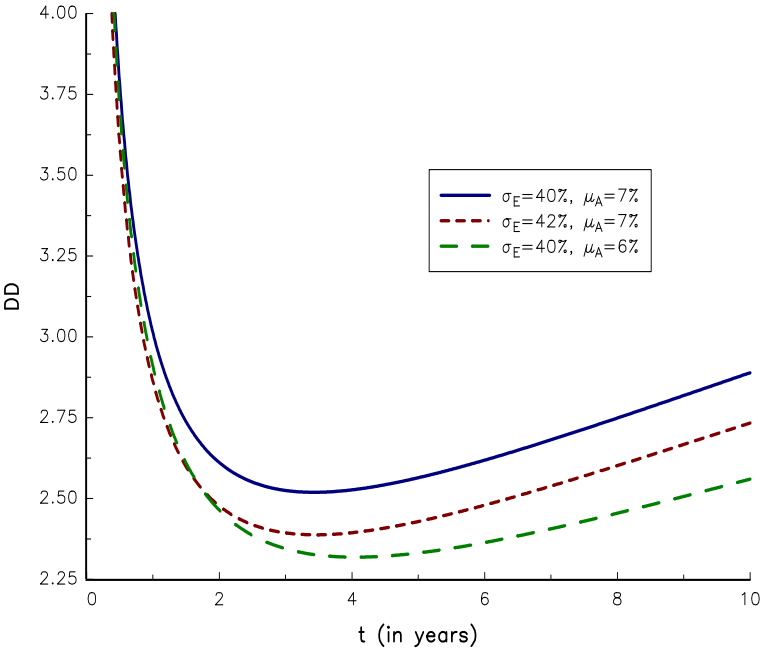


FIGURE 3.38: Distance-to-default in the KMV model

They also assume that $\boldsymbol{\mathcal{R}}$ and $A(t)$ are independent and $\boldsymbol{\mathcal{R}} \sim \mathcal{LN}(\mu_{\boldsymbol{\mathcal{R}}}, \sigma_{\boldsymbol{\mathcal{R}}})$. We recall that the default time is defined by:

$$\boldsymbol{\tau} := \inf\{t \geq 0 : t \in \mathcal{D}\}$$

where $\mathcal{D} = \{A(t) \leq B\}$. Since we have $A(t) = A_0 e^{\sigma_A W(t) - \sigma_A^2 t/2}$ and $B = De^{\mu_{\boldsymbol{\mathcal{R}}} + \sigma_{\boldsymbol{\mathcal{R}}} \varepsilon}$ where $\varepsilon \sim \mathcal{N}(0,1)$, it follows that:

$$\mathcal{D} = \left\{ A_0 e^{\sigma_A W(t) - \sigma_A^2 t/2} \leq De^{\mu_{\boldsymbol{\mathcal{R}}} + \sigma_{\boldsymbol{\mathcal{R}}} \varepsilon} \right\}$$

The authors introduce the average recovery rate $\bar{\boldsymbol{\mathcal{R}}} = \mathbb{E}[\boldsymbol{\mathcal{R}}] = e^{\mu_{\boldsymbol{\mathcal{R}}} + \sigma_{\boldsymbol{\mathcal{R}}}^2/2}$. We deduce that:

$$\begin{aligned} \mathcal{D} &= \left\{ A_0 e^{\sigma_A W(t) - \sigma_A^2 t/2} \leq \bar{\boldsymbol{\mathcal{R}}} De^{\sigma_{\boldsymbol{\mathcal{R}}} \varepsilon - \sigma_{\boldsymbol{\mathcal{R}}}^2/2} \right\} \\ &= \left\{ A_0 e^{\sigma_A W(t) - \sigma_A^2 t/2 - \sigma_{\boldsymbol{\mathcal{R}}} \varepsilon + \sigma_{\boldsymbol{\mathcal{R}}}^2/2} \leq \bar{\boldsymbol{\mathcal{R}}} D \right\} \end{aligned} \tag{3.50}$$

Finkelstein *et al.* (2002) introduce the process $X(t)$ defined by:

$$X(t) = \sigma_A W(t) - \frac{1}{2}\sigma_A^2 t - \sigma_{\boldsymbol{\mathcal{R}}} \varepsilon - \frac{1}{2}\sigma_{\boldsymbol{\mathcal{R}}}^2$$

It follows that Inequality (3.50) becomes:

$$\mathcal{D} = \left\{ X(t) \leq \ln\left(\frac{\bar{\boldsymbol{\mathcal{R}}} D}{A_0 e^{\sigma_{\boldsymbol{\mathcal{R}}}^2}} \right) \right\}$$

By assuming that $X(t)$ can be approximated by a geometric Brownian motion with drift $-\sigma_A^2/2$ and diffusion rate σ_A, we can show that[97]:

$$\mathbf{S}(t) = \Phi\left(-\frac{\sigma(t)}{2} + \frac{\ln\varphi}{\sigma(t)}\right) - \varphi\Phi\left(-\frac{\sigma(t)}{2} - \frac{\ln\varphi}{\sigma(t)}\right)$$

where $\sigma(t) = \sqrt{\sigma_A^2 t + \sigma_{\boldsymbol{\mathcal{R}}}^2}$ and:

$$\varphi = \frac{A_0 e^{\sigma_{\boldsymbol{\mathcal{R}}}^2}}{\bar{\boldsymbol{\mathcal{R}}} D}$$

This survival function is then calibrated by assuming that $A_0 = S_0 + \bar{\boldsymbol{\mathcal{R}}} D$ and:

$$\sigma_A = \sigma_S \frac{S^\star}{S^\star + \bar{\boldsymbol{\mathcal{R}}} D}$$

where S_0 is the current stock price, $S^\star$ is the reference stock price and σ_S is the stock (implied or historical) volatility. All the parameters (S_0, $S^\star$, σ_S, $\bar{\boldsymbol{\mathcal{R}}}$, D) are easy to calibrate, except the volatility of the recovery rate $\sigma_{\boldsymbol{\mathcal{R}}}$. We have:

$$\sigma_{\boldsymbol{\mathcal{R}}}^2 = \text{var}(\ln \boldsymbol{\mathcal{R}}) = \text{var}(\ln B)$$

We deduce that $\sigma_{\boldsymbol{\mathcal{R}}}$ is the uncertainty of the default barrier B.

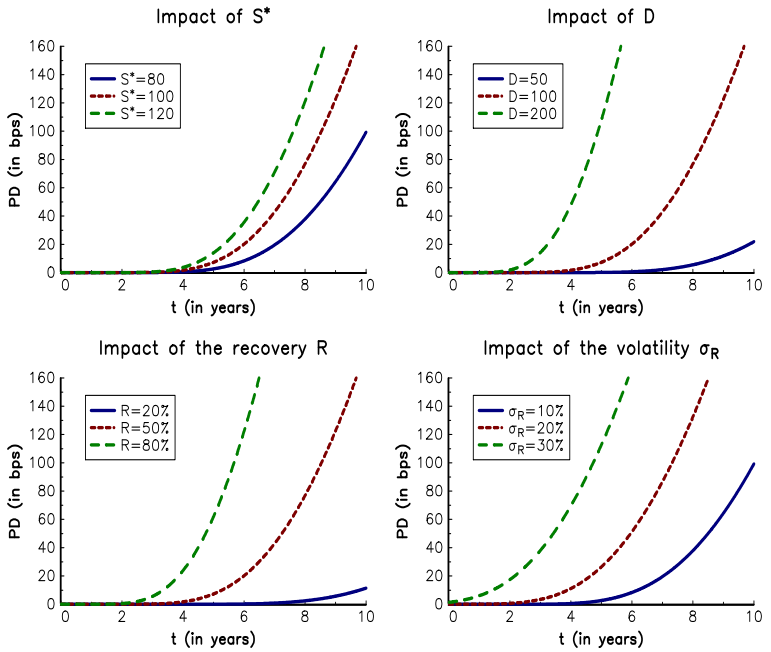


FIGURE 3.39: Probability of default in the CreditGrades model

[97] By considering the reflection principle and Equation (A.24) defined on page 1074, we deduce that:

$$\Pr\left\{\inf_{s\leq t} \mu s + \sigma W(s) > c\right\} = \Phi\left(\frac{\mu t - c}{\sigma\sqrt{t}}\right) - e^{2\mu c/\sigma^2}\Phi\left(\frac{\mu t + c}{\sigma\sqrt{t}}\right)$$

The expression of $\mathbf{S}(t)$ is obtained by setting $\mu = -\sigma_A^2/2$, $\sigma = \sigma_A$ and $c = \ln\left(\bar{\boldsymbol{\mathcal{R}}} D\right) - \ln\left(A_0 e^{\sigma_{\boldsymbol{\mathcal{R}}}^2}\right)$, and using the change of variable $u = t + \sigma_{\boldsymbol{\mathcal{R}}}^2/\sigma_A^2$.

In Figure 3.39, we illustrate the CreditGrades model by computing the probability of default when $S_0 = 100$, $S^\star = 100$, $\sigma_S = 20\%$, $\bar{\mathcal{R}} = 50\%$, $\sigma_{\mathcal{R}} = 10\%$ and $D = 100$. We notice that $\text{PD}(t)$ is an increasing function of $S^\star$, σ_S, $\bar{\mathcal{R}}$, and $\sigma_{\mathcal{R}}$. The impact of the recovery rate may be curious, but bond holders may be encouraged to cause the default when the recovery rate is high.

Relationship with intensity (or reduced-form) models Let $\lambda(s)$ be a positive continuous process. We define the default time by $\boldsymbol{\tau} := \inf\left\{t \geq 0 : \int_0^t \lambda(s)\,\mathrm{d}s \geq \theta\right\}$ where θ is a standard exponential random variable. We have:

$$\begin{aligned}
\mathbf{S}(t) &= \Pr\{\boldsymbol{\tau} > t\} \\
&= \Pr\left\{\int_0^t \lambda(s)\,\mathrm{d}s \leq \theta\right\} \\
&= \mathbb{E}\left[\exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right)\right]
\end{aligned}$$

Let $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ be the integrated hazard function. If $\lambda(s)$ is deterministic, we obtain $\mathbf{S}(t) = \exp(-\Lambda(t))$. In particular, if $\lambda(s)$ is a piecewise constant function, we obtain the piecewise exponential model.

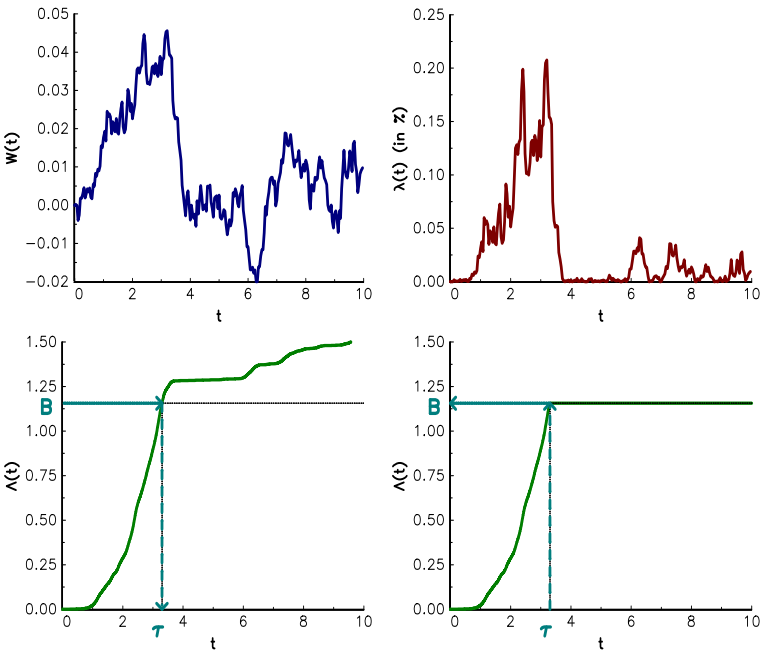


FIGURE 3.40: Intensity models and the default barrier issue

We now consider the stochastic case $\lambda(t) = \sigma W^2(t)$ where $W(t)$ is a Brownian motion. In Figure 3.40, we illustrate the simulation mechanism of defaults. First, we simulate the exponential variable B. In our example, it is equal to 1.157. Second, we simulate the Brownian motion $W(t)$ (top/left panel). Then, we calculate $\lambda(t)$ where $\sigma = 1.5\%$ (top/right panel), and the integrated hazard function $\Lambda(t)$ (bottom/left panel). Finally, we determine the default time when the integrated hazard function crosses the barrier B. In our example, $\boldsymbol{\tau}$ is equal to 3.30. In fact, the simulation mechanism may be confusing. Indeed, we have the

impression that we know the barrier B, implying that the default is predictable. In intensity models, this is the contrary. We don't know the stochastic barrier B, but the occurrence of the default unveils the barrier B as illustrated in the bottom/right panel in Figure 3.40. In structural models, we assume that the barrier B is known and we can predict the default time because we observe the distance to the barrier. Intensity and structural models are then the two faces of the same coin. They use the same concept of default barrier, but its interpretation is completely different.

3.3.4 Default correlation

In this section, we consider the modeling of default correlations, which corresponds essentially to two approaches: the copula model and the factor model. Then, we see how to estimate default correlations. Finally, we show how to consider the dependence of default times in the pricing of basket derivatives.

3.3.4.1 The copula model

Copula functions are extensively studied in Chapter 11, and we invite the reader to examine this chapter before to go further. Let $\mathbf{F}$ be the joint distribution of the random vector $(X_1, \ldots, X_n)$, we show on page 719 that $\mathbf{F}$ admits a copula representation:

$$\mathbf{F}(x_1, \ldots, x_n) = \mathbf{C}(\mathbf{F}_1(x_1), \ldots, \mathbf{F}_n(x_n))$$

where $\mathbf{F}_i$ is the marginal distribution of X_i and $\mathbf{C}$ is the copula function associated to $\mathbf{F}$. Since there is a strong relationship between probability distributions and survival functions, we can also show that the survival function $\mathbf{S}$ of the random vector $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ has a copula representation:

$$\mathbf{S}(t_1, \ldots, t_n) = \breve{\mathbf{C}}(\mathbf{S}_1(t_1), \ldots, \mathbf{S}_n(t_n))$$

where $\mathbf{S}_i$ is the survival function of $\boldsymbol{\tau}_i$ and $\breve{\mathbf{C}}$ is the survival copula associated to $\mathbf{S}$. The copula $\breve{\mathbf{C}}$ is unique if the marginals are continuous. The copula functions $\mathbf{C}$ and $\breve{\mathbf{C}}$ are not necessarily the same, except when the copula $\mathbf{C}$ is radially symmetric (Nelsen, 2006). This is for example the case of the Normal (or Gaussian) copula and the Student's t copula. Since these two copula functions are the only ones that are really used by professionals[98], we assume that $\breve{\mathbf{C}} = \mathbf{C}$ in the sequel.

The Basel model We have seen that the Basel framework for modeling the credit risk is derived from the Merton model. Let $Z_i \sim \mathcal{N}(0,1)$ be the (normalized) asset value of the i^{th} firm. In the Merton model, the default occurs when Z_i is below a non-stochastic barrier B_i:

$$D_i = 1 \Leftrightarrow Z_i \leq B_i$$

The Basel Committee assumes that $Z_i = \sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i$ where $X \sim \mathcal{N}(0,1)$ is the systematic risk factor and $\varepsilon_i \sim \mathcal{N}(0,1)$ is the specific risk factor. We have shown that the default barrier B_i is equal to $\Phi^{-1}(p_i)$ where p_i is the unconditional default probability. We have also demonstrated that the conditional default probability is equal to:

$$p_i(X) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}X}{\sqrt{1-\rho}}\right)$$

[98] They can also use some Archimedean copulas that are not radially symmetric such as the Clayton copula, but it generally concerns credit portfolios with a small number of exposures.

Remark 46 *In the Basel framework, we assume a fixed maturity. If we introduce the time dimension, we obtain:*

$$\begin{aligned} p_i(t) &= \Pr\{\boldsymbol{\tau}_i \leq t\} \\ &= 1 - S_i(t) \end{aligned}$$

and:

$$p_i(t, X) = \Phi\left(\frac{\Phi^{-1}(1 - \mathbf{S}_i(t)) - \sqrt{\rho}X}{\sqrt{1-\rho}}\right)$$

where $S_i(t)$ is the survival function of the i^{th} firm.

The vector of assets $Z = (Z_1, \ldots, Z_n)$ is Gaussian with a constant covariance matrix $\mathbb{C} = \mathbb{C}_n(\rho)$:

$$\mathbb{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

It follows that the joint default probability is:

$$\begin{aligned} p_{1,\ldots,n} &= \Pr\{D_1 = 1, \ldots, D_n = 1\} \\ &= \Pr\{Z_1 \leq B_1, \ldots, Z_n \leq B_n\} \\ &= \Phi(B_1, \ldots, B_n; \mathbb{C}) \end{aligned}$$

Since we have $B_i = \Phi^{-1}(p_i)$, we deduce that the Basel copula between the default indicator functions is a Normal copula, whose parameters are a constant correlation matrix $\mathbb{C}_n(\rho)$:

$$\begin{aligned} p_{1,\ldots,n} &= \Phi\left(\Phi^{-1}(p_1), \ldots, \Phi^{-1}(p_n); \mathbb{C}\right) \\ &= \mathbf{C}(p_1, \ldots, p_n; \mathbb{C}_n(\rho)) \end{aligned}$$

Let us now consider the dependence between the survival times:

$$\begin{aligned} \mathbf{S}(t_1, \ldots, t_n) &= \Pr\{\boldsymbol{\tau}_1 > t_1, \ldots, \boldsymbol{\tau}_n > t_n\} \\ &= \Pr\left\{Z_1 > \Phi^{-1}(p_1(t_1)), \ldots, Z_n > \Phi^{-1}(p_n(t_n))\right\} \\ &= \mathbf{C}(1 - p_1(t_1), \ldots, 1 - p_n(t_n); \mathbb{C}) \\ &= \mathbf{C}(\mathbf{S}_1(t_1), \ldots, \mathbf{S}_n(t_n); \mathbb{C}_n(\rho)) \end{aligned}$$

The dependence between the default times is again the Normal copula with the matrix of parameters $\mathbb{C}_n(\rho)$.

Extension to other copula models The Basel model assumes that the asset correlation is the same between the different firms. A first extension is to consider that the dependence between the default times remain a Normal copula, but with a general correlation matrix:

$$\mathbb{C} = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,n} \\ & 1 & & \vdots \\ & & \ddots & \rho_{n-1,n} \\ & & & 1 \end{pmatrix} \tag{3.51}$$

This approach is explicitly proposed by Li (2000), but it was already implemented in CreditMetrics (Gupton *et al.*, 1997). The correlation matrix can be estimated using a structural model or approximated by the correlation of stock returns. However, this approach is only valid for publicly traded companies and is not always stable. This is why professionals prefer to use direct extensions of the one-factor model.

Let X_j be a Gaussian factor where $j = 1, \ldots, m$. We assume that the asset value Z_i depends on one of these common risk factors:

$$Z_i = \sum_{j=1}^{m} \beta_{i,j} X_j + \varepsilon_i \tag{3.52}$$

with $\sum_{j=1}^{m} \mathbb{1}\{\beta_{i,j} > 0\} = 1$. We assume that the common risk factors are correlated with each other, but they are independent of the specific risks $(\varepsilon_1, \ldots, \varepsilon_n)$, which are by definition not correlated. For instance, X_j can represent the systematic risk factor of the j^{th} sector or industry. Of course, we can extend this approach to a higher dimension such as sector $\times$ region. For example, if we consider three sectors (S_1, S_2 and S_3) and two geographical regions (R_1 and R_2), we obtain six common risk factors:

	S_1	S_2	S_3
R_1	X_1	X_2	X_3
R_2	X_4	X_5	X_6

These risk factors can then be seen as composite sectors. We note $\text{map}(i)$ the mapping function, which indicates the composite sector j (or the risk factor j): $\text{map}(i) = j$ if $i \in X_j$. We assume that the dependence between the default times $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ is a Normal copula function, whose correlation matrix $\mathbb{C}$ is equal to:

$$\mathbb{C} = \begin{pmatrix} 1 & \rho(\text{map}(1), \text{map}(2)) & \cdots & \rho(\text{map}(1), \text{map}(n)) \\ & 1 & & \vdots \\ & & \ddots & \rho(\text{map}(n-1), \text{map}(n)) \\ & & & 1 \end{pmatrix} \tag{3.53}$$

In practice, we have $m \ll n$ and many elements of the correlation matrix $\mathbb{C}$ are equal. In fact, there are only $m \times (m+1)/2$ different values, which correspond to inter-sector and intra-sector correlations.

Example 41 *Let us consider the case of four sectors:*

Factor	X_1	X_2	X_3	X_4
X_1	*30%*	***20%***	***10%***	***0%***
X_2		*40%*	***30%***	***20%***
X_3			*50%*	***10%***
X_4				*60%*

The inter-sector correlations are indicated in ***bold****, whereas the intra-sector correlations are underlined.*

If the portfolio is composed of seven loans of corporate firms that belong to the following sectors:

i	1	2	3	4	5	6	7
$j = \text{map}(i)$	1	1	2	3	3	3	4

we obtain the following correlation matrix:

$$\mathbb{C} = \begin{pmatrix} 1.00 & 0.30 & 0.20 & 0.10 & 0.10 & 0.10 & 0.00 \\ & 1.00 & 0.20 & 0.10 & 0.10 & 0.10 & 0.00 \\ & & 1.00 & 0.30 & 0.30 & 0.30 & 0.20 \\ & & & 1.00 & 0.50 & 0.50 & 0.10 \\ & & & & 1.00 & 0.50 & 0.10 \\ & & & & & 1.00 & 0.10 \\ & & & & & & 1.00 \end{pmatrix}$$

Simulation of copula models With the exception of the Normal copula with a constant correlation matrix and an infinitely fine-grained portfolio, we cannot calculate analytically the value-at-risk or the expected shortfall of the portfolio loss. In this case, we consider Monte Carlo methods, and we use the method of transformation for simulating copula functions[99]. Since we have $\mathbf{S}_i(\boldsymbol{\tau}_i) \sim \mathcal{U}_{[0,1]}$, the simulation of correlated default times is obtained with the following algorithm:

1. we simulate the random vector $(u_1, \ldots, u_n)$ from the copula function $\mathbf{C}$;
2. we set $\boldsymbol{\tau}_i = \mathbf{S}_i^{-1}(u_i)$.

In many cases, we don't need to simulate the default time $\boldsymbol{\tau}_i$, but the indicator function $D_i(t_i) = \mathbb{1}\{\boldsymbol{\tau}_i \leq t_i\}$. Indeed, D_i is a Bernoulli random variable with parameter $\mathbf{F}_i(t) = 1 - \mathbf{S}_i(t)$, implying that $D(t) = (D_1(t_1), \ldots, D_n(t_n))$ is a Bernoulli random vector with parameter $p(t) = (p_1(t_1), \ldots, p_n(t_n))$. Since the copula of $D(t)$ is the copula of the random vector $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$, we obtain the following algorithm to simulate correlated indicator functions:

1. we simulate the random vector $(u_1, \ldots, u_n)$ from the copula function $\mathbf{C}$;
2. we set $D_i(t_i) = 1$ if $u_i > \mathbf{S}_i(t_i)$.

In the case of the Normal copula, the simulation of $u = (u_1, \ldots, u_n)$ requires calculating the Cholesky decomposition of the correlation matrix $\mathbb{C}$. However, this approach is valid for a small size n of the credit portfolio, because we are rapidly limited by the memory storage capacity of the computer. In a 32-bit computer, the storage of a double requires 8 bytes, meaning that the storage of a $n \times n$ Cholesky matrix requires 78.125 KB if $n = 100$, 7.629 MB if $n = 1\,000$, 762.94 MB if $n = 10\,000$, etc. It follows that the traditional Cholesky algorithm is not adapted when considering a large credit portfolio. However, if we consider the Basel model, we can simulate the correlated default times using the following [BASEL] algorithm:

1. we simulate $n + 1$ Gaussian independent random variables X and $(\varepsilon_1, \ldots, \varepsilon_n)$;
2. we simulate the Basel copula function:

$$(u_1, \ldots, u_n) = \left(\Phi\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_1\right), \ldots, \Phi\left(\sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_n\right)\right) \tag{3.54}$$

3. we set $\boldsymbol{\tau}_i = \mathbf{S}_i^{-1}(u_i)$.

[99] See Section 13.1.3.2 on page 802.

The [BASEL] algorithm is the efficient way to simulate the one-factor model and demonstrates that we don't always need to use the Cholesky decomposition for simulating the Normal (or the Student's t) copula function. Let us generalize the [BASEL] algorithm when we consider the Normal copula with the correlation matrix given by Equation (3.53). The eigendecomposition of $\mathbb{C}$ is equal to $V\Lambda V^\top$, where V is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. Let u be a vector of n independent Gaussian standardized random numbers. Then, $Z = V\Lambda^{1/2}u$ is a Gaussian vector with correlation $\mathbb{C}$. We note $\mathbb{C}^\star = \left(\rho^\star_{j_1,j_2}\right)$ the $m \times m$ correlation matrix based on intra- and inter-sector correlations[100] and we consider the corresponding eigendecomposition $\mathbb{C}^\star = V^\star\Lambda^\star V^{\star\top}$. Let $X^\star$ be a $m \times 1$ Gaussian standardized random vector. It follows that the random vector $Z = (Z_1, \ldots, Z_n)$ is a Gaussian random vector with correlation matrix $\mathbb{C} = \text{map}\left(\mathbb{C}^\star\right)$ where[101]:

$$Z_i = \sum_{j=1}^{m} A^\star_{\text{map}(i),j} X^\star_j + \sqrt{1 - \rho^\star_{\text{map}(i),\text{map}(i)}}\varepsilon_i$$

and $A^\star = V^\star \left(\Lambda^\star\right)^{1/2}$ and $V^\star$ are the L_2-normalized eigenvectors. The [EIG] algorithm proposed by Jouanin *et al.* (2004) consists then in replacing the second step of the [BASEL] algorithm:

1. we simulate $n + m$ Gaussian independent random variables $(X^\star_1, \ldots, X^\star_m)$ and $(\varepsilon_1, \ldots, \varepsilon_n)$;
2. for the i^{th} credit, we calculate:

$$Z_i = \sum_{j=1}^{m} A^\star_{\text{map}(i),j} X^\star_j + \sqrt{1 - \rho^\star_{\text{map}(i),\text{map}(i)}}\varepsilon_i \tag{3.55}$$

3. we simulate the copula function:

$$(u_1, \ldots, u_n) = \left(\Phi\left(Z_1\right), \ldots, \Phi\left(Z_n\right)\right)$$

4. we set $\boldsymbol{\tau}_i = \mathbf{S}_i^{-1}\left(u_i\right)$.

Here is a comparison of the efficiency of the [EIG] algorithm with respect to the traditional [CHOL] algorithm:

Algorithm	Matrix dimension	Random numbers	Number of operations +	Number of operations ×
CHOL	$n \times n$	n	$n \times (n-1)$	$n \times n$
EIG	$m \times m$	$n + m$	$n \times (m+1)$	$n \times (m+1)$
	10 000 loans + 20 sectors			
CHOL	10^8	10 000	$\simeq 10^8$	10^8
EIG	400	10 020	2.1×10^5	2.1×10^5

These results explain why the [EIG] algorithm is faster than the [CHOL] algorithm[102]. We also notice that the [EIG] algorithm corresponds to the [BASEL] algorithm in the case $m = 1$ when there is only one common factor.

[100] The diagonal elements correspond to intra-sector correlations, whereas the off-diagonal elements correspond to inter-sector correlations.

[101] Jouanin *et al.* (2004) showed that if the eigenvalues of $\mathbb{C}^\star$ are positive, then $\mathbb{C} = \text{map}\left(\mathbb{C}^\star\right)$ is a correlation matrix.

[102] On average, the computational time is divided by a factor of n/m.

Let us consider Example 41. We obtain:

$$A^\star = \begin{pmatrix} -0.2633 & 0.1302 & -0.3886 & 0.2504 \\ -0.5771 & -0.1980 & -0.1090 & 0.1258 \\ -0.5536 & 0.0943 & 0.3281 & 0.2774 \\ -0.4897 & 0.0568 & -0.0335 & -0.5965 \end{pmatrix}$$

We deduce that the second step of the [EIG] algorithm is:

- if the credit belongs to the first sector, we simulate Z_i as follows:

$$Z_i = -0.263 \cdot X_1^\star - 0.130 \cdot X_2^\star + 0.389 \cdot X_3^\star + 0.250 \cdot X_4^\star + 0.837 \cdot \varepsilon_i$$

- if the credit belongs to the second sector, we simulate Z_i as follows:

$$Z_i = -0.577 \cdot X_1^\star - 0.198 \cdot X_2^\star - 0.109 \cdot X_3^\star + 0.126 \cdot X_4^\star + 0.775 \cdot \varepsilon_i$$

- if the credit belongs to the third sector, we simulate Z_i as follows:

$$Z_i = -0.554 \cdot X_1^\star + 0.094 \cdot X_2^\star + 0.328 \cdot X_3^\star + 0.277 \cdot X_4^\star + 0.707 \cdot \varepsilon_i$$

- if the credit belongs to the fourth sector, we simulate Z_i as follows:

$$Z_i = -0.490 \cdot X_1^\star + 0.057 \cdot X_2^\star - 0.034 \cdot X_3^\star - 0.597 \cdot X_4^\star + 0.632 \cdot \varepsilon_i$$

Remark 47 *The extension to the Student's t copula is straightforward, because the multivariate Student's t distribution is related to the multivariate normal distribution*[103].

3.3.4.2 The factor model

In the previous paragraph, the multivariate survival function writes:

$$\mathbf{S}(t_1, \ldots, t_n) = \mathbf{C}(\mathbf{S}_1(t_1), \ldots, \mathbf{S}_n(t_n); \mathbb{C})$$

where $\mathbf{C}$ is the Normal copula and $\mathbb{C}$ is the matrix of default correlations. In the sector approach, we note $\mathbb{C} = \text{map}(\mathbb{C}^\star)$ where map is the mapping function and $\mathbb{C}^\star$ is the matrix of intra- and inter-correlations. In this model, we characterize the default time by the relationship $\boldsymbol{\tau}_i < t \Leftrightarrow Z_i < B_i(t)$ where $Z_i = \sum_{j=1}^m A^\star_{\text{map}(i),j} X_j^\star + \sqrt{1 - \rho^\star_{\text{map}(i),\text{map}(i)}}\varepsilon_i$ and $B_i(t) = \Phi^{-1}(\text{PD}_i(t)) = \Phi^{-1}(1 - \mathbf{S}_i(t))$.

The risk factors $X_j^\star$ are not always easy to interpret. If $m = 1$, we retrieve $Z_i = \sqrt{\rho} \cdot X + \sqrt{1-\rho} \cdot \varepsilon_i$ where ρ is the uniform correlation and X is the common factor. It generally corresponds to the economic cycle. Let us consider the case $m = 2$:

$$\mathbb{C}^\star = \begin{pmatrix} \rho_1 & \rho \\ \rho & \rho_2 \end{pmatrix}$$

where ρ_1 and ρ_2 are the intra-sector correlations and ρ is the inter-sector correlation. We have:

$$Z_i = A^\star_{\text{map}(i),1} \cdot X_1^\star + A^\star_{\text{map}(i),2} \cdot X_2^\star + \sqrt{1 - \rho_{\text{map}(i)}} \cdot \varepsilon_i$$

It is better to consider the following factor decomposition:

$$Z_i = \sqrt{\rho} \cdot X + \sqrt{\rho_{\text{map}(i)} - \rho} \cdot X_{\text{map}(i)} + \sqrt{1 - \rho_{\text{map}(i)}} \cdot \varepsilon_i \tag{3.56}$$

[103] See pages 737 and 1055.

In this case, we have three factors, and not two factors: X is the common factor, whereas X_1 and X_2 are the two specific sector factors. We can extend the previous approach to a factor model with $m+1$ factors:

$$Z_i = \sqrt{\rho} \cdot X + \sqrt{\rho_{\text{map}(i)} - \rho} \cdot X_{\text{map}(i)} + \sqrt{1 - \rho_{\text{map}(i)}} \cdot \varepsilon_i \tag{3.57}$$

Equations (3.56) and (3.57) are exactly the same, except the number of factors. However, the copula function associated to the factor model described by Equation (3.57) is the copula of the sector model, when we assume that the inter-sector correlation is the same for all the sectors, meaning that the off-diagonal elements of $\mathbb{C}^\star$ are equal. In this case, we can use the previous decomposition for simulating the default times. This algorithm called [CISC] (constant inter-sector correlation) requires simulating one additional random number compared to the [EIG] algorithm. However, the number of operations is reduced[104].

Let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be two default times, whose joint survival function is $\mathbf{S}(t_1, t_2) = \mathbf{C}(\mathbf{S}_1(t_1), \mathbf{S}_2(t_2))$. We have:

$$\begin{aligned}
\mathbf{S}_1(t \mid \boldsymbol{\tau}_2 = t^\star) &= \Pr\{\boldsymbol{\tau}_1 > t \mid \boldsymbol{\tau}_2 = t^\star\} \\
&= \partial_2 \mathbf{C}(\mathbf{S}_1(t), \mathbf{S}_2(t^\star)) \\
&= \mathbf{C}_{2|1}(\mathbf{S}_1(t), \mathbf{S}_2(t^\star))
\end{aligned}$$

where $\mathbf{C}_{2|1}$ is the conditional copula function[105]. If $\mathbf{C} \neq \mathbf{C}^\perp$, the default probability of one firm changes when another firm defaults (Schmidt and Ward, 2002). This implies that the credit spread of the first firm jumps at the default time of the second firm. This phenomenon is called spread jump or jump-to-default (JTD). Sometimes it might be difficult to explain the movements of these spread jumps in terms of copula functions. The interpretation is easier when we consider a factor model. For example, we consider the Basel model. Figures 3.41 to 3.45 show the jumps of the hazard function of the S&P one-year transition matrix for corporate bonds given in Table 3.34 on page 208. We recall that the rating $\mathfrak{R}(t) = K$ corresponds to the default state and we note $\mathfrak{R}(t) = i$ the initial rating of the firm. We have seen that $\mathbf{S}_i(t) = 1 - \mathbf{e}_i^\top \exp(t\Lambda)\,\mathbf{e}_K$ where Λ is the Markov generator. The hazard function is equal to:

$$\lambda_i(t) = \frac{f_i(t)}{\mathbf{S}_i(t)} = \frac{\mathbf{e}_i^\top \Lambda \exp(t\Lambda)\,\mathbf{e}_K}{1 - \mathbf{e}_i^\top \exp(t\Lambda)\,\mathbf{e}_K}$$

We deduce that:

$$\lambda_{i_1}(t \mid \boldsymbol{\tau}_{i_2} = t^\star) = \frac{f_{i_1}(t \mid \boldsymbol{\tau}_{i_2} = t^\star)}{\mathbf{S}_{i_1}(t \mid \boldsymbol{\tau}_{i_2} = t^\star)}$$

With the Basel copula, we have:

$$\mathbf{S}_{i_1}(t \mid \boldsymbol{\tau}_{i_2} = t^\star) = \Phi\left(\frac{\Phi^{-1}(\mathbf{S}_{i_1}(t)) - \rho\Phi^{-1}(\mathbf{S}_{i_2}(t^\star))}{\sqrt{1-\rho^2}}\right)$$

and:

$$\begin{aligned}
f_{i_1}(t \mid \boldsymbol{\tau}_{i_2} = t^\star) &= \phi\left(\frac{\Phi^{-1}(\mathbf{S}_{i_1}(t)) - \rho\Phi^{-1}(\mathbf{S}_{i_2}(t^\star))}{\sqrt{1-\rho^2}}\right) \cdot \\
&\qquad \frac{f_{i_1}(t)}{\sqrt{1-\rho^2}\,\phi(\Phi^{-1}(\mathbf{S}_{i_1}(t)))}
\end{aligned}$$

[104] For the [EIG] algorithm, we have $n \times (m+1)$ operations ($+$ and $\times$), while we have $3n$ elementary operations for the [CISC] algorithm.

[105] The mathematical analysis of conditional copulas is given on page 737.

The reference to the factor model allows an easier interpretation of the jumps of the hazard rate. For example, it is obvious that the default of a CCC-rated company in ten years implies a negative jump for the well rated companies (Figure 3.45). Indeed, this indicates that the high idiosyncratic risk of the CCC-rated firm has been compensated by a good economic cycle (the common risk factor X). If the default of the CCC-rated company has occurred at an early stage, the jumps were almost zero, because we can think that the default is due to the specific risk of the company. On the contrary, if a AAA-rated company defaults, the jump would be particularly high as the default is sudden, because it is more explained by the common risk factor than by the specific risk factor (Figure 3.42). We deduce that there is a relationship between jump-to-default and default correlation.

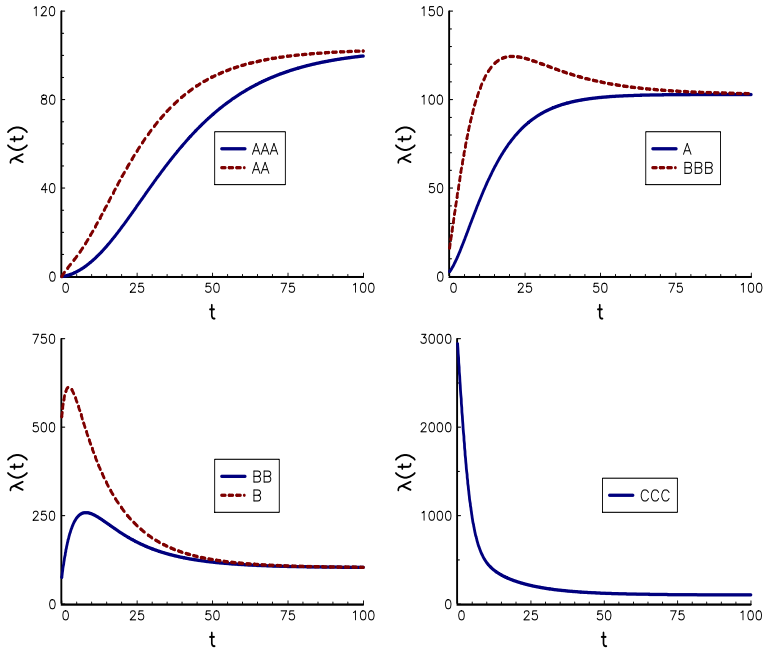


FIGURE 3.41: Hazard function $\lambda_i(t)$ (in bps)

3.3.4.3 Estimation methods

The Normal copula model with sector correlations requires the estimation of the matrix $\mathbb{C}^\star$, which is abusively called the default correlation matrix. In order to clarify this notion, we make the following distinctions:

- the '*canonical or copula correlations*' correspond to the parameter matrix of the copula function that models the dependence between the defaults;
- the '*default time correlations*' are the correlations between the default times $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$; they depend on the copula function, but also on the unidimensional survival functions;
- the '*discrete default correlations*' are the correlations between the indicator functions $(D_1(t), \ldots, D_n(t))$; they depend on the copula function, the unidimensional survival functions and the time horizon t; this is why we don't have a unique default correlation between two firms, but a term structure of default correlations;

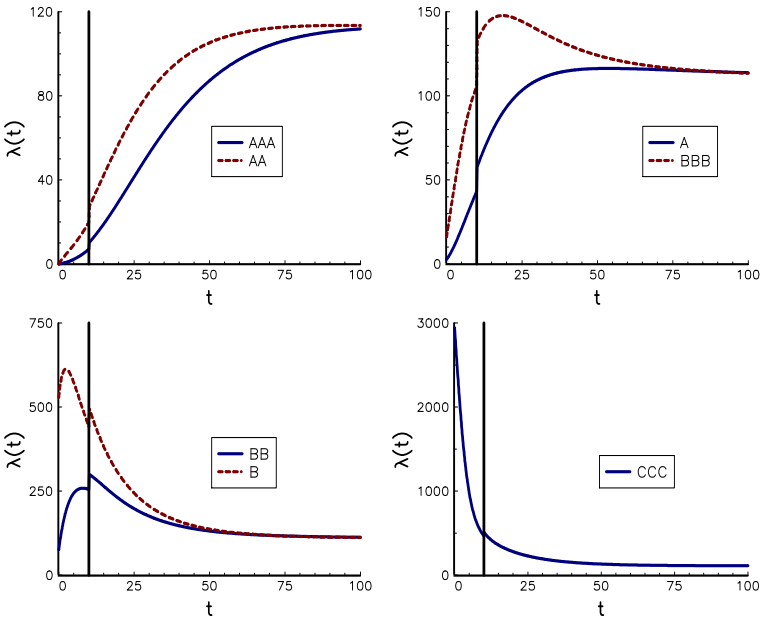


FIGURE 3.42: Hazard function $\lambda_i(t)$ (in bps) when a AAA-rated company defaults after 10 years ($\rho = 5\%$)

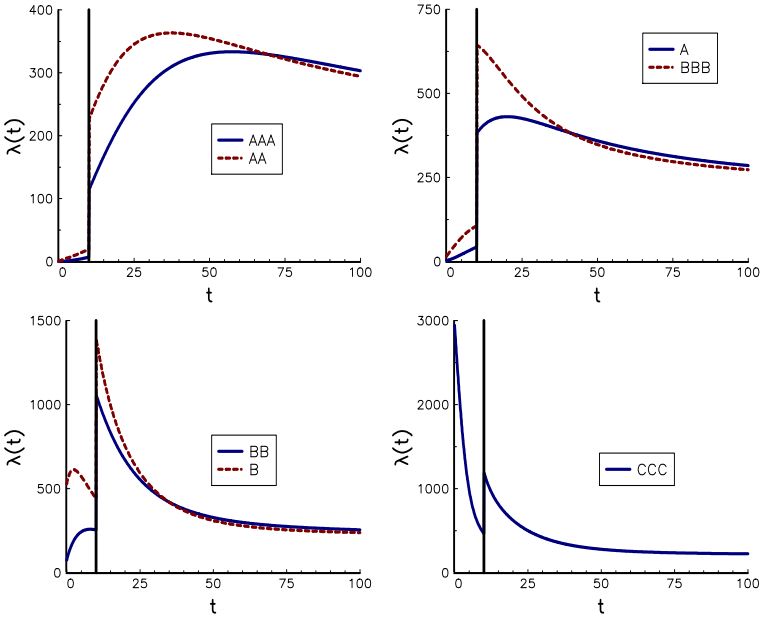


FIGURE 3.43: Hazard function $\lambda_i(t)$ (in bps) when a AAA-rated company defaults after 10 years ($\rho = 50\%$)

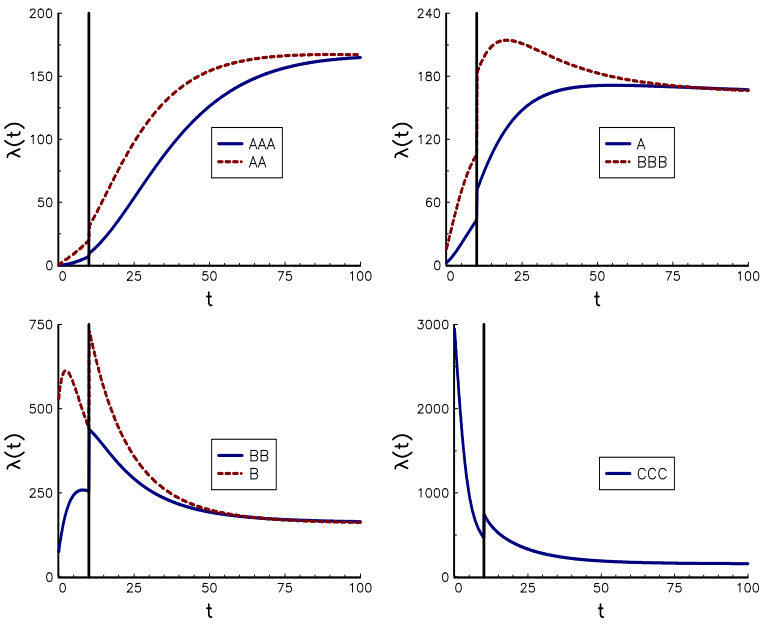


FIGURE 3.44: Hazard function $\lambda_i(t)$ (in bps) when a BB-rated company defaults after 10 years ($\rho = 50\%$)

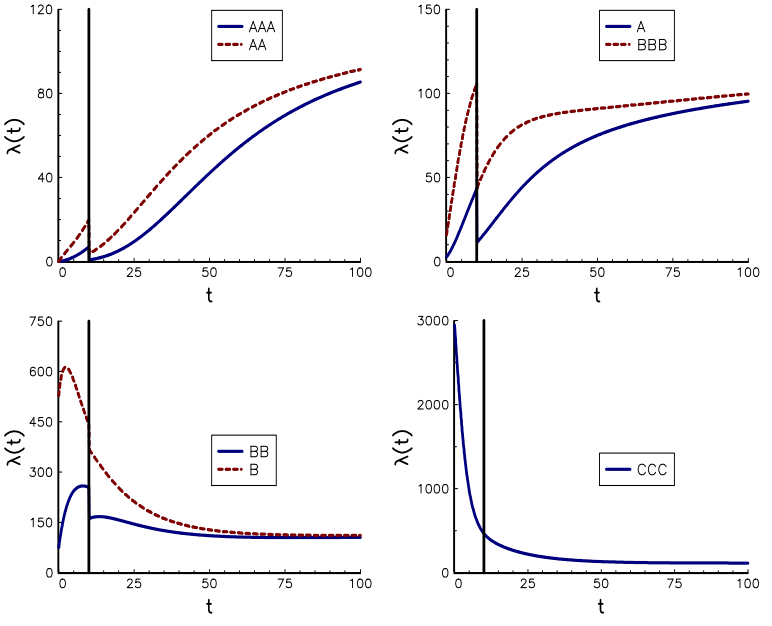


FIGURE 3.45: Hazard function $\lambda_i(t)$ (in bps) when a CCC-rated company defaults after 10 years ($\rho = 50\%$)

- the '*asset correlations*' are the correlations between the asset values in the Merton model;
- the '*equity correlations*' are the correlations between the stock returns; in a Merton-like model, they are assumed to be equal to the asset correlations.

In practice, the term '*default correlation*' is used as a generic term for these different measures.

Relationship between the different default correlations We consider two firms. Li (2000) introduces two measures of default correlation. The discrete default correlation is equal to:

$$\rho(t_1, t_2) = \frac{\mathbb{E}[D_1(t_1) D_2(t_2)] - \mathbb{E}[D_1(t_1)]\mathbb{E}[D_2(t_2)]}{\sigma(D_1(t_1))\sigma(D_2(t_2))}$$

where $D_i(t_i) = \mathbb{1}\{\boldsymbol{\tau}_i \leq t_i\}$, whereas the default (or survival) time correlation is equal to:

$$\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \frac{\mathbb{E}[\boldsymbol{\tau}_1\boldsymbol{\tau}_2] - \mathbb{E}[\boldsymbol{\tau}_1]\mathbb{E}[\boldsymbol{\tau}_2]}{\sigma(\boldsymbol{\tau}_1)\sigma(\boldsymbol{\tau}_2)}$$

These two measures give very different numerical results. Concerning the asset correlation, it is equal to:

$$\rho(Z_1, Z_2) = \frac{\mathbb{E}[Z_1 Z_2] - \mathbb{E}[Z_1]\mathbb{E}[Z_2]}{\sigma(Z_1)\sigma(Z_2)}$$

These three measures depend on the canonical correlation. Let us denote by ρ the copula parameter of the Normal copula between the two default times $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$. We have:

$$\rho(t_1, t_2) = \frac{\mathbf{C}(\mathrm{PD}_1(t_1), \mathrm{PD}_2(t_2); \rho) - \mathrm{PD}_1(t_1) \cdot \mathrm{PD}_2(t_2)}{\sqrt{\mathrm{PD}_1(t_1)(1 - \mathrm{PD}_1(t_1))} \cdot \sqrt{\mathrm{PD}_2(t_2)(1 - \mathrm{PD}_2(t_2))}}$$

and:

$$\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \frac{\mathrm{cov}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)}{\sqrt{\mathrm{var}(\boldsymbol{\tau}_1) \cdot \mathrm{var}(\boldsymbol{\tau}_1)}}$$

where $\mathrm{cov}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \int_0^\infty \int_0^\infty (\mathbf{S}(t_1, t_2) - \mathbf{S}_1(t_1)\mathbf{S}_2(t_2))\, \mathrm{d}t_1\, \mathrm{d}t_2$ and $\mathrm{var}(\boldsymbol{\tau}_i) = 2\int_0^\infty t\mathbf{S}_i(t)\, \mathrm{d}t - \left[\int_0^\infty \mathbf{S}_i(t)\, \mathrm{d}t\right]^2$. We verify that $\rho(t_1, t_2) \neq \rho$ and $\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \neq \rho$. We can also show that $\rho(t_1, t_2) < \rho$ and $\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) < \rho$ for the Normal copula. In the Basel model, we have $\rho(Z_1, Z_2) = \rho$.

We consider two exponential default times $\boldsymbol{\tau}_1 \sim \mathcal{E}(\lambda_1)$ and $\boldsymbol{\tau}_2 \sim \mathcal{E}(\lambda_2)$. In Tables 3.42, 3.43 and 3.44, we report the discrete default correlations $\rho(t_1, t_2)$ for different time horizons. We notice that $\rho(t_1, t_2)$ is much lower than 20%,, which is the copula correlation. We have also calculated $\rho(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, which is respectively equal to 17.0%, 21.5% and 18.0%. We notice that the correlations are higher for the Student's t copula than for the Normal copula[106].

Statistical inference of the default correlation In the case of a factor model, we have:

$$\tilde{Z}_{i,t} = \beta^\top \tilde{X}_{i,t} + \sqrt{1 - \|\beta\|_2^2} \cdot \tilde{\varepsilon}_{i,t}$$

where $\tilde{Z}_{i,t}$ is the standardized asset value of the i^{th} firm at time t and $\tilde{X}_{i,t}$ is the standardized vector of risk factors at time t for the i^{th} firm. We can then estimate the parameter β using OLS or GMM techniques. Let us consider the constant inter-sector correlation model:

$$Z_i = \sqrt{\rho} \cdot X + \sqrt{\rho_{\mathrm{map}(i)} - \rho} \cdot X_{\mathrm{map}(i)} + \sqrt{1 - \rho_{\mathrm{map}(i)}} \cdot \varepsilon_i$$

[106] This phenomenon is explained in the chapter dedicated to the copula functions.

TABLE 3.42: Discrete default correlation in % ($\lambda_1 = 100$ bps, $\lambda_2 = 50$ bps, Normal copula with $\rho = 20\%$)

t_1 / t_2	1	2	3	4	5	10	25	50
1	2.0	2.4	2.7	2.9	3.1	3.6	4.2	4.5
2	2.3	2.9	3.3	3.6	3.8	4.5	5.3	5.7
3	2.6	3.2	3.6	4.0	4.2	5.0	6.0	6.5
4	2.7	3.4	3.9	4.2	4.5	5.4	6.5	7.1
5	2.9	3.6	4.1	4.5	4.8	5.7	6.9	7.5
10	3.2	4.1	4.7	5.1	5.5	6.6	8.2	9.1
25	3.4	4.5	5.1	5.7	6.1	7.5	9.6	10.9
50	3.3	4.4	5.1	5.6	6.1	7.6	9.9	11.5

TABLE 3.43: Discrete default correlation in % ($\lambda_1 = 100$ bps, $\lambda_2 = 50$ bps, Student's t copula with $\rho = 20\%$ and $\nu = 4$)

t_1 / t_2	1	2	3	4	5	10	25	50
1	13.9	14.5	14.5	14.3	14.0	12.6	9.8	7.2
2	12.8	14.3	14.8	14.9	14.9	14.3	11.9	9.2
3	11.9	13.7	14.5	14.9	15.1	15.0	13.1	10.4
4	11.2	13.1	14.1	14.6	14.9	15.3	13.8	11.3
5	10.6	12.6	13.7	14.3	14.7	15.4	14.3	11.9
10	8.5	10.5	11.8	12.6	13.3	14.8	15.2	13.6
25	5.5	7.2	8.3	9.2	9.9	11.9	14.0	14.3
50	3.3	4.5	5.3	5.9	6.5	8.3	11.0	12.6

TABLE 3.44: Discrete default correlation in % ($\lambda_1 = 20\%$, $\lambda_2 = 10\%$, Normal copula with $\rho = 20\%$)

t_1 / t_2	1	2	3	4	5	10	25	50
1	8.8	10.2	10.7	11.0	11.1	10.4	6.6	2.4
2	9.4	11.0	11.8	12.1	12.3	11.9	7.9	3.1
3	9.3	11.0	11.9	12.4	12.7	12.5	8.6	3.4
4	9.0	10.8	11.7	12.2	12.6	12.6	8.9	3.7
5	8.6	10.4	11.3	11.9	12.3	12.4	9.0	3.8
10	6.3	7.8	8.7	9.3	9.7	10.3	8.1	3.7
25	1.9	2.4	2.8	3.1	3.3	3.8	3.5	1.9
50	0.2	0.3	0.3	0.3	0.4	0.5	0.5	0.3

The corresponding linear regression is:

$$\tilde{Z}_{i,t} = \beta_0 \cdot \tilde{X}_{0,t} + \beta^\top \tilde{X}_{i,t} + \sqrt{1-\rho_{\text{map}(i)}} \cdot \tilde{\varepsilon}_{i,t}$$

where $\tilde{X}_{i,t}$ is equal to $e_i \odot \tilde{X}_t$, $\tilde{X}_t$ is the set of the risk factors, which are specific to the sectors at time t and $\tilde{X}_{0,t}$ is the common risk factor. We deduce that the estimation of ρ and $\rho_1, \ldots, \rho_m$ are given by the following relationships: $\hat{\rho} = \hat{\beta}_0^2$ and $\hat{\rho}_j = \hat{\beta}_0^2 + \hat{\beta}_j^2$.

A second approach is to consider the correlation between the default rates of homogeneous cohorts[107]. This correlation converges asymptotically to the survival time correlation. Then, we have to inverse the relationship between the survival time correlation and the copula correlation for estimating the parameters of the copula function.

The third approach has been suggested by Gordy and Heitfield (2002). They consider the Basel model: $Z_i = \sqrt{\rho} \cdot X + \sqrt{1-\rho} \cdot \varepsilon_i$, where $X \sim \mathbf{H}$ and $\varepsilon_i \sim \mathcal{N}(0,1)$. The default probability conditionally to $X = x$ is equal to:

$$p_i(x; B_i, \rho) = \Phi\left(\frac{B_i - \sqrt{\rho}x}{\sqrt{1-\rho}}\right)$$

We note d_t the number of defaulted firms and n_t the total number of firms at time t. If we have a historical sample of default rates, we can estimate ρ using the method of maximum likelihood. Let $\boldsymbol{\ell}_t(\theta)$ be the log-likelihood of the observation t. If we assume that there is only one risk class $\mathcal{C}$ ($B_i = B$), the conditional number of defaults $\mathcal{D}$ is a binomial random variable:

$$\Pr\{\mathcal{D} = d_t \mid X = x\} = \binom{n_t}{d_t} p(x; B, \rho)^{d_t} (1 - p(x; B, \rho))^{n_t - d_t}$$

We deduce that:

$$\begin{aligned}
\boldsymbol{\ell}_t(\theta) &= \ln \int \Pr\{\mathcal{D} = d_t \mid X = x\} \, \mathrm{d}\mathbf{H}(x) \\
&= \ln \int \binom{n_t}{d_t} p(x; B, \rho)^{d_t} (1 - p(x; B, \rho))^{n_t - d_t} \, \mathrm{d}\mathbf{H}(x)
\end{aligned}$$

Generally, we consider a one-year time horizon for calculating default rates. Moreover, if we assume that the common factor X is Gaussian, we deduce that $B = \Phi^{-1}(\text{PD})$ where PD is the one-year default probability for the risk class $\mathcal{C}$. It follows that:

$$\boldsymbol{\ell}_t(\theta) = \ln \int \binom{n_t}{d_t} p\left(x; \Phi^{-1}(\text{PD}), \rho\right)^{d_t} \left(1 - p\left(x; \Phi^{-1}(\text{PD}), \rho\right)\right)^{n_t - d_t} \mathrm{d}\Phi(x)$$

Therefore, we can estimate the parameter ρ. If there are several risk classes, we can assume that:

$$\boldsymbol{\ell}_t(\theta) = \ln \int \binom{n_t}{d_t} p(x; B, \rho)^{d_t} (1 - p(x; B, \rho))^{n_t - d_t} \, \mathrm{d}\Phi(x)$$

In this case, we have two parameters to estimate: the copula correlation ρ and the implied default barrier B.

The underlying idea of this approach is that the distribution of the default rate depends on the default probability and the copula correlation. More specifically, the mean of the default rate of a risk class $\mathcal{C}$ is equal to the default probability of $\mathcal{C}$ whereas the volatility of the default rate is related to the default correlation. We introduce the notation:

$$f_t = \frac{d_t}{n_t}$$

[107] Each cohort corresponds to a risk class.

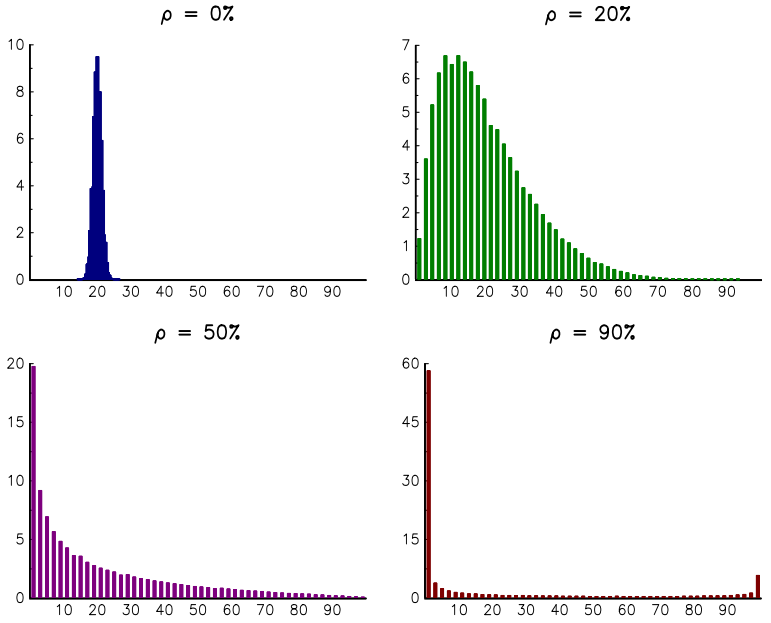


FIGURE 3.46: Distribution of the default rate (in %)

where f_t is the default rate at time t. We assume that the one-year default probability of $\mathcal{C}$ is equal to 20%. In Figure 3.46, we report the distribution of the one-year default rate for different values of ρ when the number of firms n_t is equal to 1 000. We also report some statistics (mean, standard deviation and quantile functions) in Table 3.45. By definition, the four probability distributions have the same mean, which is equal to 20%, but their standard deviations are different. If $\rho = 0\%$, $\sigma(f_t)$ is equal to 1.3% while $\sigma(f_t) = 33.2\%$ in the case $\rho = 90\%$.

TABLE 3.45: Statistics of the default rate (in %))

ρ	$\mu(f_t)$	$\sigma(f_t)$	$Q_\alpha(f_t)$						
			1%	10%	25%	50%	75%	90%	99%
0%	20.0	1.3	17.1	18.4	19.1	20.0	20.8	21.6	23.0
20%	20.0	13.0	1.7	5.6	10.0	17.4	27.3	38.3	59.0
50%	20.0	21.7	0.0	0.6	3.1	11.7	30.3	53.8	87.3
90%	20.0	33.2	0.0	0.0	0.0	0.4	26.3	88.2	100.0

Example 42 *We consider a risk class $\mathcal{C}$, whose probability of default is equal to* 200 *bps. Over the last 20 years, we have observed the following annual number of defaults:* 3, 1, 14, 0, 33, 3, 53, 1, 4, 0, 1, 8, 7, 3, 5, 5, 0, 49, 0 *and* 7. *We assume that the number of firms is equal to* 500 *every year.*

If we estimate the Basel model with the method of maximum likelihood by assuming that $B = \Phi^{-1}(\text{PD})$, we obtain $\hat{\rho} = 28.93\%$. If we estimate both the default correlation and the default barrier, we have $\hat{\rho} = 28.58\%$ and $\hat{B} = -2.063$, which is equivalent to a default probability of 195 bps. It is better to estimate the barrier if we don't trust the default probability of the risk class because the estimation can be biased. For instance, if

we assume that PD = 100 bps, we obtain $\hat{\rho} = 21.82\%$, which is relatively lower than the previous estimate.

The previous estimation method has been generalized by Demey *et al.* (2004) to the CISC model with several intra-sector correlations, but a unique inter-sector correlation. In Table 3.46, we report their results for the period between 1981 and 2002. We notice that the default correlations are relatively low between 7% and 36%. The largest correlations are observed for the sectors of energy, finance, real estate, telecom and utilities. We also notice some significant differences between the Basel model and the CISC model.

TABLE 3.46: Estimation of canonical default correlations

Sector	CISC model	Basel model
Aerospace/Automobile	11.2%	11.6%
Consumer/Service sector	8.7%	7.5%
Energy/Natural resources	21.3%	11.5%
Financial institutions	15.7%	12.2%
Forest/Building products	6.8%	14.5%
Health	8.3%	9.2%
High technology	6.8%	4.7%
Insurance	12.2%	7.6%
Leisure time/Media	7.0%	7.0%
Real estate	35.9%	27.7%
Telecom	27.1%	34.3%
Transportation	6.8%	8.3%
Utilities	18.3%	21.2%
Inter-sector	6.8%	✓

Source: Demey *et al.* (2004).

Remark 48 *There are very few publications on the default correlations. Moreover, they generally concern the one-year discrete default correlations* $\rho(1,1)$*, not the copula correlation. For example, Nagpal and Bahar (2001) estimate* $\rho(t_1,t_2)$ *for US corporates and the period 1981-1999. They distinguish the different sectors, three time horizons (1Y, 5Y and 7Y) and IG/HY credit ratings. Even if the range goes from* -5.35% *to* 39.35%*, they obtain a very low correlation on average. However, these results should be taken with caution, because we know that the default correlation has increased since the 2008 Global Financial Crisis (Christoffersen et al., 2017).*

3.3.4.4 Dependence and credit basket derivatives

Interpretation and pitfalls of the Basel copula The Basel copula is the basic model for pricing CDO tranches, just as the Black-Scholes model is for options. We define the implied correlation as the parameter ρ that gives the market spread of the CDO tranche. In some sense, the implied correlation for CDOs is the equivalent of the implied volatility for options. Since the implied correlation depends on attachment and detachment points of the CDO tranche, we don't have a single value, but a curve which is not flat. Therefore, we observe a correlation smile or skew, meaning that the correlation is not constant.

In order to understand this phenomenon, we come back to the economic interpretation of the Basel model. In Figure 3.47, we report the mapping between the economic cycle and the common risk factor X. In this case, negative values of X correspond to bad economic

times whereas positive values of X correspond to good economic times. We notice that the factor model does not encompass the dynamics of the economic cycle. The Basel model is typically a through-the-cycle approach, and not a point-in-time approach, meaning that the time horizon is the long-run (typically an economic cycle of 7 years).

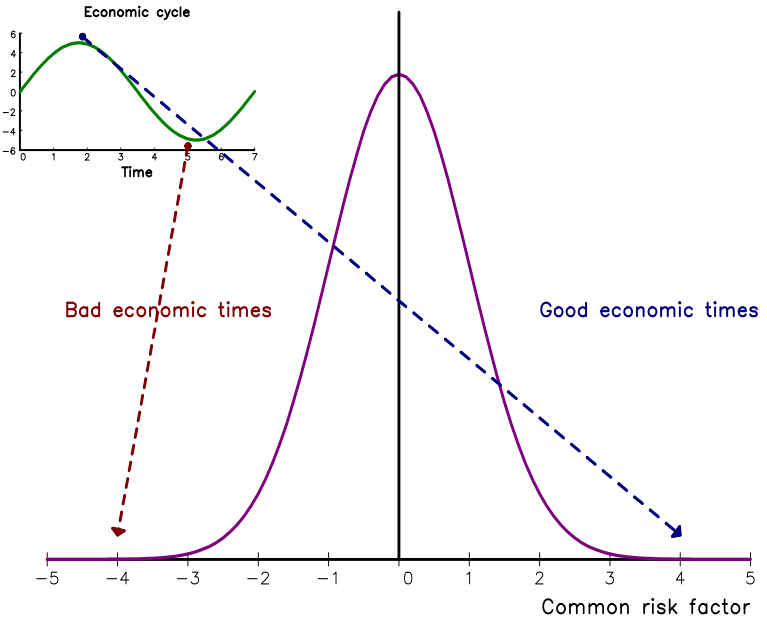


FIGURE 3.47: Economic interpretation of the common factor X

We recall that the loss function is $L = \sum_{i=1}^{n} \text{EAD}_i \cdot \text{LGD}_i \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq T_i\}$. Let A and D be the attachment and detachment points of the tranche. We have:

$$\mathbb{E}[L \mid A \leq L < D] = \mathbb{E}_X[L(X) \mid A \leq L < D]$$

where $L(X)$ is the conditional loss with respect to the common factor X. With this model, the pricing of a CDO tranche uses all the economic scenarios, which are equally weighted. In practice, we know that market participants are more sensitive to bad economic times and have a shorter time horizon than the duration of an economic cycle. From a mathematical point of view, this implies that the factor component $\sqrt{\rho}X$ is certainly not Gaussian and symmetric about 0. Two directions have then been investigated in order to introduce skewness in credit risk modeling. The first approach assumes that the copula correlation ρ is not constant but stochastic, while the second approach states that the copula correlation is a function of the common factor X. These two approaches are two visions of the link between default correlations and the economic cycle.

Stochastic correlation model We consider an extension of the Basel model:

$$Z_i = \sqrt{R_i}X + \sqrt{1 - R_i}\,\varepsilon_i$$

where $R_i \in [0, 1]$ is a random variable that represents the stochastic correlation (Andersen and Sidenius, 2005). We notice that the conditional process $Z_i \mid R_i = \rho$ remains Gaussian, whereas the conditional probability of default becomes:

$$p_i(X) = \int_0^1 \Phi\left(\frac{B_i - \sqrt{\rho}X}{\sqrt{1-\rho}}\right) \mathrm{d}\mathbf{G}(\rho)$$

where $\mathbf{G}$ is the probability distribution of R_i. Burtschell *et al.* (2007) propose to model the stochastic correlation R_i as a binary random variable:

$$R_i = (1 - Y_i)\sqrt{\rho_1} + Y_i\sqrt{\rho_2}$$

where Y_i is a Bernoulli random variable $\mathcal{B}(p)$. For example, if $p = 5\%$, $\rho_1 = 0\%$ and $\rho_2 = 100\%$, the defaults are uncorrelated most of the time and perfectly correlated in 5% of cases. The copula of default times is then a mixture of the copula functions $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$ as shown in Figure 3.48. From an economic point of view, we obtain a two-regime model.

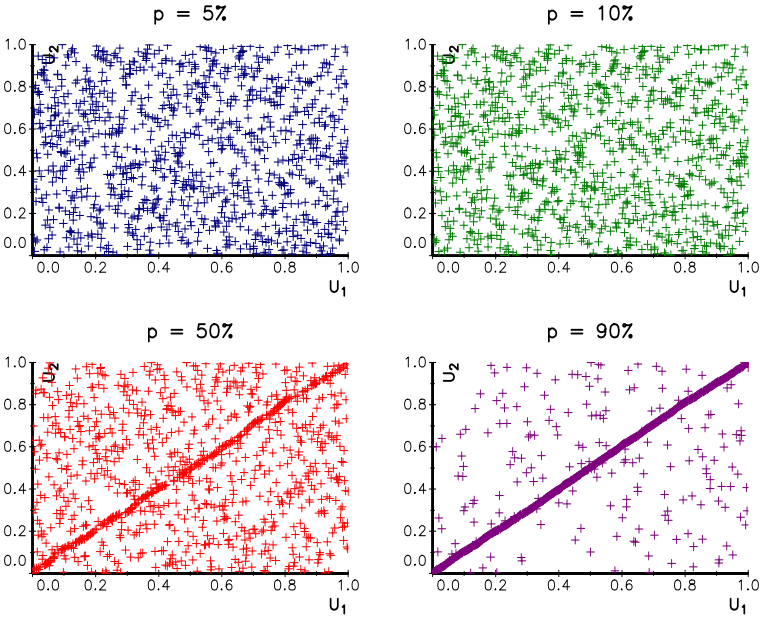


FIGURE 3.48: Dependogram of default times in the stochastic correlation model

Local correlation model In this model, we have:

$$Z_i = \beta(X)\, X + \sqrt{1 - \|\beta(X)\|_2^2}\,\varepsilon_i$$

where the factor loading $\beta(X)$ is a function of the factor X, meaning that $\beta(X)$ depends on the position in the economic cycle. In Figure 3.49, we consider two functions: $\beta_0(X)$ is constant (Basel model) and $\beta_1(X)$ decreases with the common factor. In this last case, the factor loading is high in bad economic times, meaning that the default correlation $\rho = \beta^2(X)$ is larger at the bottom of the economic cycle than at its top. This implies that the latent variable Z_i is not Gaussian and exhibits a skewness and an excess kurtosis. We verify this property on the normalized probability density function of the factor component $\beta(X)\, X$ (bottom/right panel in Figure 3.49). This specification has an impact of the joint distribution of defaults. For example, we report the empirical copula of default times in Figure 3.50 when the factor loading is $\beta_1(X)$. We notice that this copula function is not symmetric and the joint dependence of defaults is very high in bad economic times when the value of X is low. When $\beta(X)$ is a decreasing function of X, we observe a correlation skew. It is equivalent to change the probability measure in order to penalize the bad states of the economic cycle or to introduce a risk premium due to the misalignment between the time horizon of investors and the duration of the economic cycle.

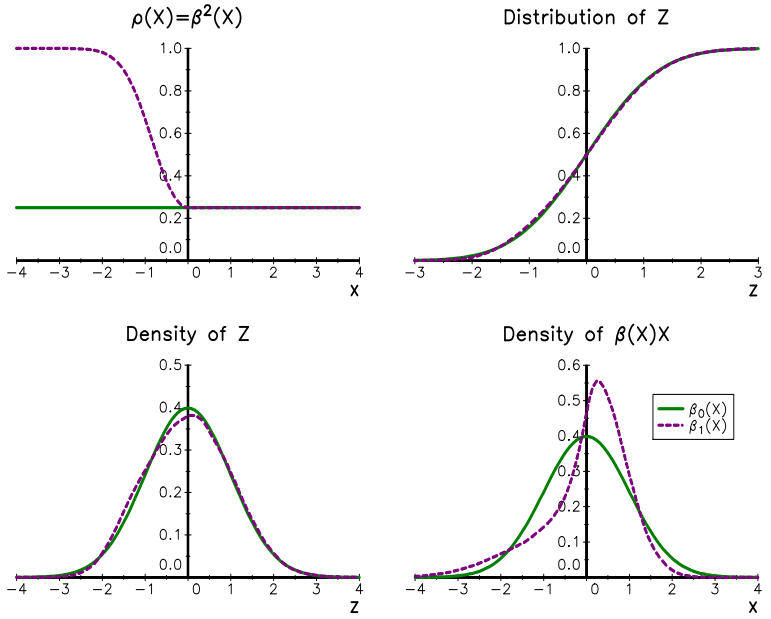


FIGURE 3.49: Distribution of the latent variable Z in the local correlation model

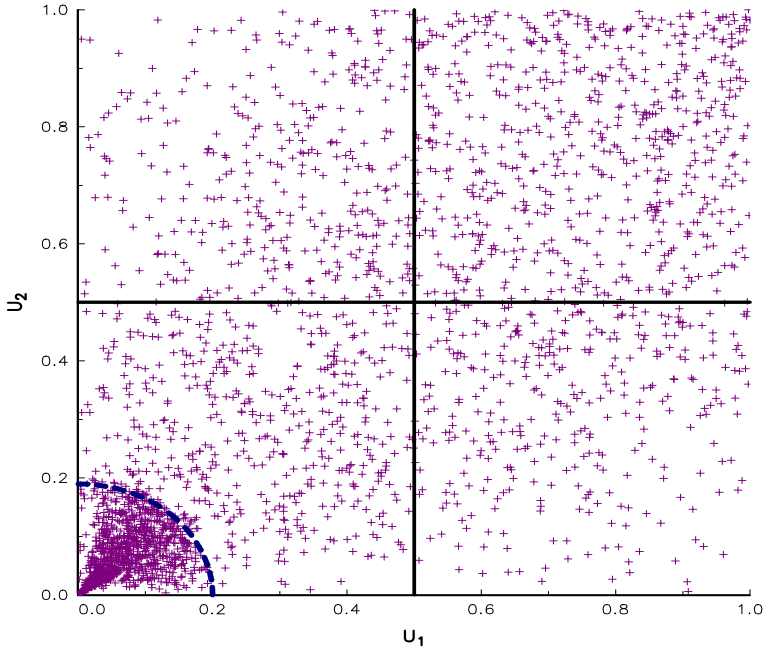


FIGURE 3.50: Dependogram of default times in the local correlation model

To implement this model, we consider the normalization $Z_i^\star = \sigma_Z^{-1}(Z_i - m_Z)$ where:

$$m_Z = \mathbb{E}[Z_i] = \int_{-\infty}^{+\infty} \beta(x)\, x\phi(x)\,\mathrm{d}x$$

and:

$$\sigma_Z^2 = \mathrm{var}(Z_i) = \int_{-\infty}^{+\infty} \left(1 - \beta^2(x) + \beta^2(x)\, x^2\right) \phi(x)\,\mathrm{d}x - m_Z^2$$

We notice that the probability distribution of the latent variable $Z_i^\star$ is equal to:

$$\begin{aligned} \mathbf{F}^\star(z) &= \Pr\{Z_i^\star \le z\} \\ &= \int_{-\infty}^{+\infty} \Phi\left(\frac{m_z + \sigma_Z z - \beta(x)\, x}{\sqrt{1 - \|\beta(x)\|_2^2}}\right) \phi(x)\,\mathrm{d}x \end{aligned}$$

To simulate correlated defaults[108], we use the inversion method such that $U_i = \mathbf{F}^\star(Z_i)$.

We consider the following parametrization:

$$\beta(x) = \begin{cases} 1 - \left(1 - \sqrt{\rho}\right) e^{-\frac{1}{2}\alpha x^2} & \text{if } x < 0 \\ \sqrt{\rho} & \text{if } x \ge 0 \end{cases}$$

The function $\beta(x)$ depends on two parameters ρ and α. The local correlation $\rho(x) = \beta^2(x)$ is given in Figure 3.51. The parameter ρ represents the default correlation when the economic cycle is good or the common factor X is positive. We also notice that the local correlation $\rho(x)$ tends to 1 when x tends to $-\infty$. This implies an absolute contagion of the default times when the economic situation is dramatic. The parameter α is then a measure of the contagion intensity when the economic cycle is unfavorable. Figure 3.52 shows the base correlation[109] which are generated by this model[110]. We observe that these concave skews are coherent with respect to those observed in the market.

In Figure 3.53, we report the base correlation of the 5Y European iTraxx index at the date of 14 June 2005. The estimation of the local correlation model gives $\rho = 0.5\%$ and $\alpha = 60\%$. We notice that the calibrated model fits well the correlation skew of the market. Moreover, the calibrated model implies an asymmetric distribution and a left fat tail of the factor component (top/right panel in Figure 3.54) and an implied economic cycle, which is more flattened than the economic cycle derived from a Gaussian distribution. In particular, we observe small differences within good economic times and large differences within bad economic times. If we consider the copula function, we find that defaults are generally weakly correlated except during deep economic crisis. Let us consider the ordinal sum of the two copula functions $\mathbf{C}^\perp$ and $\mathbf{C}^+$. This copula is represented in Figure 3.55. The 10% worst economic scenarios correspond to the perfect dependence (copula $\mathbf{C}^+$) whereas the remaining 90% economic scenarios correspond to the zero-correlation situation (copula $\mathbf{C}^\perp$). We notice that this copula function fits very well the correlation skew. We conclude that market participants underestimate default correlations in good times and overestimate default correlations in bad times.

[108] We calculate m_Z, σ_Z and $\mathbf{F}^\star(z)$. For $\mathbf{F}^\star(z)$, we consider a meshgrid (z_k). When $z \in (z_k, z_{k+1})$, we use the linear or the Gaussian interpolation.

[109] The base correlation is the implied correlation of an equity tranche, where the attachment point is equal to 0 and the detachment point is equal to the strike.

[110] We consider a CDO with a five-year maturity. The coupons are paid every quarter. The portfolio of underlying assets is homogenous with a spread of 100 bps and a recovery rate of 40%. The pricing is done with the method of Monte Carlo and one million simulations

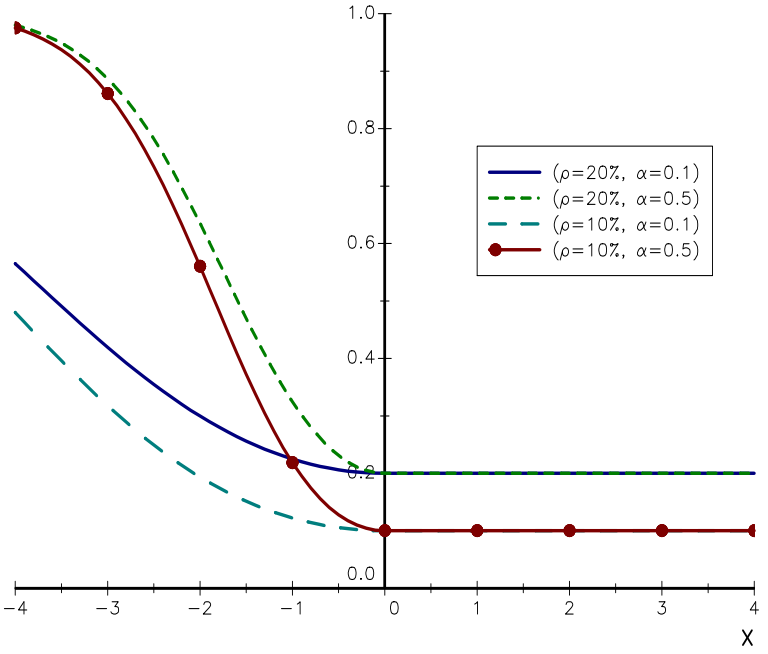


FIGURE 3.51: Local correlation $\rho(x) = \beta^2(x)$

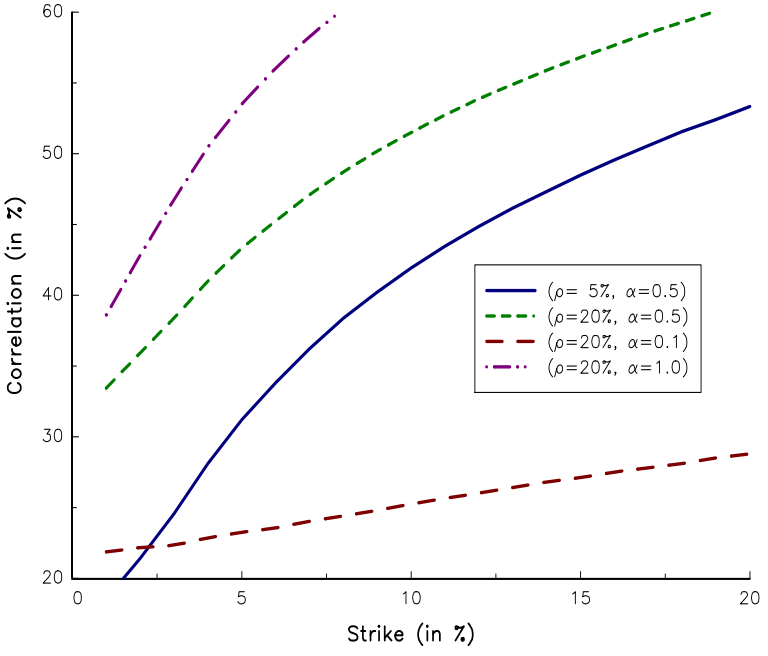


FIGURE 3.52: Correlation skew generated by the local correlation model

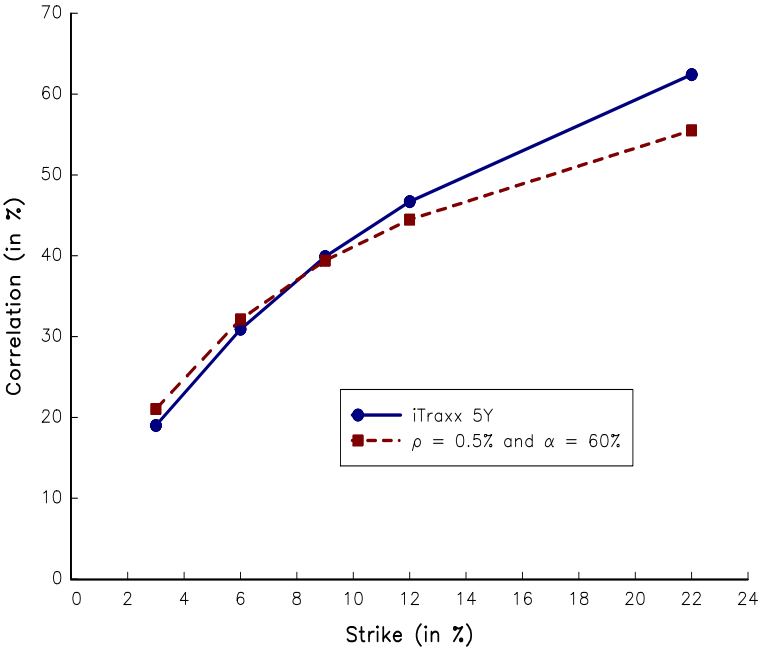


FIGURE 3.53: Calibration of the correlation skew (local correlation model)

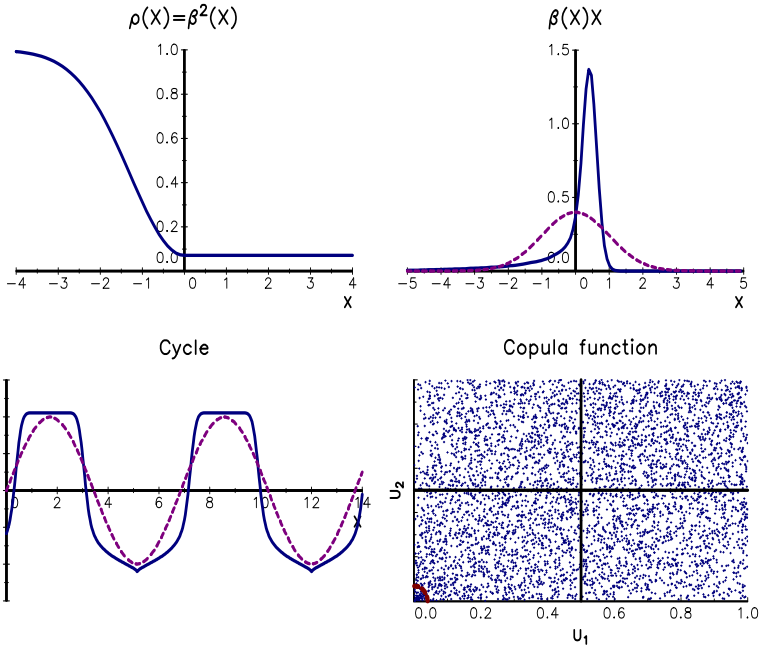


FIGURE 3.54: Implied local correlation model

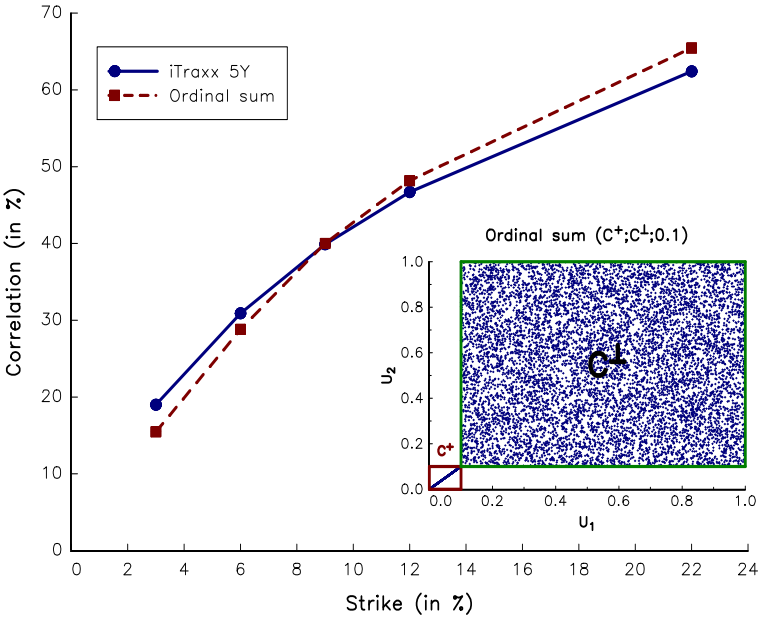


FIGURE 3.55: Calibration of the correlation skew (ordinal sum of $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$)

3.3.5 Granularity and concentration

The risk contribution of the Basel model has been obtained under the assumption that the portfolio is infinitely fine-grained. In this case, the common risk factor X largely dominates the specific risk factors ε_i. When the portfolio is concentrated in a few number of credits, the risk contribution formula, which has been derived on page 173, is not valid. In this case, the Basel regulation implies to calculate an additional capital. In the second consultative paper on the Basel II Accord (BCBS, 2001a), the Basel Committee suggested to complement the IRB-based capital by a '*granularity adjustment*' that captures the risk of concentration. Finally, the Basel Committee has abandoned the idea to calculate the additional capital in the first pillar. In fact, this granularity adjustment is today treated in the second pillar, and falls under the internal capital adequacy assessment process (ICAAP).

3.3.5.1 Difference between fine-grained and concentrated portfolios

Definition of the granularity adjustment We recall that the portfolio loss is given by:

$$L = \sum_{i=1}^{n} \text{EAD}_i \cdot \text{LGD}_i \cdot \mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\} \tag{3.58}$$

Under the assumption that the portfolio is infinitely fine-grained (IFG), we have shown that the one-year value-at-risk is given by[111]:

$$\text{VaR}_{\alpha}\left(w_{\text{IFG}}\right) = \sum_{i=1}^{n} \text{EAD}_i \cdot \mathbb{E}\left[\text{LGD}_i\right] \cdot \Phi\left(\frac{\Phi^{-1}\left(\text{PD}_i\right) + \sqrt{\rho}\Phi^{-1}\left(\text{PD}_i\right)}{\sqrt{1-\rho}}\right) \tag{3.59}$$

However, this assumption does not always hold, and the portfolio w cannot be fine-grained and present some concentration issues. In this case, the one-year value-at-risk is equal to

[111] Without any loss of generality, we assume that $T_i = 1$ in the sequel.

the quantile α of the loss distribution:

$$\text{VaR}_\alpha(w) = \mathbf{F}_L^{-1}(\alpha)$$

The granularity adjustment GA is the difference between the two risk measures. In the case of the VaR and UL credit risk measures, we obtain:

$$\text{GA} = \text{VaR}_\alpha(w) - \text{VaR}_\alpha(w_{\text{IFG}})$$

In most cases, we expect that the granularity adjustment is positive, meaning that the IRB-based capital underestimates the credit risk of the portfolio.

The case of a perfectly concentrated portfolio Let us consider a portfolio that is composed of one credit. We have:

$$L = \text{EAD} \cdot \text{LGD} \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\}$$

Let $\mathbf{G}$ be the distribution function of the loss given default. It follows that:

$$\mathbf{F}_L(\ell) = \Pr\{\text{EAD} \cdot \text{LGD} \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\} \leq \ell\}$$

Since we have $\ell = 0 \Leftrightarrow \boldsymbol{\tau} > T$, we deduce that $\mathbf{F}_L(0) = \Pr\{\boldsymbol{\tau} > T\} = 1 - \text{PD}$. If $\ell \neq 0$, this implies that the default has occurred and we have:

$$\begin{aligned}\mathbf{F}_L(\ell) &= \mathbf{F}_L(0) + \Pr\{\text{EAD} \cdot \text{LGD} \leq \ell \mid \boldsymbol{\tau} \leq T\} \\ &= (1 - \text{PD}) + \text{PD} \cdot \mathbf{G}\left(\frac{\ell}{\text{EAD}}\right)\end{aligned}$$

The value-at-risk of this portfolio is then equal to:

$$\text{VaR}_\alpha(w) = \begin{cases} \text{EAD} \cdot \mathbf{G}^{-1}\left(\dfrac{\alpha + \text{PD} - 1}{\text{PD}}\right) & \text{if } \alpha \geq 1 - \text{PD} \\ 0 & \text{otherwise} \end{cases}$$

In figure 3.56, we consider an illustration when the exposure at default is equal to one. The first panel compares the value-at-risk $\text{VaR}_\alpha(w)$ when $\text{LGD} \sim \mathcal{U}[0,1]$ and $\text{LGD} = 50\%$. Except for low default probabilities, $\text{VaR}_\alpha(w)$ is larger when the loss given default is stochastic than when the loss given default is set to the mean $\mathbb{E}[\text{LGD}]$. The next panels also shows that the IRB value-at-risk $\text{VaR}_\alpha(w_{\text{IFG}})$ underestimates the true value-at-risk $\text{VaR}_\alpha(w)$ when PD is high. We conclude that the granularity adjustment depends on two main factors:

- the discrepancy between LGD and its expectation $\mathbb{E}[\text{LGD}]$;
- the specific risk that can increase or decrease[112] the credit risk of the portfolio.

The diversification effect and the default correlation We also notice that the granularity adjustment is equal to zero when the default correlation tends to one:

$$\lim_{\rho \to 1} \text{VaR}_\alpha(w) = \text{VaR}_\alpha(w_{\text{IFG}})$$

[112] For instance, the true value-at-risk can be lower than the sum of IRB contributions for well-rated portfolios.

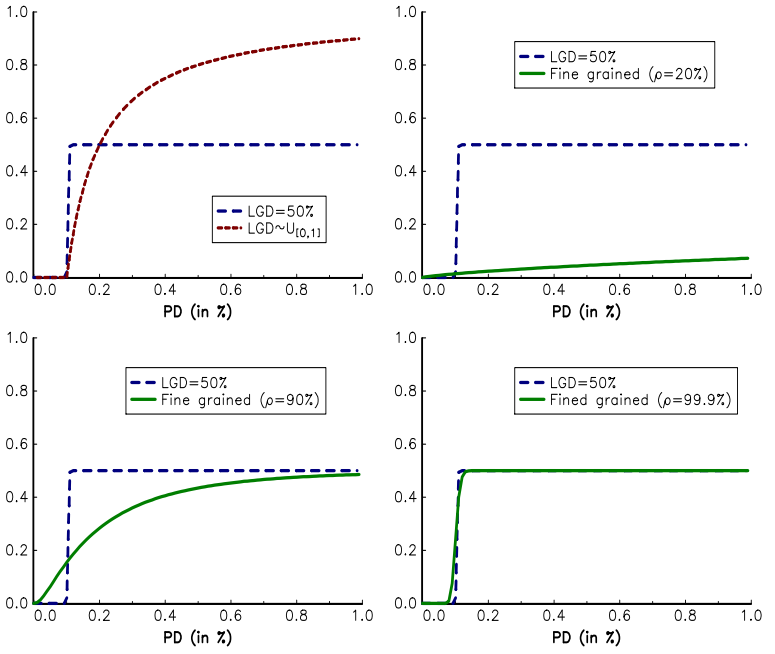


FIGURE 3.56: Comparison between the 99.9% value-at-risk of a loan and its risk contribution in an IFG portfolio

Indeed, when $\rho = 1$, there is no diversification effect. To illustrate this property, we report the loss distribution of an infinitely fine-grained portfolio[113] in Figure 3.57. When the correlation is equal to zero, the conditional expected loss does not depend on X and we have:

$$L = \mathbb{E}[L \mid X] = \text{EAD} \cdot \text{LGD} \cdot \text{PD}$$

When the correlation is different from zero, we have:

$$\begin{cases} \mathbb{E}[L \mid X] > \mathbb{E}[L] & \text{for low values of } X \\ \mathbb{E}[L \mid X] < \mathbb{E}[L] & \text{for high values of } X \end{cases}$$

Since the value-at-risk considers a bad economic scenario, it is normal that the value-at-risk increases with respect to ρ because $\mathbb{E}[L \mid X]$ is an increasing function of ρ in bad economic times.

In Figure 3.58, we compare the normalized loss distribution[114] of non fined-grained, but homogenous portfolios. We notice that the loss distribution of the portfolio converges rapidly to the loss distribution of the IFG portfolio. It suffices that the number of credits is larger than 50. However, this result assumes that the portfolio is homogenous. In the case of non-homogenous portfolio, it is extremely difficult to define a rule to know if the portfolio is fine-grained or not.

[113] This is a homogeneous portfolio of 50 credits with the following characteristics: EAD = 1, $\mathbb{E}[\text{LGD}] = 50\%$ and PD = 10%.

[114] This is the loss of the portfolio divided by the number n of credits in the portfolio.

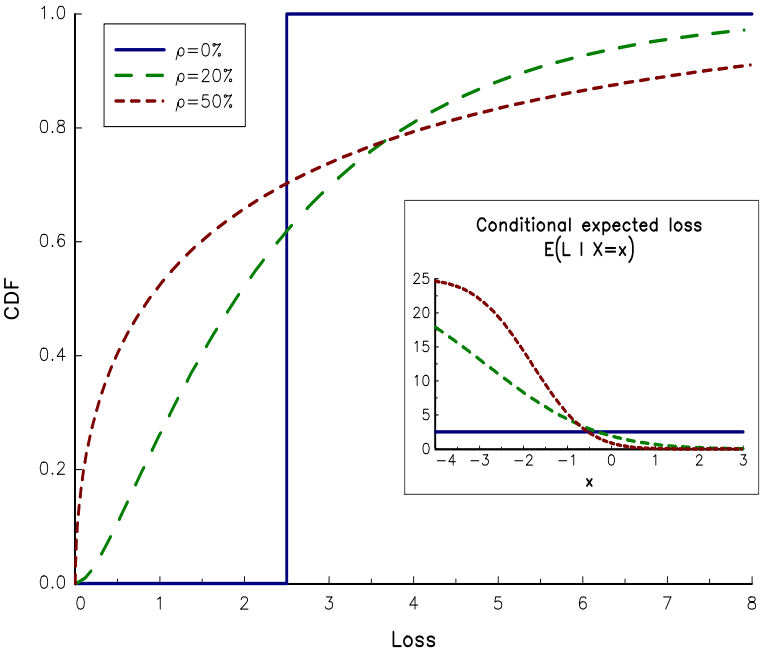


FIGURE 3.57: Loss distribution of an IFG portfolio

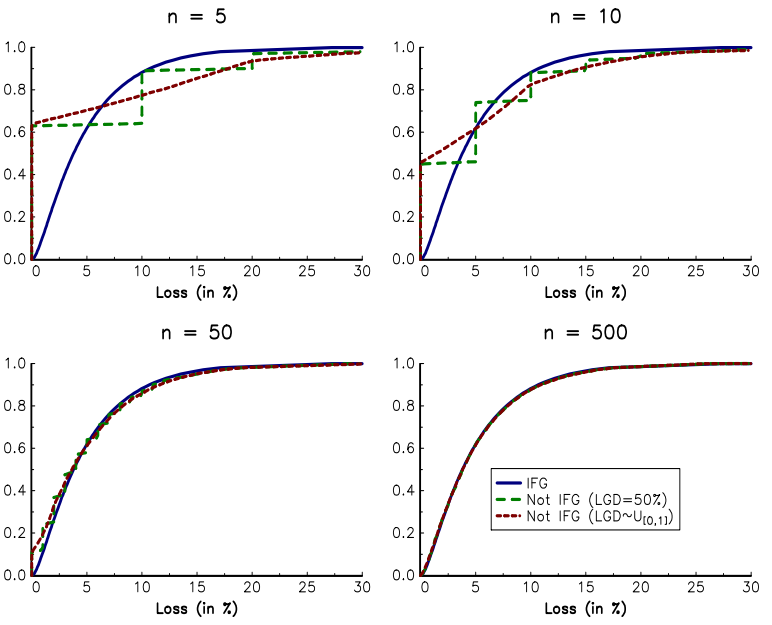


FIGURE 3.58: Comparison of the loss distribution of non-IFG and IFG portfolios

3.3.5.2 Granularity adjustment

Monte Carlo approach The first approach to compute the granularity adjustment is to estimate the quantile $\hat{\mathbf{F}}_L^{-1}(\alpha)$ of the portfolio loss using the Monte Carlo method. In Table 3.47, we have reported the (relative) granularity adjustment, which is defined as:

$$\mathrm{GA}^\star = \frac{\hat{\mathbf{F}}_L^{-1}(\alpha) - \mathrm{VaR}_\alpha(w_{\mathrm{IFG}})}{\mathrm{VaR}_\alpha(w_{\mathrm{IFG}})}$$

for different homogenous credit portfolios when EAD = 1. We consider different values of the default probability PD (1% and 10%), the size n of the portfolio (50, 100 and 500) and the confidence level α of the value-at-risk (90%, 99% and 99.9%). For the loss given default, we consider two cases: LGD = 50% and LGD $\sim \mathcal{U}[0,1]$. For each set of parameters, we use 10 million simulations for estimating the quantile $\hat{\mathbf{F}}_L^{-1}(\alpha)$ and the same seed for the random number generator[115] in order to compare the results. For example, when $n = 50$, PD = 10%, $\rho = 10\%$, LGD $\sim \mathcal{U}[0,1]$ and $\alpha = 90\%$, we obtain $\mathrm{GA}^\star = 13.8\%$. This means that the capital charge is underestimated by 13.8% if we consider the IRB formula. We notice that the granularity adjustment is positive in the different cases we have tested. We verify that it decreases with respect to the portfolio size. However, it is difficult to draw other conclusions. For instance, it is not necessarily an increasing function of the confidence level.

TABLE 3.47: Granularity adjustment $\mathrm{GA}^\star$ (in %)

n		50	100	500	50	100	500
Parameters	α	LGD $\sim \mathcal{U}_{[0,1]}$			LGD = 50%		
PD = 10% ρ = 10%	90%	13.8	7.4	1.6	12.5	6.8	1.2
	99%	19.3	10.0	2.1	13.3	6.2	1.2
	99.9%	21.5	10.9	2.3	12.2	6.9	1.6
PD = 10% ρ = 20%	90%	8.1	4.2	0.9	2.7	2.7	0.9
	99%	10.3	5.3	1.1	6.7	4.1	0.6
	99.9%	11.3	5.6	1.2	6.5	2.8	0.6
PD = 1% ρ = 20%	90%	43.7	23.5	5.0	60.1	20.1	4.0
	99%	36.7	18.8	3.9	32.9	19.6	3.7
	99.9%	30.2	15.3	3.1	23.7	9.9	1.7

Analytical approach Let w be a credit portfolio. We have the following identity:

$$\mathrm{VaR}_\alpha(w) = \mathrm{VaR}_\alpha(w_{\mathrm{IFG}}) + \underbrace{\mathrm{VaR}_\alpha(w) - \mathrm{VaR}_\alpha(w_{\mathrm{IFG}})}_{\text{Granularity adjustment GA}} \tag{3.60}$$

The granularity adjustment is then the capital we have to add to the IRB value-at-risk in order to obtain the true value-at-risk (Wilde, 2001b; Gordy, 2003). Since we have seen that $\mathrm{VaR}_\alpha(w_{\mathrm{IFG}})$ is the conditional expected loss when the risk factor X corresponds to the quantile $1-\alpha$, we obtain:

$$\begin{aligned}\mathrm{VaR}_\alpha(w) &= \mathrm{VaR}_\alpha(w_{\mathrm{IFG}}) + \mathrm{GA} \\ &= \mathbb{E}[L \mid X = x_\alpha] + \mathrm{GA}\end{aligned}$$

[115] See Chapter 13 on page 787.

where $x_\alpha = \mathbf{H}^{-1}(1-\alpha)$ and $\mathbf{H}(x)$ is the cumulative distribution function of X. In order to derive the expression of the granularity adjustment, we rewrite Equation (3.60) in terms of portfolio loss:

$$L = \mathbb{E}[L \mid X] + (L - \mathbb{E}[L \mid X])$$

Since we have $\mathrm{VaR}_\alpha(w) = \mathbf{F}^{-1}(\alpha)$ where $\mathbf{F}(\ell)$ is the loss distribution, we deduce that:

$$\begin{aligned}\mathrm{VaR}_\alpha(w) &= \mathrm{VaR}_\alpha(L) \\ &= \mathrm{VaR}_\alpha(\mathbb{E}[L \mid X] + \eta(L - \mathbb{E}[L \mid X]))\big|_{\eta=1}\end{aligned}$$

Emmer and Tasche (2005) consider the second-order Taylor expansion of the value-at-risk:

$$\begin{aligned}\mathrm{VaR}_\alpha(w) \approx\ & \mathrm{VaR}_\alpha(\mathbb{E}[L \mid X]) + \left|\frac{\partial\ \mathrm{VaR}_\alpha(\mathbb{E}[L \mid X] + \eta(L - \mathbb{E}[L \mid X]))}{\partial\,\eta}\right|_{\eta=1} \\ & + \frac{1}{2}\left|\frac{\partial^2\ \mathrm{VaR}_\alpha(\mathbb{E}[L \mid X] + \eta(L - \mathbb{E}[L \mid X]))}{\partial\,\eta^2}\right|_{\eta=1}\end{aligned}$$

Under some assumptions (homogeneous portfolio, regularity of the conditional expected loss, single factor model, etc.), Wilde (2001b) and Gordy (2003) show that the second-order Taylor expansion reduces to[116]:

$$\mathrm{VaR}_\alpha(w) \approx \mu(x_\alpha) - \frac{1}{2h(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{h(x)\,\upsilon(x)}{\partial_x\mu(x)}\right)\Bigg|_{x=x_\alpha}$$

where $h(x)$ is the probability density function of X, $\mu(x)$ is the conditional expected loss function:

$$\mu(x) = \mathbb{E}[L \mid X = x]$$

and $\upsilon(x)$ is the conditional variance function:

$$\upsilon(x) = \sigma^2(L \mid X = x)$$

Since $\mu(x_\alpha) = \mathrm{VaR}_\alpha(w_{\mathrm{IFG}})$, we deduce that:

$$\mathrm{VaR}_\alpha(w) \approx \mathrm{VaR}_\alpha(w_{\mathrm{IFG}}) + \mathrm{GA}$$

where the granularity adjustment is equal to:

$$\begin{aligned}\mathrm{GA} &= -\frac{1}{2h(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{h(x)\,\upsilon(x)}{\partial_x\mu(x)}\right)\Bigg|_{x=x_\alpha} \\ &= \frac{1}{2}\upsilon(x_\alpha)\frac{\partial_x^2\mu(x_\alpha)}{(\partial_x\mu(x_\alpha))^2} - \frac{1}{2}\frac{\partial_x\upsilon(x_\alpha)}{\partial_x\mu(x_\alpha)} - \frac{1}{2}\upsilon(x_\alpha)\frac{\partial_x \ln h(x_\alpha)}{\partial_x\mu(x_\alpha)}\end{aligned}$$

The granularity adjustment has been extensively studied[117]. Originally, the Basel Committee proposed to include the granularity adjustment in the first pillar (BCBS, 2001a), but it has finally preferred to move this issue into the second pillar[118].

[116] In fact, we can show that the first derivative vanishes (Gouriéroux *et al.*, 2000). If we remember the Euler allocation principle presented on page 105, this is not surprising since $\mathrm{VaR}_\alpha(\mathbb{E}[L \mid X])$ is the sum of risk contributions and already includes the first-order effects. In this case, it only remains the second-order effects.

[117] See for example Gordy (2003, 2004), Gordy and Marrone (2012), Gordy and Lütkebohmert (2013). The works of Wilde (2001a,b) and Emmer and Tasche (2005) are a good introduction to this topic.

[118] See Exercise 3.4.7 on page 253 for a derivation of the original Basel granularity adjustment.

3.4 Exercises

3.4.1 Single- and multi-name credit default swaps

1. We assume that the default time $\boldsymbol{\tau}$ follows an exponential distribution with parameter λ. Write the cumulative distribution function $\mathbf{F}$, the survival function $\mathbf{S}$ and the density function f of the random variable $\boldsymbol{\tau}$. How do we simulate this default time?

2. We consider a CDS 3M with two-year maturity and \$1 mn notional principal. The recovery rate $\boldsymbol{\mathcal{R}}$ is equal to 40% whereas the spread $\boldsymbol{s}$ is equal to 150 bps at the inception date. We assume that the protection leg is paid at the default time.

   (a) Give the cash flow chart. What is the P&L of the protection seller A if the reference entity does not default? What is the P&L of the protection buyer B if the reference entity defaults in one year and two months?

   (b) What is the relationship between $\boldsymbol{s}$, $\boldsymbol{\mathcal{R}}$ and λ? What is the implied one-year default probability at the inception date?

   (c) Seven months later, the CDS spread has increased and is equal to 450 bps. Estimate the new default probability. The protection buyer B decides to realize his P&L. For that, he reassigns the CDS contract to the counterparty C. Explain the offsetting mechanism if the risky PV01 is equal to 1.189.

3. We consider the following CDS spread curves for three reference entities:

Maturity	#1	#2	#3
6M	130 bps	1 280 bps	30 bps
1Y	135 bps	970 bps	35 bps
3Y	140 bps	750 bps	50 bps
5Y	150 bps	600 bps	80 bps

   (a) Define the notion of credit curve. Comment the previous spread curves.

   (b) Using the Merton Model, we estimate that the one-year default probability is equal to 2.5% for #1, 5% for #2 and 2% for #3 at a five-year time horizon. Which arbitrage position could we consider about the reference entity #2?

4. We consider a basket of n single-name CDS.

   (a) What is a first-to-default (FtD), a second-to-default (StD) and a last-to-default (LtD)?

   (b) Define the notion of default correlation. What is its impact on the three previous spreads?

   (c) We assume that $n = 3$. Show the following relationship:

$$\boldsymbol{s}_1^{\text{CDS}} + \boldsymbol{s}_2^{\text{CDS}} + \boldsymbol{s}_3^{\text{CDS}} = \boldsymbol{s}^{\text{FtD}} + \boldsymbol{s}^{\text{StD}} + \boldsymbol{s}^{\text{LtD}}$$

   where $\boldsymbol{s}_i^{\text{CDS}}$ is the CDS spread of the i^{th} reference entity.

   (d) Many professionals and academics believe that the subprime crisis is due to the use of the Normal copula. Using the results of the previous question, what could you conclude?

3.4.2 Risk contribution in the Basel II model

1. We note L the portfolio loss of n credit and w_i the exposure at default of the i^{th} credit. We have:

$$L(w) = w^\top \varepsilon = \sum_{i=1}^{n} w_i \cdot \varepsilon_i \tag{3.61}$$

where ε_i is the unit loss of the i^{th} credit. Let $\mathbf{F}$ be the cumulative distribution function of $L(w)$.

   (a) We assume that $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Compute the value-at-risk $\text{VaR}_\alpha(w)$ of the portfolio when the confidence level is equal to α.

   (b) Deduce the marginal value-at-risk of the i^{th} credit. Define then the risk contribution $\mathcal{RC}_i$ of the i^{th} credit.

   (c) Check that the marginal value-at-risk is equal to:

$$\frac{\partial\, \text{VaR}_\alpha(w)}{\partial\, w_i} = \mathbb{E}\left[\varepsilon_i \mid L(w) = \mathbf{F}^{-1}(\alpha)\right]$$

   Comment on this result.

2. We consider the Basel II model of credit risk and the value-at-risk risk measure. The expression of the portfolio loss is given by:

$$L = \sum_{i=1}^{n} \text{EAD}_i \cdot \text{LGD}_i \cdot \mathbb{1}\{\boldsymbol{\tau}_i < T_i\} \tag{3.62}$$

   (a) Define the different parameters EAD_i, LGD_i, $\boldsymbol{\tau}_i$ and T_i. Show that Model (3.62) can be written as Model (3.61) by identifying w_i and ε_i.

   (b) What are the necessary assumptions $\mathcal{H}$ to obtain this result:

$$\mathbb{E}\left[\varepsilon_i \mid L = \mathbf{F}^{-1}(\alpha)\right] = \mathbb{E}\left[\text{LGD}_i\right] \cdot \mathbb{E}\left[D_i \mid L = \mathbf{F}^{-1}(\alpha)\right]$$

   with $D_i = \mathbb{1}\{\boldsymbol{\tau}_i < T_i\}$.

   (c) Deduce the risk contribution $\mathcal{RC}_i$ of the i^{th} credit and the value-at-risk of the credit portfolio.

   (d) We assume that the credit i defaults before the maturity T_i if a latent variable Z_i goes below a barrier B_i:

$$\boldsymbol{\tau}_i \leq T_i \Leftrightarrow Z_i \leq B_i$$

   We consider that $Z_i = \sqrt{\rho} \cdot X + \sqrt{1-\rho} \cdot \varepsilon_i$ where Z_i, X and ε_i are three independent Gaussian variables $\mathcal{N}(0,1)$. X is the factor (or the systematic risk) and ε_i is the idiosyncratic risk.

      i. Interpret the parameter ρ.

      ii. Calculate the unconditional default probability:

$$p_i = \Pr\{\boldsymbol{\tau}_i \leq T_i\}$$

      iii. Calculate the conditional default probability:

$$p_i(x) = \Pr\{\boldsymbol{\tau}_i \leq T_i \mid X = x\}$$

(e) Show that, under the previous assumptions $\mathcal{H}$, the risk contribution $\mathcal{RC}_i$ of the i^{th} credit is:

$$\mathcal{RC}_i = \text{EAD}_i \cdot \mathbb{E}[\text{LGD}_i] \cdot \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{3.63}$$

when the risk measure is the value-at-risk.

3. We now assume that the risk measure is the expected shortfall:

$$\text{ES}_\alpha(w) = \mathbb{E}[L \mid L \geq \text{VaR}_\alpha(w)]$$

(a) In the case of the Basel II framework, show that we have:

$$\text{ES}_\alpha(w) = \sum_{i=1}^{n} \text{EAD}_i \cdot \mathbb{E}[\text{LGD}_i] \cdot \mathbb{E}\left[p_i(X) \mid X \leq \Phi^{-1}(1-\alpha)\right]$$

(b) By using the following result:

$$\int_{-\infty}^{c} \Phi(a+bx)\phi(x)\,\mathrm{d}x = \Phi_2\left(c, \frac{a}{\sqrt{1+b^2}}; \frac{-b}{\sqrt{1+b^2}}\right)$$

where $\Phi_2(x, y; \rho)$ is the cdf of the bivariate Gaussian distribution with correlation ρ on the space $[-\infty, x] \cdot [-\infty, y]$, deduce that the risk contribution $\mathcal{RC}_i$ of the i^{th} credit in the Basel II model is:

$$\mathcal{RC}_i = \text{EAD}_i \cdot \mathbb{E}[\text{LGD}_i] \cdot \frac{\mathbf{C}\left(1-\alpha, p_i; \sqrt{\rho}\right)}{1-\alpha} \tag{3.64}$$

where $\mathbf{C}(u_1, u_2; \theta)$ is the Normal copula with parameter θ.

(c) What do the results (3.63) and (3.64) become if the correlation ρ is equal to zero? Same question if $\rho = 1$.

4. The risk contributions (3.63) and (3.64) were obtained by considering the assumptions $\mathcal{H}$ and the default model defined in Question 2(d). What are the implications in terms of Pillar 2?

3.4.3 Calibration of the piecewise exponential model

1. We denote by $\mathbf{F}$ and $\mathbf{S}$ the distribution and survival functions of the default time $\boldsymbol{\tau}$. Define the function $\mathbf{S}(t)$ and deduce the expression of the associated density function $f(t)$.

2. Define the hazard rate $\lambda(t)$. Deduce that the exponential model corresponds to the particular case $\lambda(t) = \lambda$.

3. We assume that the interest rate r is constant. In a continuous-time model, we recall that the CDS spread is given by the following expression:

$$\boldsymbol{s}(T) = \frac{(1-\boldsymbol{\mathcal{R}}) \cdot \int_0^T e^{-rt} f(t)\,\mathrm{d}t}{\int_0^T e^{-rt}\mathbf{S}(t)\,\mathrm{d}t} \tag{3.65}$$

where $\boldsymbol{\mathcal{R}}$ is the recovery rate and T is the maturity of the CDS. Find the triangle relationship when $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$.

4. Let us assume that:
$$\lambda(t) = \begin{cases} \lambda_1 & \text{if } t \leq 3 \\ \lambda_2 & \text{if } 3 < t \leq 5 \\ \lambda_3 & \text{if } t > 5 \end{cases}$$

   (a) Give the expression of the survival function $\mathbf{S}(t)$ and calculate the density function $f(t)$. Verify that the hazard rate $\lambda(t)$ is a piecewise constant function.

   (b) Find the expression of the CDS spread using Equation (3.65).

   (c) We consider three credit default swaps, whose maturities are respectively equal to 3, 5 and 7 years. Show that the calibration of the piecewise exponential model implies to solve a set of 3 equations with the unknown variables λ_1, λ_2 and λ_3. What is the name of this calibration method?

   (d) Find an approximated solution when r is equal to zero and λ_m is small. Comment on this result.

   (e) We consider the following numerical application: $r = 5\%$, $\boldsymbol{s}(3) = 100$ bps, $\boldsymbol{s}(5) = 150$ bps, $\boldsymbol{s}(7) = 160$ bps and $\boldsymbol{\mathcal{R}} = 40\%$. Estimate the implied hazard function.

   (f) Using the previous numerical results, simulate the default time with the uniform random numbers 0.96, 0.23, 0.90 and 0.80.

3.4.4 Modeling loss given default

1. What is the difference between the recovery rate and the loss given default?

2. We consider a bank that grants 250 000 credits per year. The average amount of a credit is equal to \$50 000. We estimate that the average default probability is equal to 1% and the average recovery rate is equal to 65%. The total annual cost of the litigation department is equal to \$12.5 mn. Give an estimation of the loss given default?

3. The probability density function of the beta probability distribution $\mathcal{B}(\alpha, \beta)$ is:
$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathfrak{B}(\alpha, \beta)}$$
   where $\mathfrak{B}(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u$.

   (a) Why is the beta probability distribution a good candidate to model the loss given default? Which parameter pair (α, β) does correspond to the uniform probability distribution?

   (b) Let us consider a sample $(x_1, \ldots, x_n)$ of n losses in case of default. Write the log-likelihood function. Deduce the first-order conditions of the maximum likelihood estimator.

   (c) We recall that the first two moments of the beta probability distribution are:
$$\begin{aligned} \mathbb{E}[X] &= \frac{\alpha}{\alpha+\beta} \\ \sigma^2(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$
   Find the method of moments estimator.

4. We consider a risk class $\mathcal{C}$ corresponding to a customer/product segmentation specific to retail banking. A statistical analysis of 1 000 loss data available for this risk class gives the following results:

LGD_k	0%	25%	50%	75%	100%
n_k	100	100	600	100	100

where n_k is the number of observations corresponding to LGD_k.

   (a) We consider a portfolio of 100 homogeneous credits, which belong to the risk class $\mathcal{C}$. The notional is \$10 000 whereas the annual default probability is equal to 1%. Calculate the expected loss of this credit portfolio with a one-year time horizon if we use the previous empirical distribution to model the LGD parameter.

   (b) We assume that the LGD parameter follows a beta distribution $\mathcal{B}(\alpha, \beta)$. Calibrate the parameters α and β with the method of moments.

   (c) We assume that the Basel II model is valid. We consider the portfolio described in Question 4(a) and calculate the unexpected loss. What is the impact if we use a uniform probability distribution instead of the calibrated beta probability distribution? Why does this result hold even if we consider different factors to model the default time?

3.4.5 Modeling default times with a Markov chain

We consider a rating system with 4 risk classes (A, B, C and D), where rating D represents the default. The transition probability matrix with a two-year time horizon is equal to:

$$P(2) = \begin{pmatrix} 94\% & 3\% & 2\% & 1\% \\ 10\% & 80\% & 5\% & 5\% \\ 10\% & 10\% & 60\% & 20\% \\ 0\% & 0\% & 0\% & 100\% \end{pmatrix}$$

We also have:

$$P(4) = \begin{pmatrix} 88.860\% & 5.420\% & 3.230\% & 2.490\% \\ 17.900\% & 64.800\% & 7.200\% & 10.100\% \\ 16.400\% & 14.300\% & 36.700\% & 32.600\% \\ 0.000\% & 0.000\% & 0.000\% & 100.000\% \end{pmatrix}$$

and:

$$P(6) = \begin{pmatrix} 84.393\% & 7.325\% & 3.986\% & 4.296\% \\ 24.026\% & 53.097\% & 7.918\% & 14.959\% \\ 20.516\% & 15.602\% & 23.063\% & 40.819\% \\ 0.000\% & 0.000\% & 0.000\% & 100.000\% \end{pmatrix}$$

Let us denote by $\mathbf{S}_A(t)$, $\mathbf{S}_B(t)$ and $\mathbf{S}_C(t)$ the survival functions of each risk class A, B and C.

1. How are the matrices $P(4)$ and $P(6)$ calculated?

2. Assuming a piecewise exponential model, calibrate the hazard function of each risk class for $0 < t \leq 2$, $2 < t \leq 4$ and $4 < t \leq 6$.

3. Give the definition of a Markovian generator. How can we estimate the generator Λ associated to the transition probability matrices? Verify numerically that the direct estimator is equal to:

$$\hat{\Lambda} = \begin{pmatrix} -3.254 & 1.652 & 1.264 & 0.337 \\ 5.578 & -11.488 & 3.533 & 2.377 \\ 6.215 & 7.108 & -25.916 & 12.593 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix} \times 10^{-2}$$

4. In Figure 3.59, we show the hazard function $\lambda(t)$ deduced from Questions 2 and 3. Explain how do we calculate $\lambda(t)$ in both cases. Why do we obtain an increasing curve for rating A, a decreasing curve for rating C and an inverted U-shaped curve for rating B?

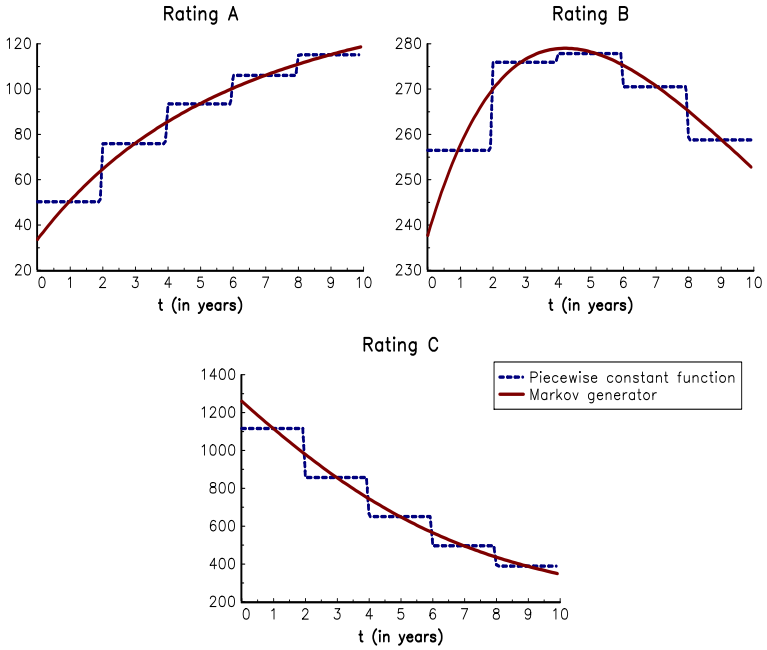


FIGURE 3.59: Hazard function $\lambda(t)$ (in bps) estimated respectively with the piecewise exponential model and the Markov generator

3.4.6 Continuous-time modeling of default risk

We consider a credit rating system with four risk classes (A, B, C and D), where rating D represents the default. The one-year transition probability matrix is equal to:

$$P = P(1) = \begin{pmatrix} 94\% & 3\% & 2\% & 1\% \\ 10\% & 80\% & 7\% & 3\% \\ 5\% & 15\% & 60\% & 20\% \\ 0\% & 0\% & 0\% & 0\% \end{pmatrix}$$

We denote by $\mathbf{S}_A(t)$, $\mathbf{S}_B(t)$ and $\mathbf{S}_C(t)$ the survival functions of each risk class A, B and C.

1. Explain how we can calculate the n-year transition probability matrix $P(n)$? Find the transition probability matrix $P(10)$.

2. Let $V = \left(V_1 \vdots V_2 \vdots V_3 \vdots V_4\right)$ and $D = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ be the matrices of eigenvectors and eigenvalues associated to P.

   (a) Show that:
   $$P(n)\, V = V D^n$$
   Deduce a second approach for calculating the n-year transition probability matrix $P(n)$.

   (b) Calculate the eigendecomposition of the transition probability matrix P. Deduce the transition probability matrix $P(10)$.

3. We assume that the default time follows a piecewise exponential model. Let $\mathbf{S}_i(n)$ and $\lambda_i(n)$ be the survival function and the hazard rate of a firm whose initial rating is the state i (A, B or C). Give the expression of $\mathbf{S}_i(n)$ and $\lambda_i(n)$. Show that:
   $$\lambda_i(1) = -\ln\left(1 - \mathbf{e}_i^\top P^n \mathbf{e}_4\right)$$
   Calculate $\mathbf{S}_i(n)$ and $\lambda_i(n)$ for $n \in \{0, \ldots, 10, 50, 100\}$.

4. Give the definition of a Markov generator. How can we estimate the generator Λ associated to the transition probability matrices? Give an estimate $\hat{\Lambda}$.

5. Explain how we can calculate the transition probability matrix $P(t)$ for the time horizon $t \geq 0$. Give the theoretical approximation of $P(t)$ based on Taylor expansion. Calculate the 6-month transition probability matrix.

6. Deduce the expression of $\mathbf{S}_i(t)$ and $\lambda_i(t)$.

3.4.7 Derivation of the original Basel granularity adjustment

In this exercise, we derive the formula of the granularity adjustment that was proposed by the Basel Committee in 2001. The mathematical proof follows Chapter 8 (§422 to §457) of BCBS (2001a) and the works of Wilde (2001a,b) and Gordy (2003, 2004). We encourage the reader to consult carefully these references. Most of the time, we use the notations of the Basel Committee[119]. We consider the Basel model that has been presented in Section 3.2.3.2 on page 169.

1. We consider the normalized loss:
   $$L_i = \text{LGD}_i \cdot D_i$$
   We assume that the conditional probability of default is given by the CreditRisk+ model (Gordy, 2000):
   $$p_i(X) = p_i\left(1 + \varpi_i (X - 1)\right)$$
   where $\varpi_i \in [0, 1]$ is the factor weight and X is the systematic risk factor, which follows the gamma distribution $\mathcal{G}(\alpha_g, \beta_g)$. Calculate the conditional expected loss[120]:
   $$\mu(x) = \mathbb{E}[L_i \mid X = x]$$

[119] When they are different, we indicate the changes in footnotes.
[120] We use the notation $E_i = \mathbb{E}[\text{LGD}_i]$.

and the conditional variance:

$$\upsilon(x) = \sigma^2(L_i \mid X = x)$$

The Basel Committee assumes that (BCBS, 2001a, §447):

$$\sigma(\text{LGD}_i) = \frac{1}{2}\sqrt{E_i(1-E_i)}$$

Deduce that we have the following approximation:

$$\upsilon(x) \approx E_i\left(\frac{1}{4} + \frac{3}{4}E_i\right) p_i(1 + \varpi_i(x-1))$$

2. Calculate the granularity adjustment function:

$$\beta(x) = \frac{1}{2h(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{h(x)\upsilon(x)}{\partial_x \mu(x)}\right)$$

3. In order to maintain the coherency with the IRB formula, the Basel Committee imposes that the conditional probabilities are the same for the IRB formula (Vasicek model) and the granularity formula (CreditRisk+ model). Show that:

$$\varpi_i = \frac{1}{(x-1)}\frac{F_i}{p_i}$$

where:

$$F_i = \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) - p_i$$

Deduce the expression of $\beta(x)$.

4. The calibration has been done by assuming that $\mathbb{E}[X] = 1$ and $\sigma(X) = 2$ (BCBS, 2001a, §445). Show that:

$$\beta(x_\alpha) = (0.4 + 1.2 \cdot E_i)\left(0.76229640 + 1.0747964 \cdot \frac{p_i}{F_i}\right)$$

We recall that the Basel Committee finds the following expression of $\beta(x_\alpha)$:

$$\beta(x_\alpha) = (0.4 + 1.2 \cdot E_i)\left(0.76 + 1.10 \cdot \frac{p_i}{F_i}\right)$$

How to obtain exactly this formula?

5. In order to transform the granularity adjustment function $\beta(x_\alpha)$ into risk-weighted assets, the Basel Committee indicates that it uses a scaling factor $c = 1.5$ (BCBS, 2001a, §457). Moreover, the Basel Committee explains that the "*the baseline IRB risk-weights for non-retail assets (i.e. the RWA before granularity adjustment) incorporate a margin of 4% to cover average granularity*". Let $w^\star$ be the equivalent homogenous portfolio of the current portfolio w. Show that the granularity adjustment is equal to[121]:

$$\text{GA} = \frac{\text{EAD}^\star}{n^\star} \cdot \text{GSF} - 0.04 \cdot \text{RWA}_{\text{NR}}$$

[121] The Basel Committee uses the notation $\square_{\text{AG}}$ instead of $\square^\star$ for the equivalent homogeneous portfolio. The global exposure $\text{EAD}^\star$ corresponds to the variable TNRE (total non-retail exposure) of the Basel Committee.

where RWA_{NR} are the risk-weighted assets for non-retail assets and:

$$\text{GSF} = (0.6 + 1.8 \cdot E^\star) \left(9.5 + 13.75 \cdot \frac{p^\star}{F^\star}\right)$$

6. The Basel Committee considers the following definition of the portfolio loss:

$$L = \sum_{j=1}^{n_{\mathcal{C}}} \sum_{i \in \mathcal{C}_j} \text{EAD}_i \cdot \text{LGD}_i \cdot D_i$$

where $\mathcal{C}_j$ is the j^{th} class of risk. Find the equivalent homogeneous portfolio $w^\star$ of size $n^\star$ and exposure $\text{EAD}^\star$. Calibrate the parameters $p^\star$, $E^\star$ and $\sigma\left(\text{LGD}^\star\right)$.

7. Using the notations of BCBS (2001a), summarize the different steps for computing the original Basel granularity adjustment.

3.4.8 Variance of the conditional portfolio loss

The portfolio loss is given by:

$$L = \sum_{i=1}^{n} w_i \cdot \text{LGD}_i \cdot D_i$$

where w_i is the exposure at default of the i^{th} credit, LGD_i is the loss given default, T_i is the residual maturity and $D_i = \mathbb{1}\{\boldsymbol{\tau}_i \leq T_i\}$ is the default indicator function. We suppose the assumptions of the Basel II model are satisfied. We note $D_i(X)$ and $p_i(X)$ the conditional default indicator function and the conditional default probability with respect to the risk factor X.

1. Define $D_i(X)$. Calculate $\mathbb{E}[D_i(X)]$, $\mathbb{E}\left[D_i^2(X)\right]$ and $\mathbb{E}[D_i(X) D_j(X)]$.
2. Define the conditional portfolio loss $L(X)$.
3. Calculate the expectation of $L(X)$.
4. Show that the variance of $L(X)$ is equal to:

$$\sigma^2(L(X)) = \sum_{i=1}^{n} w_i^2 \left(\mathbb{E}[D_i(X)]\, \sigma^2(\text{LGD}_i) + \mathbb{E}^2[\text{LGD}_i]\, \sigma^2(D_i(X))\right)$$

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 4

Counterparty Credit Risk and Collateral Risk

Counterparty credit risk and collateral risk are other forms of credit risk, where the underlying credit risk is not directly generated by the economic objective of the financial transaction. Therefore, it can reduce the P&L of the portfolio and create a loss even if the business objective is reached. A typical example is the purchase transaction of a credit default swap. In this case, we have previously seen that the protection buyer is hedged against the credit risk if the reference entity defaults. This is partially true, because the protection buyer faces the risk that the protection seller also defaults. In this example, we see that the total P&L of the financial transaction is the direct P&L of the economic objective minus the potential loss due to the transaction settlement. Another example concerns the collateral risk, since the P&L of the financial transaction is directly affected by the mark-to-market of the collateral portfolio.

In this chapter, we study the counterparty credit risk (CCR) and show its computation. We also focus on the regulatory framework that has evolved considerably since the collapse of the LTCM hedge fund in 1997, which has shocked the entire financial system, not because of the investor losses, but because of the indirect losses generated by the counterparty credit risk[1]. The second section is dedicated to the credit valuation adjustment (CVA), which can be considered as the '*little brother*' of the CCR. This risk has been mainly identified with the bankruptcy of Lehman Brothers, which has highlighted the market risk of CCR. Finally, Section three reviews different topics associated to the collateral risk management, particularly in the repo markets.

4.1 Counterparty credit risk

We generally make the distinction between credit risk (CR) and counterparty credit risk (CCR). The counterparty credit risk on market transactions is the risk that the counterparty could default before the final settlement of the transaction's cash flows. For instance, if the bank buys a CDS protection on a firm and the seller of the CDS protection defaults before the maturity of the contract, the bank could not be hedged against the default of the firm. Another example of CCR is the delivery/settlement risk. Indeed, few financial transactions are settled on the same-day basis and the difference between the payment date and the delivery date is generally between one and five business days. There is then a counterparty credit risk if one counterparty defaults when the payment date is not synchronized with the delivery date. This settlement risk is low when it is expressed as a percent of the notional amount because the maturity mismatch is short, but it concerns large amounts from an aggregate point of view. In a similar way, when an OTC contract has a positive mark-to-

[1] Chapter 8 on page 453 describes the impact of the LTCM bankruptcy on systemic risk.

market, the bank suffers a loss if the counterparty defaults. To reduce this risk, the bank can put in place bilateral netting agreements. We note that this risk disappears (or more precisely decreases) when the bank uses an exchange, because the counterparty credit risk is transferred to the central counterparty clearing house, which guarantees the expected cash flows.

4.1.1 Definition

BCBS (2004a) measures the counterparty credit risk by the replacement cost of the OTC derivative. Let us consider two banks A and B that have entered into an OTC contract $\mathfrak{C}$. We assume that the bank B defaults before the maturity of the contract. According to Pykhtin and Zhu (2006), Bank A can then face two situations:

- The current value of the contract $\mathfrak{C}$ is negative. In this case, Bank A closes out the position and pays the market value of the contract to Bank B. To replace the contract $\mathfrak{C}$, Bank A can enter with another counterparty C into a similar contract $\mathfrak{C}'$. For that, Bank A receives the market value of the contract $\mathfrak{C}'$ and the loss of the bank is equal to zero.
- The current value of the contract $\mathfrak{C}$ is positive. In this case, Bank A close out the position, but receives nothing from Bank B. To replace the contract, Bank A can enter with another counterparty C into a similar contract $\mathfrak{C}'$. For that, Bank A pays the market value of the contract $\mathfrak{C}'$ to C. In this case, the loss of the bank is exactly equal to the market value.

We note that the counterparty exposure is then the maximum of the market value and zero. Moreover, the counterparty credit risk differs from the credit risk by two main aspects (Canabarro and Duffie, 2003):

1. The counterparty credit risk is bilateral, meaning that both counterparties may face losses. In the previous example, Bank B is also exposed to the risk that Bank A defaults.
2. The exposure at default is uncertain, because we don't know what will be the replacement cost of the contract when the counterparty defaults.

Using the notations introduced in the previous chapter, we deduce that the credit loss of an OTC portfolio is:

$$L = \sum_{i=1}^{n} \text{EAD}_i(\boldsymbol{\tau}_i) \cdot \text{LGD}_i \cdot \mathbb{1}\{\boldsymbol{\tau}_i \leq T_i\}$$

This is the formula of a credit portfolio loss, except that the exposure at default is random and depends on different factors: the default time of the counterparty, the evolution of market risk factors and the correlation between the market value of the OTC contract and the default of the counterparty.

Let $\text{MtM}(t)$ be the mark-to-market value of the OTC contract at time t. The exposure at default is defined as:

$$\text{EAD} = \max(\text{MtM}(\boldsymbol{\tau}), 0)$$

If we consider a portfolio of OTC derivatives with the same counterparty entity, the exposure at default is the sum of positive market values:

$$\text{EAD} = \sum_{i=1}^{n} \max(\text{MtM}_i(\boldsymbol{\tau}), 0)$$

This is why the bank may be interested in putting in place a global netting agreement:

$$\begin{aligned}\text{EAD} &= \max\left(\sum_{i=1}^{n}\text{MtM}_i\left(\boldsymbol{\tau}\right),0\right)\\ &\leq \sum_{i=1}^{n}\max\left(\text{MtM}_i\left(\boldsymbol{\tau}\right),0\right)\end{aligned}$$

In practice, it is extremely complicated and rare that two counterparties succeed in signing such agreement. Most of the time, there are several netting agreements on different trading perimeters (equities, bonds, interest rate swaps, etc.). In this case, the exposure at default is:

$$\text{EAD}=\sum_{k}\max\left(\sum_{i\in\mathcal{N}_k}\text{MtM}_i\left(\boldsymbol{\tau}\right),0\right)+\sum_{i\notin\cup\mathcal{N}_k}\max\left(\text{MtM}_i\left(\boldsymbol{\tau}\right),0\right)$$

where $\mathcal{N}_k$ corresponds to the k^{th} netting arrangement and defines a netting set. Since the default of Lehman Brothers, we observe a strong development of (global and partial) netting agreements in order to reduce potential losses, but also the capital charge induced by counterparty credit risk.

Example 43 *Banks A and B have traded five OTC products, whose mark-to-market values*[2] *are given in the table below:*

t	1	2	3	4	5	6	7	8
$\mathfrak{C}_1$	5	5	3	0	−4	0	5	8
$\mathfrak{C}_2$	−5	10	5	−3	−2	−8	−7	−10
$\mathfrak{C}_3$	0	2	−3	−4	−6	−3	0	5
$\mathfrak{C}_4$	2	−5	−5	−5	2	3	5	7
$\mathfrak{C}_5$	−1	−3	−4	−5	−7	−6	−7	−6

If we suppose that there is no netting agreement, the counterparty exposure of Bank A corresponds to the second row in Table 4.1. We notice that the exposure changes over time. If there is a netting agreement, we obtain lower exposures. We now consider a more complicated situation. We assume that Banks A and B have two netting agreements: one on equity OTC contracts ($\mathfrak{C}_1$ and $\mathfrak{C}_2$) and one on fixed income OTC contracts ($\mathfrak{C}_3$ and $\mathfrak{C}_4$). In this case, we obtain results given in the last row in Table 4.1. For instance, the exposure at default for $t=8$ is calculated as follows:

$$\text{EAD}=\max\left(8-10,0\right)+\max\left(5+7,0\right)+\max\left(-6,0\right)=12$$

TABLE 4.1: Counterparty exposure of Bank A

t	1	2	3	4	5	6	7	8
No netting	7	17	8	0	2	3	10	20
Global netting	1	9	0	0	0	0	0	4
Partial netting	2	15	8	0	0	0	5	12

If we consider Bank B, the counterparty exposure is given in Table 4.2. This illustrates the bilateral nature of the counterparty credit risk. Indeed, except if there is a global netting arrangement, both banks have a positive counterparty exposure.

[2] They are calculated from the viewpoint of Bank A.

TABLE 4.2: Counterparty exposure of Bank B

t	1	2	3	4	5	6	7	8
No netting	6	8	12	17	19	17	14	16
Global netting	0	0	4	17	17	14	4	0
Partial netting	1	6	12	17	17	14	9	8

Remark 49 *In the previous example, we have assumed that the mark-to-market value of the OTC contract for one bank is exactly the opposite of the mark-to-market value for the other bank. In practice, banks calculate mark-to-model prices, implying that they can differ from one bank to another one.*

4.1.2 Modeling the exposure at default

In order to understand the counterparty credit risk, we begin by an example and illustrate the time-varying property of the exposure at default. Then, we introduce the different statistical measures that are useful for characterizing the EAD and show how to calculate them.

4.1.2.1 An illustrative example

Example 44 *We consider a bank that buys* 1 000 *ATM call options, whose maturity is one-year. The current value of the underlying asset is equal to* \$100*. We assume that the interest rate* r *and the cost-of-carry parameter* b *are equal to* 5%*. Moreover, the implied volatility of the option is considered as a constant and is equal to* 20%*.*

By considering the previous parameters, the value $\boldsymbol{\mathcal{C}}_0$ of the call option[3] is equal to \$10.45. At time t, the mark-to-market of this derivative exposure is defined by:

$$\mathrm{MtM}(t) = n_C \cdot (\boldsymbol{\mathcal{C}}(t) - \boldsymbol{\mathcal{C}}_0)$$

where n_C and $\boldsymbol{\mathcal{C}}(t)$ are the number and the value of call options. Let $e(t)$ be the exposure at default. We have:

$$e(t) = \max(\mathrm{MtM}(t), 0)$$

At the initial date of the trade, the mark-to-market value and the counterparty exposure are zero. When $t > 0$, the mark-to-market value is not equal to zero, implying that the counterparty exposure $e(t)$ may be positive. In Table 4.3, we have reported the values taken by $\boldsymbol{\mathcal{C}}(t)$, $\mathrm{MtM}(t)$ and $e(t)$ for two scenarios of the underlying price $S(t)$. If we consider the first scenario, the counterparty exposure is equal to zero during the first three months, because the mark-to-market value is negative. The counterparty exposure is then positive for the next four months. For instance, it is equal to \$2 519 at the end of the fourth month[4]. In the case of the second scenario, the counterparty exposure is always equal to zero except for two months. Therefore, we notice that the counterparty exposure is time-varying and depends of the trajectory of the underlying price. This implies that the counterparty exposure cannot be calculated once and for all at the initial date of the trade. Indeed, the counterparty exposure changes with time. Moreover, we don't known what the future price of the underlying asset will be. That's why we are going to simulate it.

[3] We use the Black-Scholes formula given by Equation (2.10) on page 94 to price the option.

[4] We have:

$$\mathrm{MtM}(t) = 1\,000 \times (12.969 - 10.450) = \$2\,519$$

TABLE 4.3: Mark-to-market and counterparty exposure of the call option

t	Scenario #1				Scenario #2			
	$S(t)$	$\boldsymbol{C}(t)$	$\mathrm{MtM}(t)$	$e(t)$	$S(t)$	$\boldsymbol{C}(t)$	$\mathrm{MtM}(t)$	$e(t)$
1M	97.58	8.44	−2 013	0	91.63	5.36	−5 092	0
2M	98.19	8.25	−2 199	0	89.17	3.89	−6 564	0
3M	95.59	6.26	−4 188	0	97.60	7.35	−3 099	0
4M	106.97	12.97	2 519	2 519	97.59	6.77	−3 683	0
5M	104.95	10.83	382	382	96.29	5.48	−4 970	0
6M	110.73	14.68	4 232	4 232	97.14	5.29	−5 157	0
7M	113.20	16.15	5 700	5 700	107.71	11.55	1 098	1 098
8M	102.04	6.69	−3 761	0	105.71	9.27	−1 182	0
9M	115.76	17.25	6 802	6 802	107.87	10.18	−272	0
10M	103.58	5.96	−4 487	0	108.40	9.82	−630	0
11M	104.28	5.41	−5 043	0	104.68	5.73	−4 720	0
1Y	104.80	4.80	−5 646	0	115.46	15.46	5 013	5 013

We note $\mathrm{MtM}(t_1; t_2)$ the mark-to-market value between dates t_1 and t_2. By construction, we have:

$$\mathrm{MtM}(0; t) = \mathrm{MtM}(0; t_0) + \mathrm{MtM}(t_0; t)$$

where 0 is the initial date of the trade, t_0 is the current date and t is the future date. This implies that the mark-to-market value at time t has two components:

1. the current mark-to-market value $\mathrm{MtM}(0; t_0)$ that depends on the past trajectory of the underlying price;
2. and the future mark-to-market value $\mathrm{MtM}(t_0; t)$ that depends on the future trajectory of the underlying price.

In order to evaluate the second component, we need to define the probability distribution of $S(t)$. In our example, we can assume that the underlying price follows a geometric Brownian motion:

$$\mathrm{d}S(t) = \mu S(t)\ \mathrm{d}t + \sigma S(t)\ \mathrm{d}W(t)$$

We face here an issue because we have to define the parameters μ and σ. There are two approaches:

1. the first method uses the historical probability measure $\mathbb{P}$, meaning that the parameters μ and σ are estimated using historical data;
2. the second method considers the risk-neutral probability measure $\mathbb{Q}$, which is used to price the OTC derivative.

While the first approach is more relevant to calculate the counterparty exposure, the second approach is more frequent because it is easier for a bank to implement it. Indeed, $\mathbb{Q}$ is already available because of the hedging portfolio, which is not the case of $\mathbb{P}$. In our example, this is equivalent to set μ and σ to their historical estimates $\hat{\mu}$ and $\hat{\sigma}$ if we consider the historical probability measure $\mathbb{P}$, while they are equal to the interest rate r and the implied volatility Σ if we consider the risk-neural probability measure $\mathbb{Q}$.

In Figure 4.1, we report an illustration of scenario generation when the current date t_0 is 6 months. This means that the trajectory of the asset price $S(t)$ is given when $t \leq t_0$ whereas it is simulated when $t > t_0$. At time $t_0 = 0.5$, the asset price is equal to \$114.77. We deduce

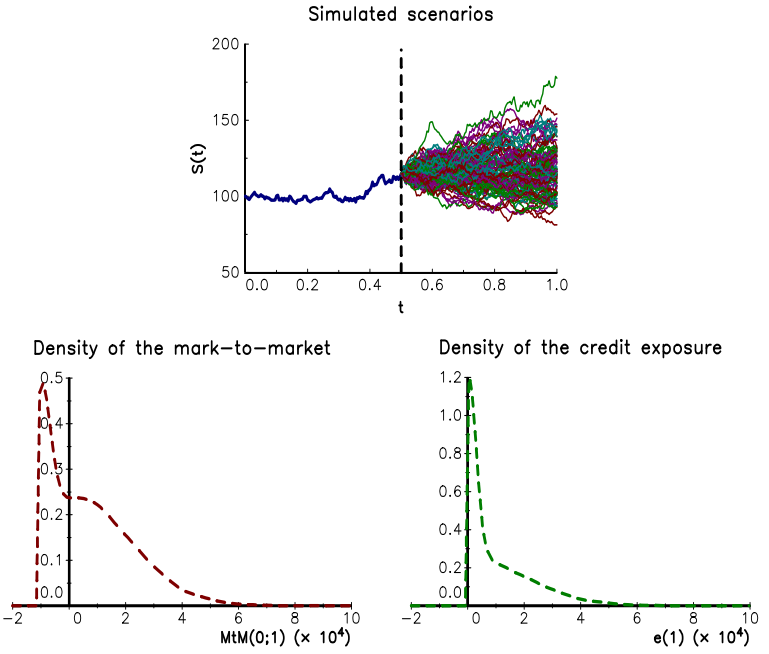


FIGURE 4.1: Probability density function of the counterparty exposure after six months

that the option price $\boldsymbol{\mathcal{C}}(t_0)$ is equal to \$18.17. The mark-to-market value is then positive and equal to \$7 716. Using 10 000 simulated scenarios, we estimate the probability density function of the mark-to-market value $\mathrm{MtM}(0;1)$ at the maturity date (bottom/left panel in Figure 4.1) and deduce the probability density function of the counterparty exposure $e(1)$ (bottom/right panel in Figure 4.1). We notice that the probability to obtain a negative mark-to-market at the maturity date is significant. Indeed, it is equal to 36% because it remains 6 months and the asset price may sufficiently decrease. Of course, this probability depends on the parameters used for simulating the trajectories, especially the trend μ. Using a risk-neutral approach has the advantage to limit the impact of this coefficient.

Remark 50 *The mark-to-market value presents a very high skew, because it is bounded. Indeed, the worst-case scenario is reached when the asset price $S(1)$ is lower than the strike $K = 100$. In this case, we obtain:*

$$\begin{aligned}\mathrm{MtM}(0;1) &= 1\,000 \times (0 - 10.45)\\ &= -\$10\,450\end{aligned}$$

We suppose now that the current date is nine months. During the last three months, the asset price has changed and it is now equal to \$129.49. The current counterparty exposure has then increased and is equal to[5] \$20 294. In Figure 4.2, we observe that the shape of the probability density function has changed. Indeed, the skew has been highly reduced, because it only remains three months before the maturity date. The price is then sufficiently high that the probability to obtain a positive mark-to-market at the settlement date is almost equal to 100%. This is why the two probability density functions are very similar.

We can use the previous approach of scenario generation in order to represent the evolution of counterparty exposure. In Figure 4.3, we consider two observed trajectories of the

[5] Using the previous parameters, the BS price of the call option is now equal to \$30.74.

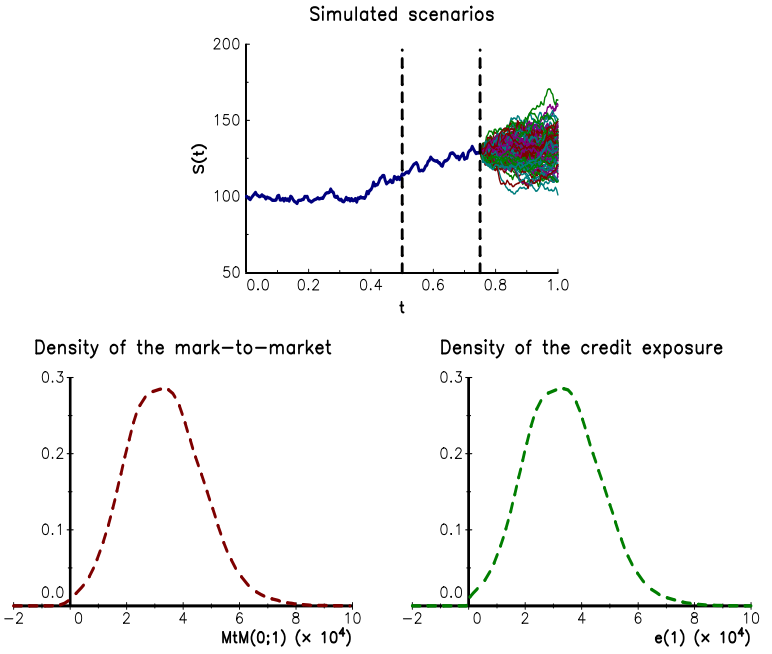


FIGURE 4.2: Probability density function of the counterparty exposure after nine months

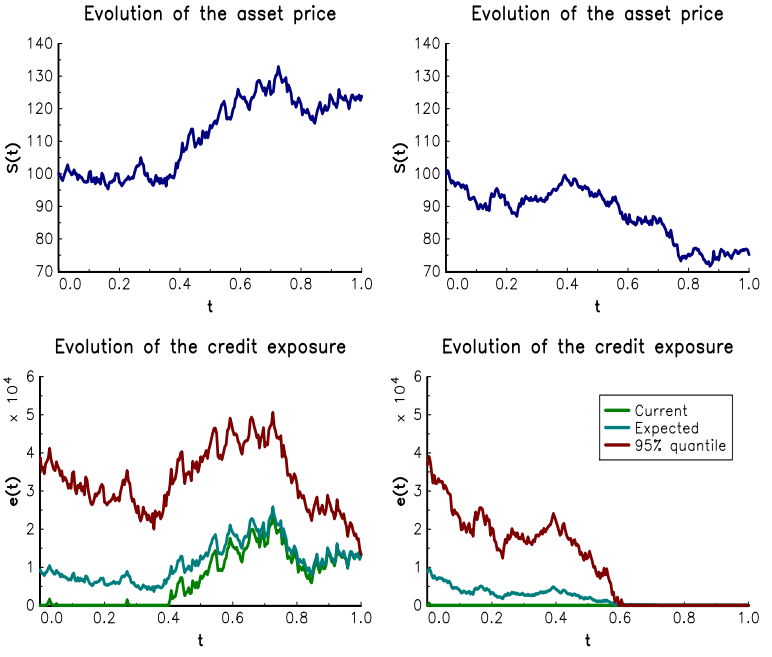


FIGURE 4.3: Evolution of the counterparty exposure

asset price. For each trajectory, we report the current exposure, the expected exposure and the 95% quantile of the counterparty exposure at the maturity date. All these counterparty measures converge at the maturity date, but differ before because of the uncertainty between the current date and the maturity date.

4.1.2.2 Measuring the counterparty exposure

We define the counterparty exposure at time t as the random credit exposure[6]:

$$e(t) = \max(\text{MtM}(0;t), 0) \tag{4.1}$$

This counterparty exposure is also known as the potential future exposure (PFE). When the current date t_0 is not equal to the initial date 0, the counterparty exposure can be decomposed in two parts:

$$\begin{aligned} e(t) &= \max(\text{MtM}(0;t_0) + \text{MtM}(t_0;t), 0) \\ &= \max(\text{MtM}(0;t_0), 0) + \\ &\quad (\max(\text{MtM}(0;t_0) + \text{MtM}(t_0;t), 0) - \max(\text{MtM}(0;t_0), 0)) \end{aligned}$$

The first component is the current exposure, which is always positive:

$$\text{CE}(t_0) = \max(\text{MtM}(0;t_0), 0)$$

The second component is the credit variation between t_0 and t. While the current mark-to-market value is negative, the second component can only be a positive value. However, the credit variation may also be negative if the future mark-to-market value is negative. Let us denote by $\mathbf{F}_{[0,t]}$ the cumulative distribution function of the potential future exposure $e(t)$. The peak exposure (PE) is the quantile of the counterparty exposure at the confidence level α:

$$\begin{aligned} \text{PE}_\alpha(t) &= \mathbf{F}_{[0,t]}^{-1}(\alpha) \\ &= \{\inf x : \Pr\{e(t) \leq x\} \geq \alpha\} \end{aligned} \tag{4.2}$$

The maximum value of the peak exposure is referred as the maximum peak exposure[7] (MPE):

$$\text{MPE}_\alpha(0;t) = \sup_s \text{PE}_\alpha(0;s) \tag{4.3}$$

We now introduce the traditional counterparty credit risk measures:

- The expected exposure (EE) is the average of the distribution of the counterparty exposure at the future date t:

$$\begin{aligned} \text{EE}(t) &= \mathbb{E}[e(t)] \\ &= \int_0^\infty x \, \mathrm{d}\mathbf{F}_{[0,t]}(x) \end{aligned} \tag{4.4}$$

- The expected positive exposure (EPE) is the weighted average over time $[0,t]$ of the expected exposure:

$$\begin{aligned} \text{EPE}(0;t) &= \mathbb{E}\left[\frac{1}{t}\int_0^t e(s)\,\mathrm{d}s\right] \\ &= \frac{1}{t}\int_0^t \text{EE}(s)\,\mathrm{d}s \end{aligned} \tag{4.5}$$

[6] The definitions introduced in this paragraph come from Canabarro and Duffie (2003) and the Basel II framework.

[7] It is also known as the maximum potential future exposure (MPFE).

- The effective expected exposure (EEE) is the maximum expected exposure that occurs at the future date t or any prior date:

$$\begin{aligned} \text{EEE}(t) &= \sup_{s \leq t} \text{EE}(s) \\ &= \max\left(\text{EEE}\left(t^-\right), \text{EE}(t)\right) \end{aligned} \tag{4.6}$$

- Finally, the effective expected positive exposure (EEPE) is the weighted average over time $[0, t]$ of the effective expected exposure:

$$\text{EEPE}(0;t) = \frac{1}{t}\int_0^t \text{EEE}(s)\ \mathrm{d}s \tag{4.7}$$

We can make several observations concerning the previous measures. Some of them are defined with respect to a future date t. This is the case of $\text{PE}_\alpha(t)$, $\text{EE}(t)$ and $\text{EEE}(t)$. The others depend on the time period $[0;t]$, typically a one-year time horizon. Previously, we have considered the counterparty measure $e(t)$, which defines the potential future exposure between the initial date 0 and the future date t. We can also use other credit measures like the potential future exposure between the current date t_0 and the future date t:

$$e(t) = \max(\text{MtM}(t_0;t), 0)$$

The counterparty exposure $e(t)$ can be defined with respect to one contract or to a basket of contracts. In this last case, we have to take into account netting arrangements.

4.1.2.3 Practical implementation for calculating counterparty exposure

We consider again Example 44 and assume that the current date t_0 is the initial date $t = 0$. Using 50 000 simulations, we have calculated the different credit measures with respect to the time t and reported them in Figure 4.4. For that, we have used the risk-neutral distribution probability $\mathbb{Q}$ in order to simulate the trajectory of the asset price $S(t)$. Let $\{t_0, t_1, \ldots, t_n\}$ be the set of discrete times. We note n_S the number of simulations and $S_j(t_i)$ the value of the asset price at time t_i for the j^{th} simulation. For each simulated trajectory, we then calculate the option price $\boldsymbol{\mathcal{C}}_j(t_i)$ and the mark-to-market value:

$$\text{MtM}_j(t_i) = n_C \cdot (\boldsymbol{\mathcal{C}}_j(t_i) - \boldsymbol{\mathcal{C}}_0)$$

Therefore, we deduce the potential future exposure:

$$e_j(t_i) = \max(\text{MtM}_j(t_i), 0)$$

The peak exposure at time t_i is estimated using the order statistics:

$$\text{PE}_\alpha(t_i) = e_{\alpha n_S:n_S}(t_i) \tag{4.8}$$

We use the empirical mean to calculate the expected exposure:

$$\text{EE}(t_i) = \frac{1}{n_S}\sum_{j=1}^{n_S} e_j(t_i) \tag{4.9}$$

For the expected positive exposure, we approximate the integral by the following sum:

$$\text{EPE}(0;t_i) = \frac{1}{t_i}\sum_{k=1}^{i} \text{EE}(t_k)\,\Delta t_k \tag{4.10}$$

If we consider a fixed-interval scheme with $\Delta t_k = \Delta t$, we obtain:

$$\begin{aligned} \text{EPE}(0; t_i) &= \frac{\Delta t}{t_i} \sum_{k=1}^{i} \text{EE}(t_k) \\ &= \frac{1}{i} \sum_{k=1}^{i} \text{EE}(t_k) \end{aligned} \tag{4.11}$$

By definition, the effective expected exposure is given by the following recursive formula:

$$\text{EEE}(t_i) = \max(\text{EEE}(t_{i-1}), \text{EE}(t_i)) \tag{4.12}$$

where $\text{EEE}(0)$ is initialized with the value $\text{EE}(0)$. Finally, the effective expected positive exposure is given by:

$$\text{EEPE}(0; t_i) = \frac{1}{t_i} \sum_{k=1}^{i} \text{EEE}(t_k) \Delta t_k \tag{4.13}$$

In the case of a fixed-interval scheme, this formula becomes:

$$\text{EEPE}(0; t_i) = \frac{1}{i} \sum_{k=1}^{i} \text{EEE}(t_k) \tag{4.14}$$

If we consider Figure 4.4, we observe that the counterparty exposure is increasing with respect to the time horizon[8]. This property is due to the fact that the credit risk evolves according to a square-root-of-time rule $\sqrt{t}$. In the case of an interest rate swap, the counterparty exposure takes the form of a bell-shaped curve. In fact, there are two opposite effects that determine the counterparty exposure (Pykhtin and Zhu, 2007):

- the diffusion effect of risk factors increases the counterparty exposure over time, because the uncertainty is greater in the future and may produce very large potential future exposures compared to the current exposure;
- the amortization effect decreases the counterparty exposure over time, because it reduces the remaining cash flows that are exposed to default.

In Figure 4.5, we have reported counterparty exposure in the case of an interest swap with a continuous amortization. The peak exposure initially increases because of the diffusion effect and generally reaches its maximum at one-third of the remaining maturity. It then decreases because of the amortization effect. This is why it is equal to zero at the maturity date when the swap is fully amortized.

4.1.3 Regulatory capital

The Basel II Accord includes three approaches to calculate the capital requirement for the counterparty credit risk: current exposure method (CEM), standardized method (SM) and internal model method (IMM). In March 2014, the Basel Committee decided to replace non-internal model approaches (CEM and SM) by a more sensitive approach called standardized approach (or SA-CCR), which is has been implemented since January 2017.

Each approach defines how the exposure at default EAD is calculated. The bank uses this estimate with the appropriated credit approach (SA or IRB) in order to measure the capital requirement. In the SA approach, the capital charge is equal to:

$$\mathcal{K} = 8\% \cdot \text{EAD} \cdot \text{RW}$$

[8] This implies that $\text{MPE}_\alpha(0; t) = \text{PE}_\alpha(t)$ and $\text{EEE}(t) = \text{EE}(t)$.

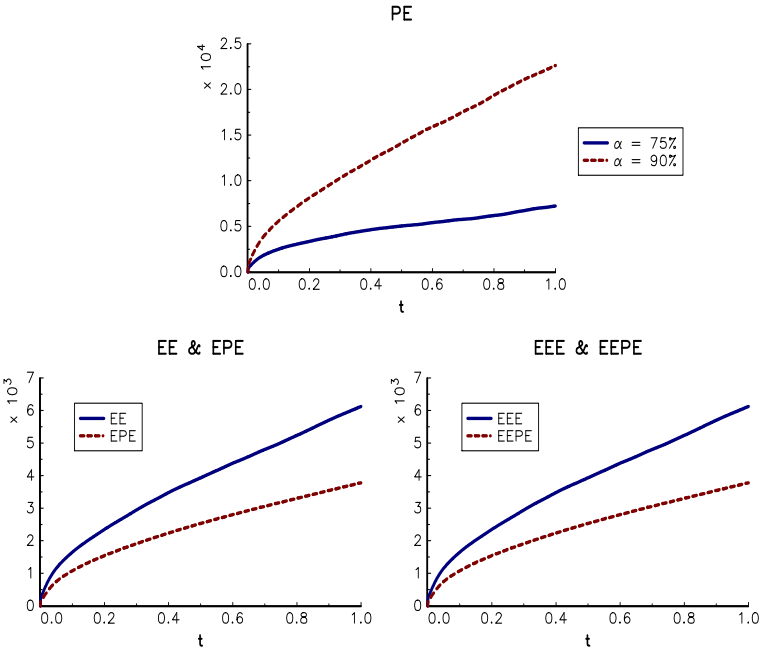


FIGURE 4.4: Counterparty exposure profile of options

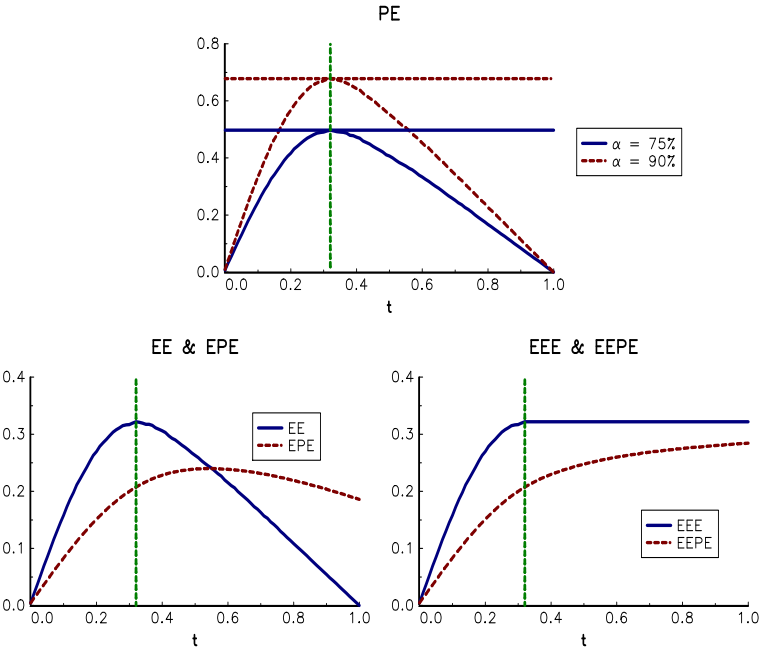


FIGURE 4.5: Counterparty exposure profile of interest rate swaps

where RW is the risk weight of the counterparty. In the IRB approach, we recall that:

$$\mathcal{K} = \text{EAD} \cdot \text{LGD} \cdot \left(\Phi \left(\frac{\Phi^{-1}(\text{PD}) + \sqrt{\rho(\text{PD})}\Phi^{-1}(0.999)}{\sqrt{1-\rho(\text{PD})}} \right) - \text{PD} \right) \cdot \varphi(\text{M})$$

where LGD and PD are the loss given default and the probability of default, which apply to the counterparty. The correlation $\rho(\text{PD})$ is calculated using the standard formula (3.35) given on page 184.

4.1.3.1 Internal model method

In the internal model method, the exposure at default is calculated as the product of a scalar α and the one-year effective expected positive exposure[9]:

$$\text{EAD} = \alpha \cdot \text{EEPE}(0;1)$$

The Basel Committee has set the value α at 1.4. The maturity M used in the IRB formula is equal to one year if the remaining maturity is less or equal than one year. Otherwise, it is calculated as follows[10]:

$$\text{M} = \min\left(1 + \frac{\sum_{k=1} \mathbb{1}\{t_k > 1\} \cdot \text{EE}(t_k) \cdot \Delta t_k \cdot B_0(t_k)}{\sum_{k=1} \mathbb{1}\{t_k \leq 1\} \cdot \text{EEE}(t_k) \cdot \Delta t_k \cdot B_0(t_k)}, 5\right)$$

Under some conditions, the bank may uses its own estimates for α. Let LEE be the loan equivalent exposure such that:

$$\mathcal{K}(\text{LEE} \cdot \text{LGD} \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\}) = \mathcal{K}(\text{EAD}(\boldsymbol{\tau}) \cdot \text{LGD} \cdot \mathbb{1}\{\boldsymbol{\tau} \leq T\}) \tag{4.15}$$

The loan equivalent exposure is then the deterministic exposure at default, which gives the same capital than the random exposure at default $\text{EAD}(\boldsymbol{\tau})$. Using a one-factor credit risk model, Canabarro *et al.* (2003) showed that:

$$\alpha = \frac{\text{LEE}}{\text{EPE}}$$

This is the formula that banks must use in order to estimate α, subject to a floor of 1.2.

Example 45 *We assume that the one-year effective expected positive exposure with respect to a given counterparty is equal to* \$50.2 *mn.*

In Table 4.4, we have reported the required capital $\mathcal{K}$ for different values of PD under the foundation IRB approach. The maturity M is equal to one year and we consider the 45% supervisory factor for the loss given default. The exposure at default is calculated with $\alpha = 1.4$. We show the impact of the Basel III multiplier applied to the correlation. In this example, if the default probability of the counterparty is equal to 1%, this induces an additional required capital of 27.77%.

[9] If the remaining maturity τ of the product is less than one year, the exposure at default becomes:

$$\text{EAD} = \alpha \cdot \text{EEPE}(0;\tau)$$

[10] The maturity has then a cap of five years.

TABLE 4.4: Capital charge of counterparty credit risk under the FIRB approach

	PD	1%	2%	3%	4%	5%
Basel II	$\rho\,(\text{PD})$ (in %)	19.28	16.41	14.68	13.62	12.99
	$\mathcal{K}$ (in \$ mn)	4.12	5.38	6.18	6.82	7.42
Basel III	$\rho\,(\text{PD})$ (in %)	24.10	20.52	18.35	17.03	16.23
	$\mathcal{K}$ (in \$ mn)	5.26	6.69	7.55	8.25	8.89
	$\Delta\mathcal{K}$ (in %)	27.77	24.29	22.26	20.89	19.88

4.1.3.2 Non-internal model methods (Basel II)

Under the current exposure method (CEM), we have:

$$\text{EAD} = \text{CE}\,(0) + A$$

where $\text{CE}\,(0)$ is the current exposure and A is the add-on value. In the views of the Basel Committee, $\text{CE}\,(0)$ represents the replacement cost, whereas the add-on reflects the potential future exposure of the contract. For a single OTC transaction, A is the product of the notional and the add-on factor, which is given in Table 4.5. For a portfolio of OTC transactions with netting agreements, the exposure at default is the sum of the current net exposure plus a net add-one value A_N, which is defined as follows:

$$A_N = (0.4 + 0.6 \cdot \text{NGR}) \cdot A_G$$

where $A_G = \sum_i A_i$ is the gross add-on, A_i is the add-on of the i^{th} transaction and NGR is the ratio between the current net and gross exposures.

TABLE 4.5: Regulatory add-on factors for the current exposure method

Residual Maturity	Fixed Income	FX and Gold	Equity	Precious Metals	Other Commodities
0−1Y	0.0%	1.0%	8.0%	7.0%	10.0%
1Y−5Y	0.5%	5.0%	8.0%	7.0%	12.0%
5Y+	1.5%	7.5%	10.0%	8.0%	15.0%

Example 46 *We consider a portfolio of four OTC derivatives, which are traded with the same counterparty:*

Contract	$\mathfrak{C}_1$	$\mathfrak{C}_2$	$\mathfrak{C}_3$	$\mathfrak{C}_4$
Asset class	*Fixed income*	*Fixed income*	*Equity*	*Equity*
Notional (in \$ mn)	100	40	20	10
Maturity	2Y	6Y	6M	18M
Mark-to-market (in \$ mn)	3.0	−2.0	2.0	−1.0

We assume that there are two netting arrangements: one concerning fixed income derivatives and another one for equity derivatives.

In the case where there is no netting agreement, we obtain these results:

Contract	$\mathfrak{C}_1$	$\mathfrak{C}_2$	$\mathfrak{C}_3$	$\mathfrak{C}_4$	Sum
$\text{CE}\,(0)$ (in \$ mn)	3.0	0.0	2.0	0.0	5.0
Add-on (in %)	0.5	1.5	8.0	8.0	
A (in \$ mn)	0.5	0.6	1.6	0.8	3.5

The exposure at default is then equal to \$8.5 mn. If we take into account the two netting agreements, the current net exposure becomes:

$$\mathrm{CE}(0) = \max(3-2,0) + \max(2-1,0) = \$2 \text{ mn}$$

We deduce that NGR is equal to 2/5 or 40%. It follows that:

$$A_N = (0.4 + 0.6 \times 0.4) \times 3.5 = \$2.24 \text{ mn}$$

Finally, the exposure at default is equal to \$4.24 mn.

The standardized method was designed for banks that do not have the approval to apply the internal model method, but would like to have a more sensitive approach that the current exposure method. In this framework, the exposure at default is equal to:

$$\mathrm{EAD} = \beta \cdot \max\left(\sum_i \mathrm{CMV}_i, \sum_j \mathrm{CCF}_j \cdot \left|\sum_{i \in j} \mathrm{RPT}_i\right|\right)$$

where CMV_i is the current market value of transaction i, CCF_j is the supervisory credit conversion factor with respect to the hedging set j and RPT_i is the risk position from transaction i. The supervisory scaling factor β is set to 1.4. In this approach, the risk positions have to be grouped into hedging sets, which are defined by similar instruments (e.g. same commodity, same issuer, same currency, etc.). The risk position $\sum_{i \in j} \mathrm{RPT}_i$ is the sum of notional values of linear instruments and delta-equivalent notional values of non-linear instruments, which belong to the hedging set j. The credit conversion factors ranges from 0.3% to 10%. The initial goal of the Basel Committee was to provide an approach which mimics the internal model method[11]. However, the SM approach was never really used by banks. Indeed, it didn't interest advanced banks that preferred to implement the IMM, and it was too complicated for the other banks that have used the CEM.

4.1.3.3 SA-CCR method (Basel III)

The SA-CCR has been adopted by the Basel Committee in March 2014 in order to replace non-internal models approaches since January 2017. The main motivation the Basel Committee was to propose a more-sensitive approach, which can easily be implemented:

> "*Although being more risk-sensitive than the CEM, the SM was also criticized for several weaknesses. Like the CEM, it did not differentiate between margined and unmargined transactions or sufficiently capture the level of volatilities observed over stress periods in the last five years. In addition, the definition of hedging set led to operational complexity resulting in an inability to implement the SM, or implementing it in inconsistent ways*" (BCBS, 2014b, page 1).

The exposure at default under the SA-CCR is defined as follows:

$$\mathrm{EAD} = \alpha \cdot (\mathrm{RC} + \mathrm{PFE})$$

where RC is the replacement cost (or the current exposure), PFE is the potential future exposure and α is equal to 1.4. We can view this formula as an approximation of the IMM calculation, meaning that RC + PFE represents a stylized EEPE value. The PFE add-on is given by:

$$\mathrm{PFE} = \gamma \cdot \sum_{q=1}^{5} A^{(\mathcal{C}_q)}$$

[11] Indeed, the β multiplier coefficient is the equivalent of the α multiplier coefficient, whereas the rest of the expression can be interpreted as an estimate of the effective expected positive exposure.

where γ is the multiplier and $A^{(\mathcal{C}_q)}$ is the add-on of the asset class $\mathcal{C}_q$ (interest rate, foreign exchange, credit, equity and commodity). We have:

$$\gamma = \min\left(1, 0.05 + 0.95 \cdot \exp\left(\frac{\text{MtM}}{1.90 \cdot \sum_{q=1}^{5} A^{(\mathcal{C}_q)}}\right)\right)$$

where MtM is the mark-to-market value of the derivative transactions minus the haircut value of net collateral held. We notice that γ is equal to 1 when the mark-to-market is positive and $\gamma \in [5\%, 1]$ when the net mark-to-market is negative. Figure 4.6 shows the relationship between the ratio $\text{MtM} \Big/ \sum_{q=1}^{5} A^{(\mathcal{C}_q)}$ and the multiplier γ. The role of γ is then to reduce the potential future exposure in the case of negative mark-to-market.

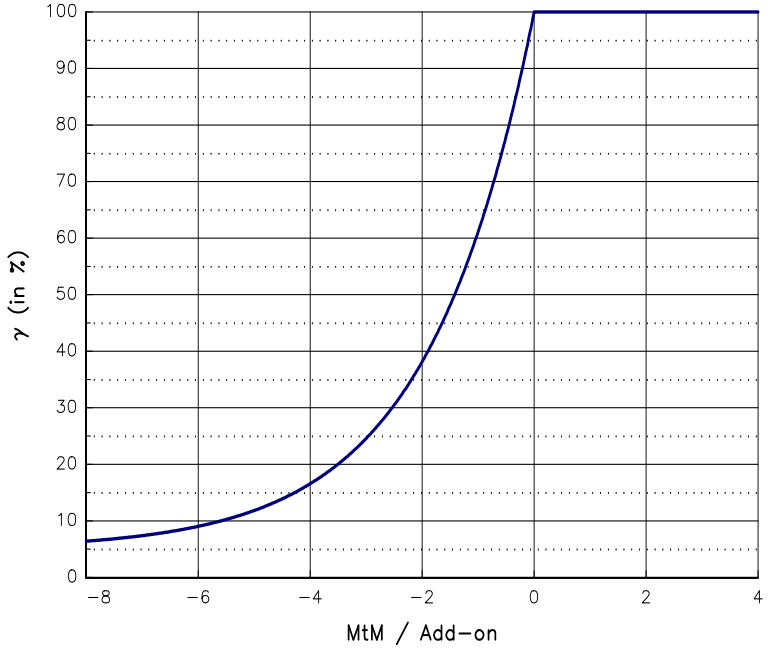


FIGURE 4.6: Impact of negative mark-to-market on the PFE multiplier

The general steps for calculating the add-on are the following. First, we have to determine the primary risk factors of each transaction in order to classify the transaction into one or more asset classes. Second, we calculate an adjusted notional amount d_i at the transaction level[12] and a maturity factor $\mathcal{MF}_i$, which reflects the time horizon appropriate for this type of transactions. For unmargined transactions, we have:

$$\mathcal{MF}_i = \sqrt{\min\left(\text{M}_i, 1\right)}$$

[12]The trade-level adjusted notional d_i is defined as the product of current price of one unit and the number of units for equity and commodity derivatives, the notional of the foreign currency leg converted to domestic currency for foreign exchange derivatives and the product of the trade notional amount and the supervisory duration $\mathcal{SD}_i$ for interest rate and credit derivatives. The supervisory duration $\mathcal{SD}_i$ is defined as follows:

$$\mathcal{SD}_i = 20 \cdot \left(e^{-0.05 \cdot S_i} - e^{-0.05 \cdot E_i}\right)$$

where S_i and E_i are the start and end dates of the time period referenced by the derivative instrument.

where M_i is the remaining maturity of the transaction and is floored by 10 days. For margined transactions, we have:

$$\mathcal{MF}_i = \frac{3}{2}\sqrt{\mathrm{M}_i^\star}$$

where $\mathrm{M}_i^\star$ is the appropriate margin period of risk (MPOR). Then, we apply a supervisory delta adjustment $\boldsymbol{\Delta}_i$ to each transaction[13] and a supervisory factor $\mathcal{SF}_j$ to each hedging set j in order to take the volatility into account. The add-on of one transaction i has then the following expression:

$$A_i = \mathcal{SF}_j \cdot (\boldsymbol{\Delta}_i \cdot d_i \cdot \mathcal{MF}_i)$$

Finally, we apply an aggregation method to calculate the add-on $A^{(\mathcal{C}_q)}$ of the asset class $\mathcal{C}_q$ by considering correlations between hedging sets. Here are the formulas that determine the add-on values:

- The add-on for interest rate derivatives is equal to:

$$A^{(\mathrm{ir})} = \sum_j \mathcal{SF}_j \cdot \sqrt{\sum_{k=1}^{3}\sum_{k'=1}^{3} \rho_{k,k'} \cdot D_{j,k} \cdot D_{j,k'}}$$

  where notations j and k refer to currency j and maturity bucket[14] k and the effective notional $D_{j,k}$ is calculated according to:

$$D_{j,k} = \sum_{i \in (j,k)} \boldsymbol{\Delta}_i \cdot d_i \cdot \mathcal{MF}_i$$

- For foreign exchange derivatives, we obtain:

$$A^{(\mathrm{fx})} = \sum_j \mathcal{SF}_j \cdot \left| \sum_{i \in j} \boldsymbol{\Delta}_i \cdot d_i \cdot \mathcal{MF}_i \right|$$

  where the hedging set j refers to currency pair j.

- The add-on for credit and equity derivatives use the same formula:

$$A^{(\mathrm{credit/equity})} = \sqrt{\left(\sum_k \rho_k \cdot A_k\right)^2 + \sum_k (1 - \rho_k^2) \cdot A_k^2}$$

  where k represents entity k and:

$$A_k = \mathcal{SF}_k \cdot \sum_{i \in k} \boldsymbol{\Delta}_i \cdot d_i \cdot \mathcal{MF}_i$$

- In the case of commodity derivatives, we have:

$$A^{(\mathrm{commodity})} = \sum_j \sqrt{\left(\rho_j \cdot \sum_k A_{j,k}\right)^2 + \left(1 - \rho_j^2\right) \cdot \sum_k A_{j,k}^2}$$

  where j indicates the hedging set, k corresponds to the commodity type and:

$$A_{j,k} = \mathcal{SF}_{j,k} \cdot \sum_{i \in (j,k)} \boldsymbol{\Delta}_i \cdot d_i \cdot \mathcal{MF}_i$$

[13] For instance $\boldsymbol{\Delta}_i$ is equal to -1 for a short position, $+1$ for a long position, the Black-Scholes delta for an option position, etc.

[14] The three maturity buckets k are (1) less than one year, (2) between one and five years and (3) more than five years.

TABLE 4.6: Supervisory parameters for the SA-CCR approach

Asset class		$\mathcal{SF}_j$	ρ_k			Σ_i
Interest rate	0−1Y	0.50%	100%			50%
	1Y−5Y	0.50%	70%	100%		50%
	5Y+	0.50%	30%	70%	100%	50%
Foreign exchange		4.00%				15%
Credit	AAA	0.38%		50%		100%
	AA	0.38%		50%		100%
	A	0.42%		50%		100%
	BBB	0.54%		50%		100%
	BB	1.06%		50%		100%
	B	1.60%		50%		100%
	CCC	6.00%		50%		100%
	IG index	0.38%		80%		80%
	SG index	1.06%		80%		80%
Equity	Single name	32.00%		50%		120%
	Index	20.00%		80%		75%
Commodity	Electricity	40.00%		40%		150%
	Oil & gas	18.00%		40%		70%
	Metals	18.00%		40%		70%
	Agricultural	18.00%		40%		70%
	Other	18.00%		40%		70%

Source: BCBS (2014b).

For interest rate derivatives, hedging sets correspond to all derivatives in the same currency (e.g. USD, EUR, JPY). For currency, they consists of all currency pairs (e.g. USD/EUR, USD/JPY, EUR/JPY). For credit and equity, there is a single hedging set, which contains all the entities (both single names and indices). Finally, there are four hedging sets for commodity derivatives: energy (electricity, oil & gas), metals, agricultural and other. In Table 4.6, we give the supervisory parameters[15] for the factor $\mathcal{SF}_j$, the correlation[16] ρ_k and the implied volatility Σ_i in order to calculate Black-Scholes delta exposures. We notice that the value of the supervisory factor can differ within one hedging set. For instance, it is equal to 0.38% for investment grade (IG) indices, while it takes the value 1.06% for speculative grade (SG) indices.

Example 47 *The netting set consists of four interest rate derivatives*[17]*:*

Trade	*Instrument*	*Currency*	*Maturity*	*Swap*	*Notional*	MtM
1	*IRS*	*USD*	*9M*	*Payer*	4	0.10
2	*IRS*	*USD*	*4Y*	*Receiver*	20	−0.20
3	*IRS*	*USD*	*10Y*	*Payer*	20	0.70
4	*Swaption 10Y*	*USD*	*1Y*	*Receiver*	5	0.50

This netting set consists of only one hedging set, because the underlying assets of all these derivative instruments are USD interest rates. We report the different calculations in

[15] Source: BCBS (2014b).
[16] We notice that we consider cross-correlations between the three time buckets for interest rate derivatives.
[17] For the swaption, the forward rate swap and the strike value are equal to 6% and 5%.

the following table:

i	k	S_i	E_i	$\mathcal{SD}_i$	Δ_i	d_i	$\mathcal{MF}_i$	D_i
1	1	0.00	0.75	0.74	1.00	2.94	0.87	2.55
2	2	0.00	4.00	3.63	−1.00	72.51	1.00	−72.51
3	3	0.00	10.00	7.87	1.00	157.39	1.00	157.39
4	3	1.00	11.00	7.49	−0.27	37.43	1.00	−10.08

where k indicates the time bucket, S_i is the start date, E_i is the end date, $\mathcal{SD}_i$ is the supervisory duration, $\boldsymbol{\Delta}_i$ is the delta, d_i is the adjusted notional, $\mathcal{MF}_i$ is the maturity factor and D_i is the effective notional. For instance, we obtain the following results for the swaption transaction:

$$\begin{aligned}
\mathcal{SD}_i &= 20 \times \left(e^{-0.05\times 1} - e^{-0.05\times 10}\right) = 7.49 \\
\boldsymbol{\Delta}_i &= -\Phi\left(-\frac{\ln(6\%/5\%)}{0.5\times\sqrt{1}} + \frac{1}{2}\times 0.5\times\sqrt{1}\right) = -0.27 \\
d_i &= 7.49 \times 5 = 37.43 \\
\mathcal{MF}_i &= \sqrt{1} = 1 \\
D_i &= -0.27 \times 37.43 \times 1 = -10.08
\end{aligned}$$

We deduce that the effective notional of time buckets is respectively equal to $D_1 = 2.55$, $D_2 = -72.51$ and $D_3 + D_4 = 147.30$. It follows that:

$$\begin{aligned}
\sum_{k=1}^{3}\sum_{k'=1}^{3}\rho_{k,k'}D_{j,k}D_{j,k'} &= 2.55^2 - 2\times 70\%\times 2.55\times 72.51 + \\
&\quad 72.51^2 - 2\times 70\%\times 72.51\times 147.30 + \\
&\quad 147.30^2 + 2\times 30\%\times 2.55\times 147.30 \\
&= 11\,976.1
\end{aligned}$$

While the supervisory factor is 0.50%, the add-on value $A^{(\text{ir})}$ is then equal to 0.55. The replacement cost is:

$$\text{RC} = \max(0.1 - 0.2 + 0.7 + 0.5, 0) = 1.1$$

Because the mark-to-market of the netting set is positive, the PFE multiplier is equal to 1. We finally deduce that:

$$\text{EAD} = 1.4 \times (1.1 + 1 \times 0.55) = 2.31$$

Remark 51 *Annex 4 of BCBS (2014b) contains four examples of SA-CCR calculations and presents also several applications including different hedging sets, netting sets and asset classes.*

Even if SA-CCR is a better approach for measuring the counterparty credit risk than CEM and SM, its conservative calibration has been strongly criticized, in particular the value of α. For instance, the International Swaps and Derivatives Association reports many examples, where the EAD calculated with SA-CCR is a multiple of the EAD calculated with CEM and IMM[18]. This is particularly true when the mark-to-market is negative and the hedging set is unmargined. In fact, the industry considers that $\alpha \approx 1$ is more appropriate than $\alpha = 1.4$.

[18] www.isda.org/a/qTiDE/isda-letter-to-the-bcbs-on-sa-ccr-march-2017.pdf

4.1.4 Impact of wrong way risk

According to ISDA (2014b), the wrong way risk (WWR) is defined as the risk that "*occurs when exposure to a counterparty or collateral associated with a transaction is adversely correlated with the credit quality of that counterparty*". This means that the exposure at default of the OTC contract and the default risk of the counterparty are not independent, but positively correlated. Generally, we distinguish two types of wrong way risk:

1. general (or conjectural) wrong way risk occurs when the credit quality of the counterparty is correlated with macroeconomic factors, which also impact the value of the transaction;

2. specific wrong way risk occurs when the correlation between the exposure at default and the probability of default is mainly explained by some idiosyncratic factors.

For instance, general WWR arises when the level of interest rates both impacts the mark-to-market of the transaction and the creditworthiness of the counterparty. An example of specific WWR is when Bank A buys a CDS protection on Bank B from Bank C, and the default probabilities of B and C are highly correlated. In this case, if the credit quality of B deteriorates, both the mark-to-market of the transaction and the default risk of C increase.

Remark 52 *Right way risk (RWR) corresponds to the situation where the counterparty exposure and the default risk are negatively correlated. In this case, the mark-to-market of the transaction decreases as the counterparty approaches the default. By definition, RWR is less a concern from a regulation point of view.*

4.1.4.1 An example

Let us assume that the mark-to-market of the OTC contract is given by a Brownian motion:

$$\mathrm{MtM}(t) = \mu + \sigma W(t)$$

If we note $e(t) = \max(\mathrm{MtM}(t), 0)$, we have:

$$\begin{aligned}
\mathbb{E}[e(t)] &= \int_{-\infty}^{\infty} \max\left(\mu + \sigma\sqrt{t}x, 0\right) \phi(x)\, \mathrm{d}x \\
&= \mu \int_{-\mu/(\sigma\sqrt{t})}^{\infty} \phi(x)\, \mathrm{d}x + \sigma\sqrt{t} \int_{-\mu/(\sigma\sqrt{t})}^{\infty} x\phi(x)\, \mathrm{d}x \\
&= \mu\left(1 - \Phi\left(-\frac{\mu}{\sigma\sqrt{t}}\right)\right) + \sigma\sqrt{t}\left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right]_{-\mu/(\sigma\sqrt{t})}^{\infty} \\
&= \mu\Phi\left(\frac{\mu}{\sigma\sqrt{t}}\right) + \sigma\sqrt{t}\phi\left(\frac{\mu}{\sigma\sqrt{t}}\right)
\end{aligned}$$

We consider the Merton approach for modeling the default time $\boldsymbol{\tau}$ of the counterparty. Let $B(t) = \Phi^{-1}(1 - \mathbf{S}(t))$ be the default barrier, where $\mathbf{S}(t)$ is the survival function of the counterparty. We assume that the dependence between the mark-to-market $\mathrm{MtM}(t)$ and the survival time is equal to the Normal copula $\mathbf{C}(u_1, u_2; \rho)$ with parameter ρ. Redon (2006)

shows that[19]:

$$\begin{aligned}\mathbb{E}\left[e\left(t\right) \mid \boldsymbol{\tau} = t\right] &= \mathbb{E}\left[e\left(t\right) \mid B\left(t\right) = B\right] \\ &= \mu_B \Phi\left(\frac{\mu_B}{\sigma_B}\right) + \sigma_B \phi\left(\frac{\mu_B}{\sigma_B}\right)\end{aligned}$$

where $\mu_B = \mu + \rho\sigma\sqrt{t}B$ and $\sigma_B = \sqrt{1-\rho^2}\sigma\sqrt{t}$. With the exception of $\rho = 0$, we have:

$$\mathbb{E}\left[e\left(t\right)\right] \neq \mathbb{E}\left[e\left(t\right) \mid \boldsymbol{\tau} = t\right]$$

In Figure 4.7, we report the conditional distribution of the mark-to-market given that the default occurs at time $t = 1$. The parameters are $\mu = 0$, $\sigma = 1$ and $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$ where λ is calibrated to fit the one-year probability of default PD[20]. We notice that the exposure at default decreases with the correlation ρ when PD is equal to 1% (top/left panel), whereas it increases with the correlation ρ when PD is equal to 99% (top/right panel). We verify the stochastic dominance of the mark-to-market with respect to the default probability. Figure 4.8 shows the relationship between the conditional expectation $\mathbb{E}\left[e\left(t\right) \mid \boldsymbol{\tau} = t\right]$ and the different parameters[21]. As expected, the exposure at default is an increasing function of μ, σ, ρ and PD.

4.1.4.2 Calibration of the α factor

In the internal model method, the exposure at default is computed by scaling the effective expected positive exposure:

$$\text{EAD} = \alpha \cdot \text{EEPE}\left(0;1\right)$$

where α is the scaling factor. In this framework, we assume that the mark-to-market of the OTC transaction and the default risk of the counterparty are not correlated. Therefore, the Basel Committee requires that the calibration of the scaling factor α incorporates the general wrong way risk. According to BCBS (2006), we have[22]:

$$\alpha = \frac{\boldsymbol{\mathcal{K}}\left(\text{EAD}\left(\tau\right) \cdot \text{LGD} \cdot \mathbb{1}\left\{\tau \leq T\right\}\right)}{\boldsymbol{\mathcal{K}}\left(\text{EPE} \cdot \text{LGD} \cdot \mathbb{1}\left\{\tau \leq T\right\}\right)}$$

[19] Since we have $1 - \mathbf{S}\left(t\right) \sim \mathcal{U}_{[0,1]}$, it follows that $B\left(t\right) \sim \mathcal{N}\left(0,1\right)$. We deduce that the random vector $\left(\text{MtM}\left(t\right), B\left(t\right)\right)$ is normally distributed:

$$\begin{pmatrix} \text{MtM}\left(t\right) \\ B\left(t\right) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 t & \rho\sigma\sqrt{t} \\ \rho\sigma\sqrt{t} & 1 \end{pmatrix}\right)$$

because the correlation $\rho\left(\text{MtM}\left(t\right), B\left(t\right)\right)$ is equal to the Normal copula parameter ρ. Using the conditional expectation formula given on page 1062, it follows that:

$$\text{MtM}\left(t\right) \mid B\left(t\right) = B \sim \mathcal{N}\left(\mu_B, \sigma_B^2\right)$$

where:

$$\mu_B = \mu + \rho\sigma\sqrt{t}\left(B - 0\right)$$

and:

$$\sigma_B^2 = \sigma^2 t - \rho^2\sigma^2 t = \left(1 - \rho^2\right)\sigma^2 t$$

[20] We have $1 - e^{-\lambda} = \text{PD}$.

[21] The default values are $\mu = 0$, $\sigma = 1$, PD = 90% and $\rho = 50\%$.

[22] Using standard assumptions (single factor model, fined-grained portfolio, etc.), the first-order approximation is:

$$\alpha \approx \frac{\text{LEE}}{\text{EPE}}$$

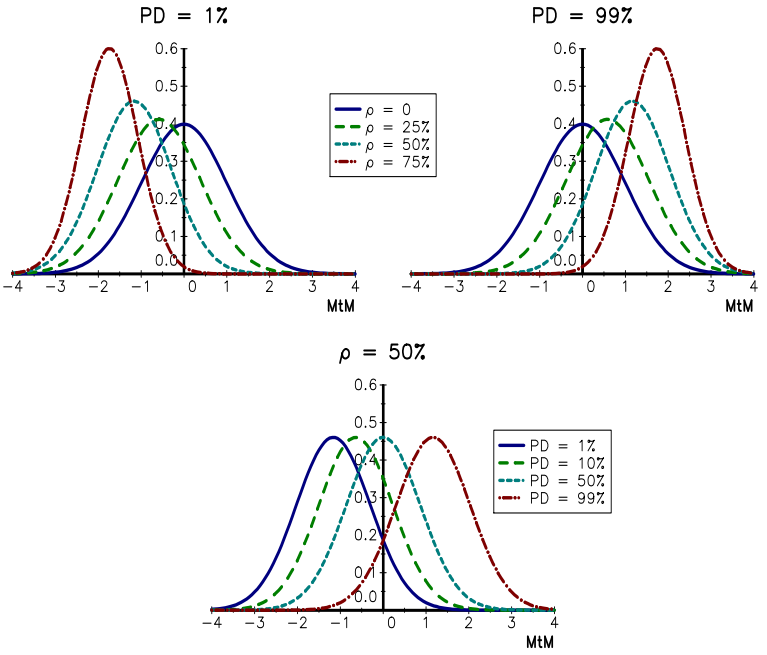


FIGURE 4.7: Conditional distribution of the mark-to-market

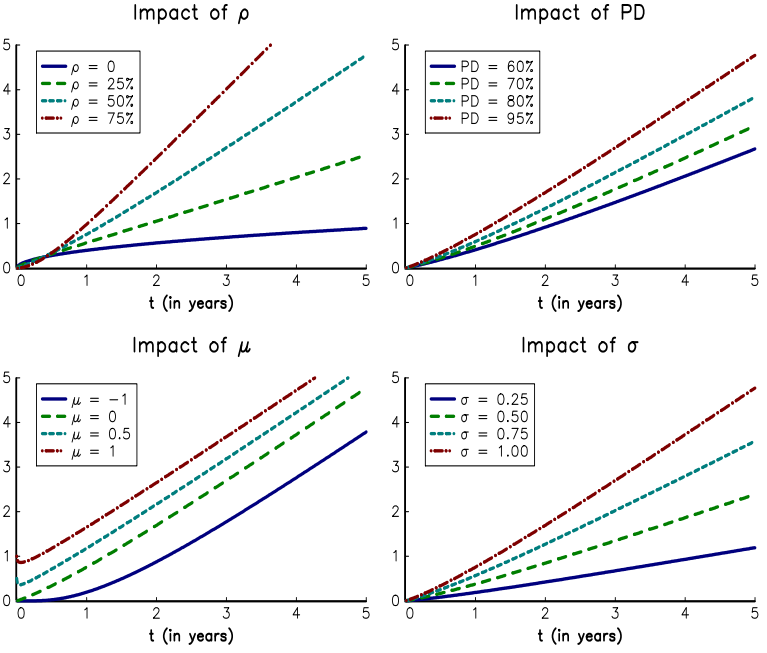


FIGURE 4.8: Conditional expectation of the exposure at default

Again, the Basel Committee considers a conservative approach, since they use EPE instead of EEPE for defining the denominator of α.

The calibration of α for a bank portfolio is a difficult task, because it is not easy to consider a joint modeling of market and credit risk factors. Let us write the portfolio loss as follows:

$$L = \sum_{i=1}^{n} \text{EAD}\left(\boldsymbol{\tau}_i, \mathcal{F}_1, \ldots, \mathcal{F}_m\right) \cdot \text{LGD}_i \cdot \mathbb{1}\left\{\boldsymbol{\tau}_i \leq T_i\right\}$$

where $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_m)$ are the market risk factors and $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ are the default times. Wrong way risk implies to correlate the random vectors $\mathcal{F}$ and $\boldsymbol{\tau}$. Given a small portfolio with a low number of transactions and counterparty entities, we can simulate the portfolio loss and calculate the corresponding α, but this Monte Carlo exercise is unrealistic for a comprehensive bank portfolio. Nevertheless, we can estimate α for more or less canonical portfolios. For instance, according to Cespedes *et al.* (2010), the scaling factor α may range from 0.7 to 1.4. When market and credit risks are uncorrelated, α is close to one. α is less than one for general right way risks, while it is larger than one for general wrong way risks. However, for realistic market-credit correlations, α is below 1.2.

Remark 53 *The treatment of specific wrong way risk is different. First, the bank must identify all the counterparty entities where specific WWR is significant, and monitor these operations. Second, the bank must calculate a conservative* EAD *figure.*

Remark 54 *The modeling of wrong way risk implies to correlate market and credit risk factors. The main approach is to specify a copula model. As the dimension of the problem is high (m risk factors and n counterparties), Cespedes et al. (2010) propose to consider a resampling approach. Another way is to relate the hazard rate of survival functions with the value of the contract (Hull and White, 2012). These two approaches will be discussed in the next section.*

4.2 Credit valuation adjustment

CVA is the adjustment to the risk-free (or fair) value of derivative instruments to account for counterparty credit risk. Thus, CVA is commonly viewed as the market price of CCR. The concept of CVA was popularized after the 2008 Global Financial Crisis, even if investments bank started to use CVA in the early 1990s (Litzenberger, 1992; Duffie and Huang, 1996). Indeed, during the global financial crisis, banks suffered significant counterparty credit risk losses on their OTC derivatives portfolios. However, according to BCBS (2010), roughly two-thirds of these losses came from CVA markdowns on derivatives and only one-third were due to counterparty defaults. In a similar way, the Financial Service Authority concluded that CVA losses were five times larger than CCR losses for UK banks during the period 2007-2009. In this context, BCBS (2010) included CVA capital charge in the Basel III framework, whereas credit-related adjustments were introduced in the accounting standard IFRS 13 also called *Fair Value Measurement*[23]. Nevertheless, the complexity of CVA raises several issues (EBA, 2015a). This is why questions around the CVA are not stabilized and new standards are emerging, but they only provide partial answers.

[23] IFRS 13 was originally issued in May 2011 and became effective after January 2013.

4.2.1 Definition

4.2.1.1 Difference between CCR and CVA

In order to understand the credit valuation adjustment, it is important to make the distinction between CCR and CVA. CCR is the credit risk of OTC derivatives associated to the default of the counterparty, whereas CVA is the market risk of OTC derivatives associated to the credit migration of the two counterparties. This means that CCR occurs at the default time. On the contrary, CVA impacts the market value of OTC derivatives before the default time.

Let us consider an example with two banks A and B and an OTC contract $\mathfrak{C}$. The P&L $\Pi_{A|B}$ of Bank A is equal to:

$$\Pi_{A|B} = \text{MtM} - \text{CVA}_B$$

where MtM is the risk-free mark-to-market value of $\mathfrak{C}$ and CVA_B is the CVA with respect to Bank B. We assume that Bank A has traded the same contract with Bank C. It follows that:

$$\Pi_{A|C} = \text{MtM} - \text{CVA}_C$$

In a world where there is no counterparty credit risk, we have:

$$\Pi_{A|B} = \Pi_{A|C} = \text{MtM}$$

If we take into account the counterparty credit risk, the two P&Ls of the same contract are different because Bank A does not face the same risk:

$$\Pi_{A|B} \neq \Pi_{A|C}$$

In particular, if Bank A wants to close the two exposures, it is obvious that the contact $\mathfrak{C}$ with the counterparty B has more value than the contact $\mathfrak{C}$ with the counterparty C if the credit risk of B is lower than the credit risk of C. In this context, the notion of mark-to-market is complex, because it depends on the credit risk of the counterparties.

Remark 55 *If the bank does not take into account CVA to price its OTC derivatives, it does not face CVA risk. This situation is now marginal because of the accounting standards IFRS 13.*

4.2.1.2 CVA, DVA and bilateral CVA

Previously, we have defined the CVA as the market risk related to the credit risk of the counterparty. According to EBA (2015a), it should reflect today's best estimate of the potential loss on the OTC derivative due to the default of the counterparty. In a similar way, we can define the debit value adjustment (DVA) as the credit-related adjustment capturing the entity's own credit risk. In this case, DVA should reflect the potential gain on the OTC derivative due to the entity's own default. If we consider our previous example, the expression of the P&L becomes:

$$\Pi_{A|B} = \text{MtM} + \underbrace{\text{DVA}_A - \text{CVA}_B}_{\text{Bilateral CVA}}$$

The combination of the two credit-related adjustments is called the bivariate CVA. We then obtain the following cases:

1. if the credit risk of Bank A is lower than the credit risk of Bank B ($\text{DVA}_A < \text{CVA}_B$), the bilateral CVA of Bank A is negative and reduces the value of the OTC portfolio from the perspective of Bank A;

2. if the credit risk of Bank A is higher than the credit risk of Bank B ($\text{DVA}_A > \text{CVA}_B$), the bilateral CVA of Bank A is positive and increases the value of the OTC portfolio from the perspective of Bank A;

3. if the credit risk of Bank A is equivalent to the credit risk of Bank B, the bilateral CVA is equal to zero.

We notice that the DVA of Bank A is the CVA of Bank A from the perspective of Bank B:

$$\text{CVA}_A = \text{DVA}_A$$

We also have $\text{DVA}_B = \text{CVA}_B$, which implies that the P&L of Bank B is equal to:

$$\begin{aligned}\Pi_{B|A} &= -\text{MtM} + \text{DVA}_B - \text{CVA}_A \\ &= -\text{MtM} + \text{CVA}_B - \text{DVA}_A \\ &= -\Pi_{A|B}\end{aligned}$$

We deduce that the P&Ls of Banks A and B are coherent in the bilateral CVA framework as in the risk-free MtM framework. This is not true if we only consider the (unilateral or one-sided) CVA or DVA adjustment.

In order to define more precisely CVA and DVA, we introduce the following notations:

- The positive exposure $e^+(t)$ is the maximum between 0 and the risk-free mark-to-market:
$$e^+(t) = \max(\text{MtM}(t), 0)$$
This quantity was previously denoted by $e(t)$ and corresponds to the potential future exposure in the CCR framework.

- The negative exposure $e^-(t)$ is the difference between the risk-free mark-to-market and the positive exposure:
$$e^-(t) = \text{MtM}(t) - e^+(t)$$
We also have:
$$\begin{aligned}e^-(t) &= -\min(\text{MtM}(t), 0) \\ &= \max(-\text{MtM}(t), 0)\end{aligned}$$
The negative exposure is then the equivalent of the positive exposure from the perspective of the counterparty.

The credit value adjustment is the risk-neutral discounted expected value of the potential loss:

$$\text{CVA} = \mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}\{\boldsymbol{\tau}_B \leq T\} \cdot e^{-\int_0^{\boldsymbol{\tau}_B} r_t \,\mathrm{d}t} \cdot L\right]$$

where T is the maturity of the OTC derivative, $\boldsymbol{\tau}_B$ is the default time of Bank B and L is the counterparty loss:

$$L = (1 - \boldsymbol{\mathcal{R}}_B) \cdot e^+(\boldsymbol{\tau}_B)$$

Using usual assumptions[24], we obtain:

$$\text{CVA} = (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T B_0(t)\,\text{EpE}(t)\,\mathrm{d}\mathbf{F}_B(t) \tag{4.16}$$

[24]The default time and the discount factor are independent and the recovery rate is constant.

where $\mathrm{EpE}(t)$ is the risk-neutral discounted expected positive exposure:

$$\mathrm{EpE}(t) = \mathbb{E}^{\mathbb{Q}}\left[e^{+}(t)\right]$$

and $\mathbf{F}_B$ is the cumulative distribution function of $\boldsymbol{\tau}_B$. Knowing that the survival function $\mathbf{S}_B(t)$ is equal to $1 - \mathbf{F}_B(t)$, we deduce that:

$$\mathrm{CVA} = (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T -B_0(t)\,\mathrm{EpE}(t)\,\mathrm{d}\mathbf{S}_B(t) \tag{4.17}$$

In a similar way, the debit value adjustment is defined as the risk-neutral discounted expected value of the potential gain:

$$\mathrm{DVA} = \mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}\{\boldsymbol{\tau}_A \leq T\} \cdot e^{-\int_0^{\boldsymbol{\tau}_A} r_t\,\mathrm{d}t} \cdot G\right]$$

where $\boldsymbol{\tau}_A$ is the default time of Bank A and:

$$G = (1 - \boldsymbol{\mathcal{R}}_A) \cdot e^{-}(\boldsymbol{\tau}_A)$$

Using the same assumptions than previously, it follows that:

$$\mathrm{DVA} = (1 - \boldsymbol{\mathcal{R}}_A) \cdot \int_0^T -B_0(t)\,\mathrm{EnE}(t)\,\mathrm{d}\mathbf{S}_A(t) \tag{4.18}$$

where $\mathrm{EnE}(t)$ is the risk-neutral discounted expected negative exposure:

$$\mathrm{EnE}(t) = \mathbb{E}^{\mathbb{Q}}\left[e^{-}(t)\right]$$

We deduce that the bilateral CVA is:

$$\begin{aligned}
\mathrm{BCVA} &= \mathrm{DVA} - \mathrm{CVA} \\
&= (1 - \boldsymbol{\mathcal{R}}_A) \cdot \int_0^T -B_0(t)\,\mathrm{EnE}(t)\,\mathrm{d}\mathbf{S}_A(t) - \\
&\quad (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T -B_0(t)\,\mathrm{EpE}(t)\,\mathrm{d}\mathbf{S}_B(t)
\end{aligned} \tag{4.19}$$

When we calculate the bilateral CVA as the difference between the DVA and the CVA, we consider that the DVA does not depend on $\boldsymbol{\tau}_B$ and the CVA does not depend on $\boldsymbol{\tau}_A$. In the more general case, we have:

$$\mathrm{BCVA} = \mathbb{E}^{\mathbb{Q}}\left[\begin{array}{l}
\mathbb{1}\{\boldsymbol{\tau}_A \leq \min(T, \boldsymbol{\tau}_B)\} \cdot e^{-\int_0^{\boldsymbol{\tau}_A} r_t\,\mathrm{d}t} \cdot G - \\
\mathbb{1}\{\boldsymbol{\tau}_B \leq \min(T, \boldsymbol{\tau}_A)\} \cdot e^{-\int_0^{\boldsymbol{\tau}_B} r_t\,\mathrm{d}t} \cdot L
\end{array}\right] \tag{4.20}$$

In this case, the calculation of the bilateral CVA requires considering the joint survival function of $(\boldsymbol{\tau}_A, \boldsymbol{\tau}_B)$.

Remark 56 *If we assume that the yield curve is flat and* $\mathbf{S}_B(t) = e^{-\lambda_B t}$*, we have* $\mathrm{d}\mathbf{S}_B(t) = -\lambda_B e^{-\lambda_B t}\,\mathrm{d}t$ *and:*

$$\begin{aligned}
\mathrm{CVA} &= (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T e^{-rt}\,\mathrm{EpE}(t)\,\lambda_B e^{-\lambda_B t}\,\mathrm{d}t \\
&= \boldsymbol{\mathcal{s}}_B \cdot \int_0^T e^{-(r+\lambda_B)t}\,\mathrm{EpE}(t)\,\mathrm{d}t
\end{aligned}$$

We notice that the CVA is the product of the CDS spread and the discounted value of the expected positive exposure.

Example 48 *Let us assume that the mark-to-market value is given by:*

$$\mathrm{MtM}(t) = N\int_t^T f(t,T)\,B_t(s)\,\mathrm{d}s - N\int_t^T f(0,T)\,B_t(s)\,\mathrm{d}s$$

where N and T are the notional and the maturity of the swap, and $f(t,T)$ is the instantaneous forward rate which follows a geometric Brownian motion:

$$\mathrm{d}f(t,T) = \mu f(t,T)\,\mathrm{d}t + \sigma f(t,T)\,\mathrm{d}W(t)$$

We also assume that the yield curve is flat – $B_t(s) = e^{-r(s-t)}$ – and the risk-neutral survival function is $\mathbf{S}(t) = e^{-\lambda t}$.

Syrkin and Shirazi (2015) show that[25]:

$$\mathrm{EpE}(t) = Nf(0,T)\,\varphi(t,T)\left(e^{\mu t}\Phi\left(\left(\frac{\mu}{\sigma}+\frac{1}{2}\sigma\right)\sqrt{t}\right) - \Phi\left(\left(\frac{\mu}{\sigma}-\frac{1}{2}\sigma\right)\sqrt{t}\right)\right)$$

where:

$$\varphi(t,T) = \frac{1-e^{-r(T-t)}}{r}$$

It follows that the CVA at time t is equal to:

$$\mathrm{CVA}(t) = \mathcal{s}_B \cdot \int_t^T e^{-(r+\lambda)(u-t)}\,\mathrm{EpE}(u)\,\mathrm{d}u$$

We consider the following numerical values: $N = 1000$, $f(0,T) = 5\%$, $\mu = 2\%$, $\sigma = 25\%$, $T = 10$ years and $\boldsymbol{\mathcal{R}}_B = 50\%$. In Figure 4.9, we have reported the value of $\mathrm{CVA}(t)$ when λ is respectively equal to 20 and 100 bps. By construction, the CVA is maximum at the starting date.

4.2.1.3 Practical implementation for calculating CVA

In practice, we calculate CVA and DVA by approximating the integral by a sum:

$$\mathrm{CVA} = (1-\boldsymbol{\mathcal{R}}_B)\cdot\sum_{t_i\leq T} B_0(t_i)\cdot\mathrm{EpE}(t_i)\cdot(\mathbf{S}_B(t_{i-1}) - \mathbf{S}_B(t_i))$$

and:

$$\mathrm{DVA} = (1-\boldsymbol{\mathcal{R}}_A)\cdot\sum_{t_i\leq T} B_0(t_i)\cdot\mathrm{EnE}(t_i)\cdot(\mathbf{S}_A(t_{i-1}) - \mathbf{S}_A(t_i))$$

where $\{t_i\}$ is a partition of $[0,T]$. For the bilateral CVA, the expression (4.20) can be evaluated using Monte Carlo methods.

We notice that the approximation of $\mathrm{d}\mathbf{S}_B(t)$ is equal to the default probability of Bank B between two consecutive trading dates:

$$\begin{aligned}\mathbf{S}_B(t_{i-1}) - \mathbf{S}_B(t_i) &= \Pr\{t_{i-1} < \boldsymbol{\tau}_B \leq t_i\}\\ &= \mathrm{PD}_B(t_{i-1},t_i)\end{aligned}$$

and we may wonder what is the best approach for estimating $\mathrm{PD}_B(t_{i-1},t_i)$. A straightforward solution is to use the default probabilities computed by the internal credit system.

[25] See Exercise 4.4.5 on page 303.

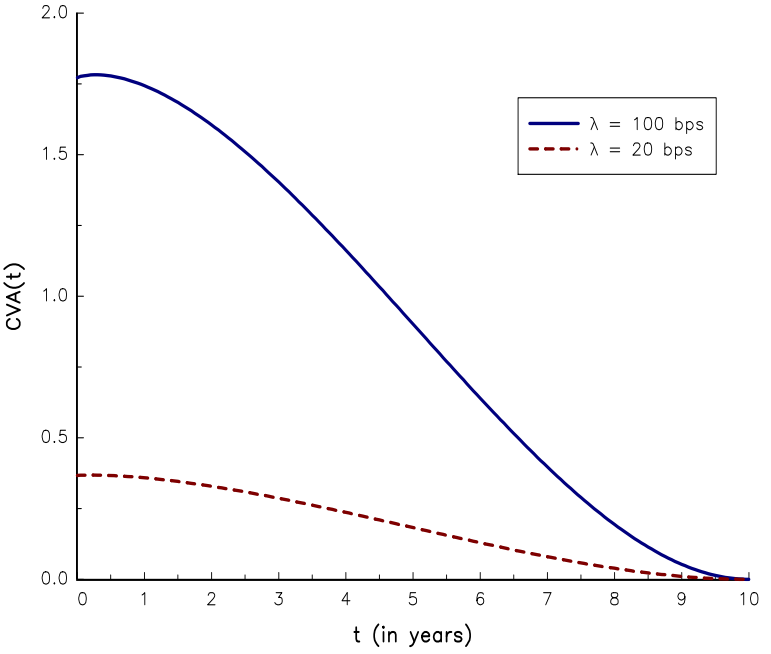


FIGURE 4.9: CVA of fixed-float swaps

However, there is a fundamental difference between CCR and CVA. Indeed, CCR is a default risk and must then be calculated using the historical probability measure $\mathbb{P}$. On the contrary, CVA is a market price, implying that it is valued under the risk-neutral probability measure $\mathbb{Q}$. Therefore, $\text{PD}_B(t_{i-1}, t_i)$ is a risk-neutral probability. Using the credit triangle relationship, we know that the CDS spread $\mathcal{s}$ is related to the intensity λ:

$$\mathcal{s}_B(t) = (1 - \mathcal{R}_B) \cdot \lambda_B(t)$$

We deduce that:

$$\begin{aligned} \mathbf{S}_B(t) &= \exp(-\lambda_B(t) \cdot t) \\ &= \exp\left(-\frac{\mathcal{s}_B(t) \cdot t}{1 - \mathcal{R}_B}\right) \end{aligned}$$

It follows that the risk-neutral probability of default $\text{PD}_B(t_{i-1}, t_i)$ is equal to:

$$\text{PD}_B(t_{i-1}, t_i) = \exp\left(-\frac{\mathcal{s}_B(t_{i-1}) \cdot t_{i-1}}{1 - \mathcal{R}_B}\right) - \exp\left(-\frac{\mathcal{s}_B(t_i) \cdot t_i}{1 - \mathcal{R}_B}\right)$$

4.2.2 Regulatory capital

The capital charge for the CVA risk has been introduced by the Basel Committee in December 2010 after the Global Financial Crisis. At that moment, banks had the choice between two approaches: the advanced method (AM-CVA) and the standardized method (SM-CVA). However, the Basel Committee completely changed the CVA framework in December 2017 with two new approaches (BA-CVA and SA-CVA) that will replace the previous approaches (AM-CVA and SM-CVA) with effect from January 2022. It is the first time that the Basel Committee completely flip-flopped within the same accord, since these different approaches are all part of the Basel III Accord.

4.2.2.1 The 2010 version of Basel III

Advanced method The advanced method (or AM-CVA) can be considered by banks that use IMM and VaR models. In this approach, we approximate the integral by the middle Riemann sum:

$$\text{CVA} = \text{LGD}_B \cdot \sum_{t_i \leq T} \left(\frac{\text{EpE}\left(t_{i-1}\right) B_0\left(t_{i-1}\right) + B_0\left(t_i\right) \text{EpE}\left(t_i\right)}{2} \right) \cdot \text{PD}_B\left(t_{i-1}, t_i\right)$$

where $\text{LGD} = 1 - \boldsymbol{\mathcal{R}}_B$ is the risk-neutral loss given default of the counterparty B and $\text{PD}_B\left(t_{i-1}, t_i\right)$ is the risk neutral probability of default between t_{i-1} and t_i:

$$\text{PD}_B\left(t_{i-1}, t_i\right) = \max\left(\exp\left(-\frac{\boldsymbol{\mathcal{s}}\left(t_{i-1}\right)}{\text{LGD}_B} \cdot t_{i-1} \right) - \exp\left(-\frac{\boldsymbol{\mathcal{s}}\left(t_i\right)}{\text{LGD}_B} \cdot t_i \right), 0 \right)$$

We notice that a zero floor is added in order to verify that $\text{PD}_B\left(t_{i-1}, t_i\right) \geq 0$. The capital charge is then equal to:

$$\boldsymbol{\mathcal{K}} = 3 \cdot (\text{CVA} + \text{SCVA})$$

where CVA is calculated using the last one-year period and SCVA is the stressed CVA based on a one-year stressed period of credit spreads.

Standardized method In the standardized method (or SM-CVA), the capital charge is equal to:

$$\boldsymbol{\mathcal{K}} = 2.33 \cdot \sqrt{h} \cdot \sqrt{\left(\frac{1}{2} \sum_i w_i \cdot \Omega_i - w^\star_{\text{index}} \cdot \Omega^\star_{\text{index}} \right)^2 + \frac{3}{4} \sum_i w_i^2 \cdot \Omega_i^2} \tag{4.21}$$

where:

$$\Omega_i = \text{M}_i \cdot \text{EAD}_i \cdot \frac{1 - e^{-0.05 \cdot \text{M}_i}}{0.05 \cdot \text{M}_i} - \text{M}^\star_i \cdot H^\star_i \cdot \frac{1 - e^{-0.05 \cdot \text{M}^\star_i}}{0.05 \cdot \text{M}^\star_i}$$

$$\Omega^\star_{\text{index}} = \text{M}^\star_{\text{index}} \cdot H^\star_{\text{index}} \cdot \frac{1 - e^{-0.05 \cdot \text{M}^\star_{\text{index}}}}{0.05 \cdot \text{M}^\star_{\text{index}}}$$

In this formula, h is the time horizon (one year), w_i is the weight of the i^{th} counterparty based on its rating, M_i is the effective maturity of the i^{th} netting set, EAD_i is the exposure at default of the i^{th} netting set, $\text{M}^\star_i$ is the maturity adjustment factor for the single name hedge, $H^\star_i$ is the hedging notional of the single name hedge, $w^\star_{\text{index}}$ is the weight of the index hedge, $\text{M}^\star_{\text{index}}$ is the maturity adjustment factor for the index hedge and $H^\star_{\text{index}}$ is the hedging notional of the index hedge. In this formula, EAD_i corresponds to the CCR exposure at default calculated with the CEM or IMM approaches.

Remark 57 *We notice that the Basel Committee recognizes credit hedges (single-name CDS, contingent CDS and CDS indices) for reducing CVA volatility. If there is no hedge, we obtain:*

$$\boldsymbol{\mathcal{K}} = 2.33 \cdot \sqrt{h} \cdot \sqrt{\frac{1}{4} \left(\sum_i w_i \cdot \text{M}_i \cdot \text{EAD}_i \right)^2 + \frac{3}{4} \sum_i w_i^2 \cdot \text{M}_i^2 \cdot \text{EAD}_i^2}$$

The derivation of Equation (4.21) is explained in Pykhtin (2012). We consider a Gaussian random vector $X = (X_1, \ldots, X_n)$ with $X_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$. We assume that the random variables $X_1, \ldots, X_n$ follow a single risk factor model such that the correlation $\rho\left(X_i, X_j\right)$

is constant and equal to ρ. We consider another random variable $X_{n+1} \sim \mathcal{N}\left(0, \sigma_{n+1}^2\right)$ such that $\rho\left(X_i, X_{n+1}\right)$ is also constant and equal to ρ_{n+1}. Let Y be the random variable defined as the sum of X_i's minus X_{n+1}:

$$Y = \sum_{i=1}^{n} X_i - X_{n+1}$$

It follows that $Y \sim \mathcal{N}\left(0, \sigma_Y^2\right)$ where:

$$\sigma_Y^2 = \sum_{i=1}^{n} \sigma_i^2 + 2\rho \sum_{i=1}^{n} \sum_{j=1}^{i} \sigma_i \sigma_j - 2\rho_{n+1}\sigma_{n+1} \sum_{i=1}^{n} \sigma_i + \sigma_{n+1}^2$$

We finally deduce that:

$$\mathbf{F}_Y^{-1}(\alpha) = \Phi^{-1}(\alpha) \sqrt{\sum_{i=1}^{n} \sigma_i^2 + 2\rho \sum_{i=1}^{n} \sum_{j=1}^{i} \sigma_i \sigma_j - 2\rho_{n+1}\sigma_{n+1} \sum_{i=1}^{n} \sigma_i + \sigma_{n+1}^2}$$

Equation (4.21) is obtained by setting $\sigma_i = w_i \Omega_i$, $\sigma_{n+1} = w_{\text{index}}^\star \Omega_{\text{index}}^\star$, $\rho = 25\%$, $\rho_{n+1} = 50\%$ and $\alpha = 99\%$. This means that X_i is the CVA net exposure of the i^{th} netting set (including individual hedges) and X_{n+1} is the macro hedge of the CVA based on credit indices.

4.2.2.2 The 2017 version of Basel III

There are now two approaches available for calculating CVA risk: the basic approach (BA-CVA) and the standardized approach (SA-CVA). However, if the bank has a few exposure on counterparty credit risk[26], it may choose to set its CVA capital requirement equal to its CCR capital requirement.

Basic approach Under the basic approach, the capital requirement is equal to:

$$\boldsymbol{\mathcal{K}} = \beta \cdot \boldsymbol{\mathcal{K}}^{\text{Reduced}} + (1 - \beta) \cdot \boldsymbol{\mathcal{K}}^{\text{Hedged}}$$

where $\boldsymbol{\mathcal{K}}^{\text{Reduced}}$ and $\boldsymbol{\mathcal{K}}^{\text{Hedged}}$ are the capital requirements without and with hedging recognition. The reduced version of the BA-CVA is obtained by setting β to 100%. A bank that actively hedges CVA risks may choose the full version of the BA-CVA. In this case, β is set to 25%.

For the reduced version, we have:

$$\boldsymbol{\mathcal{K}}^{\text{Reduced}} = \sqrt{\left(\rho \cdot \sum_j \text{SCVA}_j\right)^2 + (1 - \rho^2) \cdot \sum_j \text{SCVA}_j^2}$$

where $\rho = 50\%$ and SCVA_j is the CVA capital requirement for the j^{th} counterparty:

$$\text{SCVA}_j = \frac{1}{\alpha} \cdot \text{RW}_j \cdot \sum_k \text{DF}_k \cdot \text{EAD}_k \cdot \text{M}_k$$

In this formula, α is set to 1.4, RW_j is the risk weight for counterparty j, k is the netting set, DF_k is the discount factor, EAD_k is the CCR exposure at default and M_k is the effective

[26] The materiality threshold is €100 bn for the notional amount of non-centrally cleared derivatives.

maturity. These last three quantities are calculated at the netting set level. If the bank use the IMM to calculate the exposure at default, DF_k is equal to one, otherwise we have:

$$\text{DF}_k = \frac{1 - e^{-0.05 \cdot \text{M}_k}}{0.05 \cdot \text{M}_k}$$

RW_j depends on the credit quality of the counterparty (IG/HY) and its sector and is given in Table 4.7.

TABLE 4.7: Supervisory risk weights (BA-CVA)

Sector	Credit quality	
	IG	HY/NR
Sovereign	0.5%	3.0%
Local government	1.0%	4.0%
Financial	5.0%	12.0%
Basic material, energy, industrial, agriculture, manufacturing, mining and quarrying	3.0%	7.0%
Consumer goods and services, transportation and storage, administrative and support service activities	3.0%	8.5%
Technology, telecommunication	2.0%	5.5%
Health care, utilities, professional and technical activities	1.5%	5.0%
Other sector	5.0%	12.0%

Source: BCBS (2017c).

The full version of the BA-CVA recognizes eligible hedging transactions that are used for mitigating the credit spread component of the CVA risk. They correspond to single-name CDS and index CDS transactions. $\boldsymbol{\mathcal{K}}^{\text{Hedged}}$ depends on three components:

$$\boldsymbol{\mathcal{K}}^{\text{Hedged}} = \sqrt{K_1 + K_2 + K_3}$$

According to BCBS (2017c), the first term aggregates the systematic components of the CVA risk:

$$K_1 = \left(\rho \cdot \sum_j \left(\text{SCVA}_j - \text{SNH}_j \right) - \text{IH} \right)^2$$

where SNH_j is the CVA reduction for counterparty j due to single-name hedging and IH is the global CVA reduction due to index hedging. The second term aggregates the idiosyncratic components of the CVA risk:

$$K_2 = \left(1 - \rho^2\right) \cdot \sum_j \left(\text{SCVA}_j - \text{SNH}_j \right)^2$$

Finally, the third term corresponds to the hedging misalignment risk because of the mismatch between indirect hedges and single-name hedges:

$$K_3 = \sum_j \text{HMA}_j$$

The single-name hedge SNH_j is calculated as follows:

$$\text{SNH}_j = \sum_{h \in j} \varrho_{h,j} \cdot (\text{RW}_h \cdot \text{DF}_h \cdot N_h \cdot \text{M}_h)$$

where h represents the single-name CDS transaction, $\varrho_{h,j}$ is the supervisory correlation, DF_h is the discount factor[27], N_h is the notional and M_h is the remaining maturity. These quantities are calculated at the single-name CDS level. The correlation $\varrho_{h,j}$ between the credit spread of the counterparty and the credit spread of the CDS can take three values: 100% if CDS h directly refers to counterparty j, 80% if CDS h has a legal relation with counterparty j, and 50% if CDS h and counterparty j are of the same sector and region. For the index hedge IH, we have a similar formula:

$$\text{IH} = \sum_{h'} \text{RW}_{h'} \cdot \text{DF}_{h'} \cdot N_{h'} \cdot \text{M}_{h'}$$

where h' represents the index CDS transaction. The other quantities $\text{RW}_{h'}$, $\text{DF}_{h'}$, $N_{h'}$ and $\text{M}_{h'}$ are defined exactly as previously except that they are applied at the index CDS level. For the risk weight, its value is the weighted average of risk weights of RW_j:

$$\text{RW}_{h'} = 0.7 \cdot \sum_{j \in h'} w_j \cdot \text{RW}_j$$

where w_j is the weight of the counterparty/sector j in the index CDS h'. We notice that this formula reduces to $\text{RW}_{h'} = 0.7 \cdot \text{RW}_j$ when we consider a sector-specific index. Finally, we have

$$\text{HMA}_j = \sum_{h \in j} \left(1 - \varrho_{h,j}^2\right) \cdot (\text{RW}_h \cdot \text{DF}_h \cdot N_h \cdot \text{M}_h)^2$$

Remark 58 *In the case where there is no hedge, we have* $\text{SNH}_j = 0$, $\text{HMA}_j = 0$, $\text{IH} = 0$, *and* $\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{K}}^{\text{Reduced}}$. *If there is no hedging misalignment risk and no index CDS hedging, we have:*

$$\boldsymbol{\mathcal{K}} = \sqrt{\left(\rho \cdot \sum_j \boldsymbol{\mathcal{K}}_j\right)^2 + (1 - \rho^2) \cdot \sum_j \boldsymbol{\mathcal{K}}_j^2}$$

where $\boldsymbol{\mathcal{K}}_j = \text{SCVA}_j - \text{SNH}_j$ *is the single-name capital requirement for counterparty* j.

Example 49 *We assume that the bank has three financial counterparties A, B and C, that are respectively rated IG, IG and HY. There are 4 OTC transactions, whose characteristics are the following:*

Transaction k	1	2	3	4
Counterparty	*A*	*A*	*B*	*C*
EAD_k	100	50	70	20
M_k	1	1	0.5	0.5

In order to reduce the counterparty credit risk, the bank has purchased a CDS protection on A for an amount of \$75 *mn, a CDS protection on B for an amount of* \$10 *mn and a HY Financial CDX for an amount of* \$10 *mn. The maturity of hedges exactly matches the maturity of transactions. However, the CDS protection on B is indirect, because the underlying name is not B, but* B' *which is the parent company of B.*

[27] We have:

$$\text{DF}_h = \frac{1 - e^{-0.05 \cdot \text{M}_h}}{0.05 \cdot \text{M}_h}$$

where M_h is the remaining maturity.

We first begin to calculate the discount factors DF_k for the four transactions. We obtain $\mathrm{DF}_1 = \mathrm{DF}_2 = 0.9754$ and $\mathrm{DF}_3 = \mathrm{DF}_4 = 0.9876$. Then we calculate the single-name capital for each counterparty. For example, we have:

$$\begin{aligned}\mathrm{SCVA}_A &= \frac{1}{\alpha} \times \mathrm{RW}_A \times (\mathrm{DF}_1 \times \mathrm{EAD}_1 \times \mathrm{M}_1 + \mathrm{DF}_2 \times \mathrm{EAD}_2 \times \mathrm{M}_2) \\ &= \frac{1}{1.4} \times 5\% \times (0.9754 \times 100 \times 1 + 0.9754 \times 50 \times 1) \\ &= 5.225\end{aligned}$$

We also find that $\mathrm{SCVA}_B = 1.235$ and $\mathrm{SCVA}_C = 0.847$. It follows that $\sum_j \mathrm{SCVA}_j = 7.306$ and $\sum_j \mathrm{SCVA}_j^2 = 29.546$. The capital requirement without hedging is equal to:

$$\boldsymbol{\mathcal{K}}^{\text{Reduced}} = \sqrt{(0.5 \times 7.306)^2 + (1 - 0.5^2) \times 29.546} = 5.959$$

We notice that it is lower than the sum of individual capital charges. In order to take into account the hedging effect, we calculate the single-name hedge parameters:

$$\mathrm{SNH}_A = 5\% \times 100\% \times 0.9754 \times 75 \times 1 = 3.658$$

and:

$$\mathrm{SNH}_B = 5\% \times 80\% \times 0.9876 \times 10 \times 0.5 = 0.198$$

Since the CDS protection is on B' and not B, there is a hedging misalignment risk:

$$\mathrm{HMA}_B = 0.05^2 \times \left(1 - 0.80^2\right) \times (0.9876 \times 10 \times 0.5)^2 = 0.022$$

For the CDX protection, we have:

$$\mathrm{IH} = (0.7 \times 12\%) \times 0.9876 \times 10 \times 0.5 = 0.415$$

Then, we obtain $K_1 = 1.718$, $K_2 = 3.187$, $K_3 = 0.022$ and $\boldsymbol{\mathcal{K}}^{\text{Hedged}} = 2.220$. Finally, the capital requirement is equal to \$3.154 mn:

$$\boldsymbol{\mathcal{K}} = 0.25 \times 5.959 + 0.75 \times 2.220 = 3.154$$

Standardized approach The standardized approach for CVA follows the same principles than the standardized approach SA-TB for the market risk of the trading book. The main difference is that SA-CVA is only based on delta and vega risks, and does not include curvature, jump-to-default and residual risks:

$$\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{K}}^{\text{Delta}} + \boldsymbol{\mathcal{K}}^{\text{Vega}}$$

For computing the capital charge, we first consider two portfolios: the CVA portfolio and the hedging portfolio. For each risk (delta and vega), we calculate the weighted CVA sensitivity of each risk factor $\mathcal{F}_j$:

$$\mathrm{WS}_j^{\text{CVA}} = S_j^{\text{CVA}} \cdot \mathrm{RW}_j$$

and:

$$\mathrm{WS}_j^{\text{Hedge}} = S_j^{\text{Hedge}} \cdot \mathrm{RW}_j$$

where S_j and RW_j are the net sensitivity of the CVA or hedging portfolio with respect to the risk factor and the risk weight of $\mathcal{F}_j$. Then, we aggregate the weighted sensitivity in order to obtain a net figure:

$$\mathrm{WS}_j = \mathrm{WS}_j^{\text{CVA}} + \mathrm{WS}_j^{\text{Hedge}}$$

Second, we calculate the capital requirement for the risk bucket $\mathcal{B}_k$:

$$\mathcal{K}_{\mathcal{B}_k} = \sqrt{\sum_j \mathrm{WS}_j^2 + \sum_{j' \neq j} \rho_{j,j'} \cdot \mathrm{WS}_j \cdot \mathrm{WS}_{j'} + 1\% \cdot \sum_j \left(\mathrm{WS}_j^{\mathrm{Hedge}}\right)^2}$$

where $\mathcal{F}_j \in \mathcal{B}_k$. Finally, we aggregate the different buckets for a given risk class:

$$\mathcal{K}^{\mathrm{Delta/Vega}} = m_{\mathrm{CVA}} \cdot \sqrt{\sum_k \mathcal{K}_{\mathcal{B}_k}^2 + \sum_{k' \neq k} \gamma_{k,k'} \cdot \mathcal{K}_{\mathcal{B}_k} \cdot \mathcal{K}_{\mathcal{B}_{k'}}}$$

where $m_{\mathrm{CVA}} = 1.25$ is the multiplier factor. As in the case of SA-TB, SA-CVA is then based on the following set of parameters: the sensitivities S_j of the risk factors that are calculated by the bank; the risk weights RW_j of the risk factors; the correlation $\rho_{j,j'}$ between risk factors within a bucket; the correlation $\gamma_{k,k'}$ between the risk buckets. The values of these parameters are not necessarily equal to those of SA-TB[28]. For instance, the correlations $\rho_{j,j'}$ and $\gamma_{k,k'}$ are generally lower. The reason is that these correlations reflect the dependence between credit risk factors and not market risk factors.

Remark 59 *Contrary to the SA-TB, the bank must have the approval of the supervisory authority to use the SA-CVA. Otherwise, it must use the BA-CVA framework.*

4.2.3 CVA and wrong/right way risk

The wrong way or right way risk is certainly the big challenge when modeling CVA. We have already illustrated this point in the case of the CCR capital requirement, but this is even more relevant when computing the CVA capital requirement. The reason is that the bank generally manages the CVA risk because it represents a huge cost in terms of regulatory capital and it impacts on a daily basis the P&L of the trading book. For that, the bank generally puts in place a CVA trading desk, whose objective is to mitigate CVA risks. Therefore, the CVA desk must develop a fine modeling of WWR/RWR risks in order to be efficient and to be sure that the hedging portfolio does not create itself another source of hidden wrong way risk. This is why the CVA modeling is relatively complex, because we cannot assume in practice that market and credit risks are not correlated.

We reiterate that the definition of the CVA is[29]:

$$\mathrm{CVA} = \mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\tau} \leq T\right\} \cdot e^{-\int_0^{\boldsymbol{\tau}} r_t \,\mathrm{d}t} \cdot (1 - \boldsymbol{\mathcal{R}}) \cdot e^{+}(\boldsymbol{\tau})\right]$$

where $e^{+}(t) = \max(\omega, 0)$ and ω is the random variable that represents the mark-to-market[30]. If we assume that the recovery rate is constant and interest rates are deterministic, we obtain:

$$\begin{aligned}
\mathrm{CVA} &= (1 - \boldsymbol{\mathcal{R}}) \cdot \int_0^T \int_{-\infty}^{+\infty} B_0(t) \max(\omega, 0)\ \mathrm{d}\mathbf{F}(\omega, t) \\
&= (1 - \boldsymbol{\mathcal{R}}) \cdot \int_0^T \int_{-\infty}^{+\infty} B_0(t) \max(\omega, 0)\ \mathrm{d}\mathbf{C}\left(\mathbf{F}_\omega(\omega), \mathbf{F}_\tau(t)\right)
\end{aligned}$$

[28] See BCBS (2017c) on pages 119-127.

[29] In order to obtain more concise formulas, we delete the reference to the counterparty B and we write $\boldsymbol{\mathcal{R}}$ instead of $\boldsymbol{\mathcal{R}}_B$.

[30] We implicitly assume that the mark-to-market is a stationary process. In fact, this assumption is not verified. However, we use this simplification to illustrate how the dependence between the counterparty exposure and the default times changes the CVA figure.

where $\mathbf{F}(\omega, t)$ is the joint distribution of the mark-to-market and the default time and $\mathbf{C}$ is the copula between ω and $\boldsymbol{\mathcal{R}}$. If we assume that $\mathbf{C} = \mathbf{C}^{\perp}$, we retrieve the traditional CVA formula[31]:

$$\begin{aligned}\text{CVA} &= (1-\boldsymbol{\mathcal{R}}) \cdot \int_0^T \int_{-\infty}^{+\infty} B_0(t) \max(\omega, 0)\ \mathrm{d}\mathbf{F}_{\omega}(\omega)\ \mathrm{d}\mathbf{F}_{\tau}(t) \\ &= (1-\boldsymbol{\mathcal{R}}) \cdot \int_0^T B_0(t)\, \mathrm{EpE}(t)\ \mathrm{d}\mathbf{F}_{\tau}(t)\end{aligned}$$

where $\mathrm{EpE}(t)$ is the expected positive exposure:

$$\mathrm{EpE}(t) = \int_{-\infty}^{+\infty} \max(\omega, 0)\ \mathrm{d}\mathbf{F}_{\omega}(\omega) = \mathbb{E}\left[e^{+}(t)\right]$$

Otherwise, we have to model the dependence between the mark-to-market and the default time. In what follows, we consider two approaches: the copula model introduced by Cespedes *et al.* (2010) and the hazard rate model of Hull and White (2012).

The copula approach The Monte Carlo CVA is calculated as following:

$$\text{CVA} = (1-\boldsymbol{\mathcal{R}}) \cdot \sum_{t_i \leq T} B_0(t_i) \left(\frac{1}{n_S} \sum_{s=1}^{n_S} e_s^{+}(t_i; \omega_s) \right) \left(\mathbf{F}_{\boldsymbol{\tau}}(t_i) - \mathbf{F}_{\boldsymbol{\tau}}(t_{i-1})\right)$$

where $e^{+}(t_i; \omega_s)$ is the counterparty exposure of the s^{th} simulated scenario ω_s and n_S is the number of simulations. If market and credit risk factors are correlated, the Monte Carlo CVA becomes:

$$\text{CVA} = (1-\boldsymbol{\mathcal{R}}) \cdot \sum_{t_i \leq T} \sum_{s=1}^{n_S} B_0(t_i)\, e_s^{+}(t_i; \omega_s)\, \pi_{s,i} \tag{4.22}$$

where[32]:

$$\pi_{s,i} = \Pr\left\{\omega = \omega_s, t_i < \boldsymbol{\tau} \leq t_i\right\}$$

The objective is then to calculate the joint probability by assuming a copula function $\mathbf{C}$ between ω and $\boldsymbol{\tau}$. For that, we assume that the scenarios ω_s are ordered. Let $U = \mathbf{F}_{\omega}(\omega)$ and $V = \mathbf{F}_{\boldsymbol{\tau}}(\boldsymbol{\tau})$ be the integral transform of ω and $\boldsymbol{\tau}$. Since U and V are uniform random variables, we obtain:

$$\begin{aligned}\pi_{s,i} &= \Pr\left\{\omega_{s-1} < \omega \leq \omega_s, t_i < \boldsymbol{\tau} \leq t_i\right\} \\ &= \Pr\left\{u_{s-1} < U \leq u_s, v_{i-1} < V \leq v_i\right\} \\ &= \mathbf{C}(u_s, v_i) - \mathbf{C}(u_{s-1}, v_i) - \mathbf{C}(u_s, v_{i-1}) + \mathbf{C}(u_{s-1}, v_{i-1})\end{aligned} \tag{4.23}$$

Generally, we don't know the analytical expression of $\mathbf{F}_{\omega}$. This is why we replace it by the empirical distribution $\hat{\mathbf{F}}_{\omega}$ where the probability of each scenario is equal to $1/n_S$.

In order to define the copula function $\mathbf{C}$, Rosen and Saunders (2012) consider a market-credit version of the Basel model. Let $Z_m = \Phi^{-1}(\mathbf{F}_{\omega}(\omega))$ and $Z_c = \Phi^{-1}(\mathbf{F}_{\boldsymbol{\tau}}(\boldsymbol{\tau}))$ be the

[31] See Equation (4.16) on page 280.

[32] In the case where ω and $\boldsymbol{\tau}$ are independent, we retrieve the previous formula because we have:

$$\begin{aligned}\pi_{s,i} &= \Pr\left\{\omega = \omega_s\right\} \cdot \Pr\left\{t_i \leq \boldsymbol{\tau} \leq t_i\right\} \\ &= \frac{\mathbf{F}_{\boldsymbol{\tau}}(t_i) - \mathbf{F}_{\boldsymbol{\tau}}(t_{i-1})}{n_S}\end{aligned}$$

normalized latent random variables for market and credit risks. Rosen and Saunders use the one-factor model specification:

$$\begin{cases} Z_m = \rho_m X + \sqrt{1-\rho_m^2}\,\varepsilon_m \\ Z_c = \rho_c X + \sqrt{1-\rho_c^2}\,\varepsilon_c \end{cases}$$

where X is the systematic risk factor that impacts both market and credit risks, ε_m and ε_c are the idiosyncratic market and credit risk factors, and ρ_m and ρ_c are the market and credit correlations with the common risk factor. It follows that the market-credit correlation is equal to:

$$\rho_{m,c} = \mathbb{E}\left[Z_m Z_c\right] = \rho_m \rho_c$$

We deduce that the dependence between Z_m and Z_c is a Normal copula with parameter $\rho_{m,c} = \rho_m \rho_c$, and we can write:

$$Z_m = \rho_{m,c} Z_c + \sqrt{1-\rho_{m,c}^2}\,\varepsilon_{m,c}$$

where $\varepsilon_{m,c} \sim \mathcal{N}(0,1)$ is an independent specific risk factor. Since the expression of the Normal copula is $\mathbf{C}(u,v;\rho_{m,c}) = \Phi_2\left(\Phi^{-1}(u), \Phi^{-1}(\mathbf{v}); \rho_{m,c}\right)$, Equation (4.23) becomes[33]:

$$\begin{aligned} \pi_{s,i} &= \Phi_2\left(\Phi^{-1}\left(\frac{s}{n_S}\right), \Phi^{-1}\left(\mathbf{F}_{\boldsymbol{\tau}}(t_i)\right); \rho_{m,c}\right) - \\ &\quad \Phi_2\left(\Phi^{-1}\left(\frac{s-1}{n_S}\right), \Phi^{-1}\left(\mathbf{F}_{\boldsymbol{\tau}}(t_i)\right); \rho_{m,c}\right) - \\ &\quad \Phi_2\left(\Phi^{-1}\left(\frac{s}{n_S}\right), \Phi^{-1}\left(\mathbf{F}_{\boldsymbol{\tau}}(t_{i-1})\right); \rho_{m,c}\right) + \\ &\quad \Phi_2\left(\Phi^{-1}\left(\frac{s-1}{n_S}\right), \Phi^{-1}\left(\mathbf{F}_{\boldsymbol{\tau}}(t_{i-1})\right); \rho_{m,c}\right) \end{aligned}$$

This approach is called the ordered-scenario copula model (OSC), because it is based on the ordering trick of the scenarios ω_s. Rosen and Saunders (2012) also propose different versions of the CVA discretization leading to different expressions of Equation (4.22). For instance, if we assume that the default occurs exactly at time t_i and not in the interval $[t_{i-1}, t_i]$, we have:

$$\pi_{s,i} \approx \pi_{s|i} \cdot \Pr\left\{t_i < \boldsymbol{\tau} \leq t_i\right\}$$

and:

$$\begin{aligned} \pi_{s|i} &= \Pr\left\{\omega = \omega_s \mid \boldsymbol{\tau} = t_i\right\} \\ &= \Pr\left\{\omega_{s-1} < \omega \leq \omega_s \mid \boldsymbol{\tau} = t_i\right\} \\ &= \Pr\left\{u_{s-1} < U \leq u_s \mid V = v_i\right\} \\ &= \partial_2 \mathbf{C}(u_s, v_i) - \partial_2 \mathbf{C}(u_{s-1}, v_i) \\ &= \partial_2 \mathbf{C}\left(\frac{s}{n_S}, \mathbf{F}_{\boldsymbol{\tau}}(t_i)\right) - \partial_2 \mathbf{C}\left(\frac{s-1}{n_S}, \mathbf{F}_{\boldsymbol{\tau}}(t_i)\right) \end{aligned}$$

In the case of the Rosen-Saunders model, we use the expression of the conditional Normal copula given on page 737:

$$\partial_2 \mathbf{C}(u,v;\rho_{m,c}) = \Phi\left(\frac{\Phi^{-1}(u) - \rho_{m,c}\Phi^{-1}(v)}{\sqrt{1-\rho_{m,c}^2}}\right)$$

[33] By definition, we have $\mathbf{F}_{\omega}^{-1}(\varpi_s) = s/n_S$ because the scenarios are ordered.

The hazard rate approach In Basel II, wrong way risk is addressed by introducing the multiplier $\alpha = 1.4$, which is equivalent to change the values of the mark-to-market. In the Rosen-Saunders model, wrong way risk is modeled by changing the joint probability of the mark-to-market and the default times. Hull and White (2012) propose a third approach, which consists in changing the values of the default probabilities. They consider that the hazard rate is a deterministic function of the mark-to-market: $\lambda(t) = \lambda(t, \text{MtM}(t))$. For instance, they use two models:

$$\lambda(t, \text{MtM}(t)) = e^{a(t)+b\cdot\text{MtM}(t)} \tag{4.24}$$

and:

$$\lambda(t, \text{MtM}(t)) = \ln\left(1 + e^{a(t)+b\cdot\text{MtM}(t)}\right) \tag{4.25}$$

The case $b < 0$ corresponds to the right way risk, whereas $b > 0$ corresponds to the wrong way risk. When $b = 0$, the counterparty exposure is independent from the credit risk of the counterparty.

Hull and White (2012) propose a two-step procedure to calibrate $a(t)$ and b. First, they assume that the term structure of the hazard rate is flat. Given two pairs $(\text{MtM}_1, \mathcal{s}_1)$ and $(\text{MtM}_2, \mathcal{s}_2)$, $a(0)$ and b satisfy the following system of equations:

$$\begin{cases} (1-\mathcal{R})\cdot\lambda(0, \text{MtM}_1) = \mathcal{s}_1 \\ (1-\mathcal{R})\cdot\lambda(0, \text{MtM}_2) = \mathcal{s}_2 \end{cases}$$

The solution is:

$$\begin{cases} b = \dfrac{\ln\lambda_2 - \ln\lambda_1}{\text{MtM}_2 - \text{MtM}_1} \\ a(0) = \ln\lambda_1 - b\cdot\text{MtM}_1 \end{cases}$$

where $\lambda_i = \mathcal{s}_i/(1-\mathcal{R})$ for Model (4.24) and $\lambda_i = \exp(\mathcal{s}_i/(1-\mathcal{R})) - 1$ for Model (4.25). Hull and White (2012) consider the following example. They assume that the 5Y CDS spread of the counterparty is 300 bps when the mark-to-market is \$3 mn, and 600 bps when the mark-to-market is \$20 mn. If the recovery rate is set to 40%, the calibrated parameters are $a(0) = -3.1181$ and $b = 0.0408$ for Model (4.24) and $a(0) = -3.0974$ and $b = 0.0423$ for Model (4.25). The second step of the procedure consists in calibrating the function $a(t)$ given the value of b estimated at the first step. Since we have:

$$\mathbf{S}(t) = e^{-\int_0^t \lambda(s,\text{MtM}(s))\,\mathrm{d}s}$$

and:

$$\mathbf{S}(t) = \exp\left(-\frac{\mathcal{s}(t)\cdot t}{1-\mathcal{R}}\right)$$

the function $a(t)$ must verify that the survival probability calculated with the model is equal to the survival probability calculated with the credit spread:

$$e^{-\sum_{k=0}^{i}\lambda(t_k,\text{MtM}(t_k))\cdot(t_k-t_{k-1})} = \exp\left(-\frac{\mathcal{s}(t_i)\cdot t_i}{1-\mathcal{R}}\right)$$

In the case where the CVA is calculated with the Monte Carlo method, we have:

$$\frac{1}{n_S}\sum_{s=1}^{n_S}\prod_{k=0}^{i} e^{-\lambda(t_k,\omega_s(t_k))\cdot(t_k-t_{k-1})} = \exp\left(-\frac{\mathcal{s}(t_i)\cdot t_i}{1-\mathcal{R}}\right)$$

where $\omega_s(t_k)$ is the s^{th} simulated value of $\text{MtM}(t_k)$. Therefore, $a(t)$ is specified as a piecewise linear function and we use the bootstrap method[34] for calibrating $a(t)$ given the available market CDS spreads[35].

4.3 Collateral risk

4.3.1 Definition

When there is a margin agreement, the counterparty needs to post collateral and the exposure at default becomes:

$$e^+(t) = \max(\text{MtM}(t) - C(t), 0) \tag{4.26}$$

where $C(t)$ is the collateral value at time t. Generally, the collateral transfer occurs when the mark-to-market exceeds a threshold H:

$$C(t) = \max(\text{MtM}(t - \delta_C) - H, 0) \tag{4.27}$$

H is the minimum collateral transfer amount whereas $\delta_C \geq 0$ is the margin period of risk (MPOR). According to the Financial Conduct Authority (FCA), the margin period of risk "*stands for the time period from the most recent exchange of collateral covering a netting set of financial instruments with a defaulting counterparty until the financial instruments are closed out and the resulting market risk is re-hedged*". It can be seen as the necessary time period for posting the collateral. In many models, δ_C is set to zero in order to obtain analytical formulas. However, this is not realistic from a practical point of view. From a regulatory point of view, δ_C is generally set to five or ten days (Cont, 2018).

If we combine Equations (4.26) and (4.27), it follows that:

$$\begin{aligned} e^+(t) &= \max(\text{MtM}(t) - \max(\text{MtM}(t - \delta_C) - H, 0), 0) \\ &= \text{MtM}(t) \cdot \mathbb{1}\{0 \leq \text{MtM}(t), \text{MtM}(t - \delta_C) < H\} + \\ &\quad (\text{MtM}(t) - \text{MtM}(t - \delta_C) + H) \cdot \\ &\quad \mathbb{1}\{H \leq \text{MtM}(t - \delta_C) \leq \text{MtM}(t) + H\} \end{aligned}$$

We obtain some special cases:

- When $H = +\infty$, $C(t)$ is equal to zero and we obtain:

$$e^+(t) = \max(\text{MtM}(t), 0)$$

- When $H = 0$, the collateral $C(t)$ is equal to $\text{MtM}(t - \delta_C)$ and the counterparty exposure becomes:

$$\begin{aligned} e^+(t) &= \max(\text{MtM}(t) - \text{MtM}(t - \delta_C), 0) \\ &= \max(\text{MtM}(t - \delta_C, t), 0) \end{aligned}$$

  The counterparty credit risk corresponds to the variation of the mark-to-market $\text{MtM}(t - \delta_C, t)$ during the liquidation period $[t - \delta_C, t]$.

[34] This method is presented on page 204.
[35] Generally, they correspond to the following maturities: 1Y, 3Y, 5Y, 7Y and 10Y.

- When δ_C is set to zero, we deduce that:

$$\begin{aligned} e^+(t) &= \max(\mathrm{MtM}(t) - \max(\mathrm{MtM}(t) - H, 0), 0) \\ &= \mathrm{MtM}(t) \cdot \mathbb{1}\{0 \leq \mathrm{MtM}(t) < H\} + H \cdot \mathbb{1}\{H \leq \mathrm{MtM}(t)\} \end{aligned}$$

- When δ_C is set to zero and there is no minimum collateral transfer amount, the counterparty credit risk vanishes:

$$e^+(t) = 0$$

This last case is interesting, because it gives an indication how to reduce the counterparty risk:

$$H \searrow 0 \text{ or } \delta_C \searrow 0 \Rightarrow e^+(t) \searrow 0$$

In the first panel in Figure 4.10, we have simulated the mark-to-market of a portfolio for a two-year period. In the second panel, we have reported the counterparty exposure when there is no collateral. The other panels show the collateral $C(t)$ and the counterparty exposure $e^+(t)$ for different values of δ_C and H. When there is no margin period of risk, we verify that the exposure is capped at the collateral threshold H in the fourth panel. When the threshold is equal to zero, the counterparty exposure corresponds to the lag effect due to the margin period of risk as illustrated in the sixth panel. The riskier situation corresponds to the combination of the threshold risk and the margin period of risk (eighth panel).

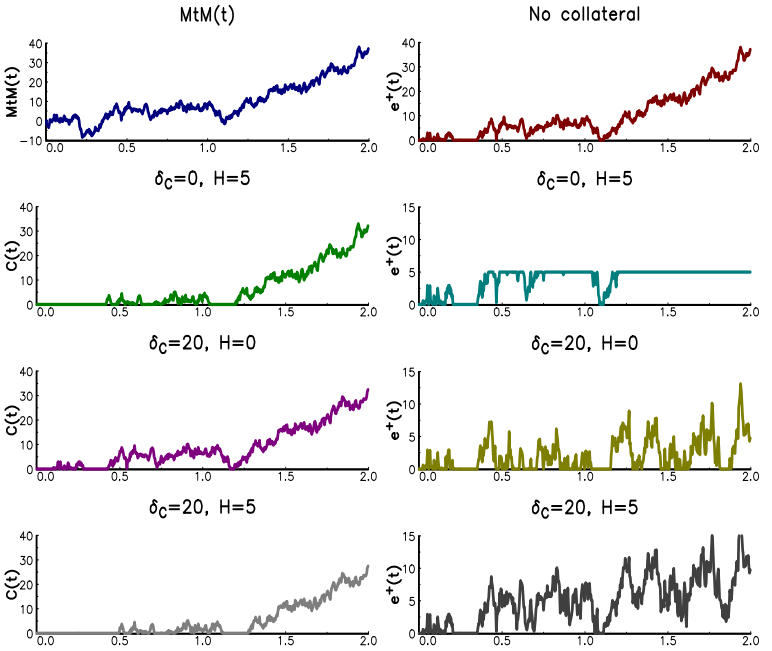


FIGURE 4.10: Impact of collateral on the counterparty exposure

4.3.2 Capital allocation

Taking into account collateral in the CVA computation is relatively straightforward when we use Monte Carlo simulations. In fact, the CVA formula remains the same, only the

computation of the expected positive exposure $\mathrm{EpE}(t)$ is changed. However, as mentioned by Pykhtin and Rosen (2010), the big issue is the allocation of the capital. In Section 2.3 on page 104, we have seen that the capital allocation is given by the Euler allocation principle. Let $\mathcal{R}(w)$ be the risk measure of Portfolio $w = (w_1, \ldots, w_n)$. Under some assumptions, we reiterate that:

$$\mathcal{R}(w) = \sum_{i=1}^{n} \mathcal{RC}_i$$

where $\mathcal{RC}_i$ is the risk contribution of the i^{th} component:

$$\mathcal{RC}_i = w_i \cdot \frac{\partial \mathcal{R}(w)}{\partial w_i}$$

The components can be assets, credits, trading desks, etc. For instance, in the case of credit risk, the IRB formula gives the risk contribution of a loan within a portfolio. In the case of a CVA portfolio, we have:

$$\mathrm{CVA}(w) = (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T -B_0(t)\,\mathrm{EpE}(t;w)\,\mathrm{d}\mathbf{S}_B(t)$$

where $\mathrm{EpE}(t;w)$ is the expected positive exposure with respect to the portfolio w. The Euler allocation principle becomes:

$$\mathrm{CVA}(w) = \sum_{i=1}^{n} \mathrm{CVA}_i(w)$$

where $\mathrm{CVA}_i(w)$ is the CVA risk contribution of the i^{th} component:

$$\mathrm{CVA}_i(w) = (1 - \boldsymbol{\mathcal{R}}_B) \cdot \int_0^T -B_0(t)\,\mathrm{EpE}_i(t;w)\,\mathrm{d}\mathbf{S}_B(t)$$

and $\mathrm{EpE}_i(t;w)$ is the EpE risk contribution of the i^{th} component:

$$\mathrm{EpE}_i(t;w) = w_i \cdot \frac{\partial\,\mathrm{EpE}(t;w)}{\partial w_i}$$

Therefore, the difficulty for computing the CVA risk contribution is to compute the EpE risk contribution.

We consider the portfolio $w = (w_1, \ldots, w_n)$, which is composed of n OTC contracts. The mark-to-market of the portfolio is equal:

$$\mathrm{MtM}(t) = \sum_{i=1}^{n} w_i \cdot \mathrm{MtM}_i(t)$$

where $\mathrm{MtM}_i(t)$ is the mark-to-market for the contract $\mathfrak{C}_i$. In the general case, the counterparty exposure is given by:

$$e^+(t) = \mathrm{MtM}(t) \cdot \mathbb{1}\{0 \leq \mathrm{MtM}(t) < H\} + H \cdot \mathbb{1}\{\mathrm{MtM}(t) \geq H\}$$

If there is no collateral, we have:

$$\begin{aligned} e^+(t) &= \mathrm{MtM}(t) \cdot \mathbb{1}\{\mathrm{MtM}(t) \geq 0\} \\ &= \sum_{i=1}^{n} w_i \cdot \mathrm{MtM}_i(t) \cdot \mathbb{1}\{\mathrm{MtM}(t) \geq 0\} \end{aligned}$$

We deduce that:

$$\frac{\partial \mathbb{E}\left[e^{+}(t)\right]}{\partial w_i} = \mathbb{E}\left[\mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq 0\right\}\right]$$

and:

$$\mathrm{EpE}_i(t;w) = \mathbb{E}\left[w_i \cdot \mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq 0\right\}\right]$$

Computing the EpE (or CVA) risk contribution is then straightforward in this case. In the general case, Pykhtin and Rosen (2010) notice that $\mathrm{EpE}(t;w)$ is not a homogeneous function of degree one because of the second term $\mathbb{E}\left[H \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]$. The idea of these authors is then to allocate the threshold risk to the individual contracts:

$$\mathbb{E}\left[H \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right] = H \cdot \sum_{i=1}^{n} \mathbb{E}\left[\omega_i \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]$$

by choosing an appropriate value of ω_i such that $\sum_{i=1}^{n} \omega_i = 1$. They consider two propositions. Type A Euler allocation is given by:

$$\begin{aligned}\mathrm{EpE}_i(t;w) &= \mathbb{E}\left[w_i \cdot \mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{0 \leq \mathrm{MtM}(t) < H\right\}\right] + \\ &\quad H \cdot \frac{\mathbb{E}\left[\mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right] \cdot \mathbb{E}\left[w_i \cdot \mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]}{\mathbb{E}\left[\mathrm{MtM}(t) \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]}\end{aligned}$$

whereas type B Euler allocation is given by:

$$\begin{aligned}\mathcal{RC}_i &= \mathbb{E}\left[w_i \cdot \mathrm{MtM}_i(t) \cdot \mathbb{1}\left\{0 \leq \mathrm{MtM}(t) < H\right\}\right] + \\ &\quad H \cdot \mathbb{E}\left[\frac{w_i \cdot \mathrm{MtM}_i(t)}{\mathrm{MtM}(t)} \cdot \mathbb{1}\left\{\mathrm{MtM}(t) \geq H\right\}\right]\end{aligned}$$

Pykhtin and Rosen (2010) consider the Gaussian case when the mark-to-market for the contract $\mathfrak{C}_i$ is given by:

$$\mathrm{MtM}_i(t) = \mu_i(t) + \sigma_i(t) X_i$$

where $(X_1, \ldots, X_n) \sim \mathcal{N}(0_n, \rho)$ and $\rho = (\rho_{i,j})$ is the correlation matrix. Let $\mu_w(t)$ and $\sigma_w(t)$ be the expected value and volatility of the portfolio mark-to-market $\mathrm{MtM}(t)$. The authors show that[36] the expected positive exposure is the sum of three components:

$$\mathrm{EpE}(t;w) = \mathrm{EpE}_{\mu}(t;w) + \mathrm{EpE}_{\sigma}(t;w) + \mathrm{EpE}_{H}(t;w)$$

where $\mathrm{EpE}_{\mu}(t;w)$ is the mean component:

$$\mathrm{EpE}_{\mu}(t;w) = \mu_w(t) \cdot \left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right)$$

$\mathrm{EpE}_{\sigma}(t;w)$ is the volatility component:

$$\mathrm{EpE}_{\sigma}(t;w) = \sigma_w(t) \cdot \left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)\right)$$

and $\mathrm{EpE}_{H}(t;w)$ is the collateral threshold component:

$$\mathrm{EpE}_{H}(t;w) = H \cdot \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)$$

[36] See Exercise 4.4.6 on page 303.

We notice that $\mathrm{EpE}_{\mu}(t;w)$ and $\mathrm{EpE}_{H}(t;w)$ are always positive, while $\mathrm{EpE}_{\sigma}(t;w)$ may be positive or negative. When there is no collateral agreement, $\mathrm{EpE}_{H}(t;w)$ is equal to zero and $\mathrm{EpE}(t;w)$ depends on the ratio $\mu_w(t)/\sigma_w(t)$. Concerning the risk contributions, Pykhtin and Rosen (2010) obtain a similar decomposition:

$$\mathrm{EpE}_i(t;w) = \mathrm{EpE}_{\mu,i}(t;w) + \mathrm{EpE}_{\sigma,i}(t;w) + \mathrm{EpE}_{H,i}(t;w)$$

where:

$$\begin{aligned}
\mathrm{EpE}_{\mu,i}(t;w) &= w_i \cdot \mu_i(t) \cdot \left(\Phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right) \\
\mathrm{EpE}_{\sigma,i}(t;w) &= w_i \cdot \gamma_i(t) \cdot \sigma_i(t) \cdot \left(\phi\left(\frac{\mu_w(t)}{\sigma_w(t)}\right) - \phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)\right) \\
\mathrm{EpE}_{H,i}(t;w) &= H \cdot \Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) \cdot \frac{\psi_i}{\psi_w}
\end{aligned}$$

$\gamma_i(t) = \sigma(t)^{-1} \sum_{j=1}^{n} w_j \cdot \rho_{i,j} \cdot \sigma_j(t)$ and:

$$\frac{\psi_i}{\psi_w} = \frac{w_i \cdot \mu_i \cdot (t)\, \Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) + w_i \cdot \gamma_i(t) \cdot \sigma_i(t) \cdot \phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)}{\mu_w(t) \cdot \Phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right) + \sigma_w(t) \cdot \phi\left(\frac{\mu_w(t)-H}{\sigma_w(t)}\right)}$$

Example 50 *We consider a portfolio of two contracts* $\mathfrak{C}_1$ *and* $\mathfrak{C}_2$ *with the following characteristics:* $\mu_1(t) = \$1$ *mn,* $\sigma_1(t) = \$1$ *mn,* $\mu_2(t) = \$1$ *mn,* $\sigma_2(t) = \$1$ *mn and* $\rho_{1,2} = 0\%$.

We first calculate the expected positive exposure $\mathrm{EpE}(t;w)$ when we change the value of $\mu_2(t)$ and there is no collateral agreement. Results are given in Figure 4.11. In the first panel, we observe that $\mathrm{EpE}(t;w)$ increases with respect to $\mu_2(t)$. We notice that the mean component is the most important contributor when the expected value of the portfolio mark-to-market is high and positive[37]:

$$\frac{\mu_w(t)}{\sigma_w(t)} \to \infty \Rightarrow \begin{cases} \mathrm{EpE}_{\mu}(t;w) \to \mathrm{EpE}(t;w) \\ \mathrm{EpE}_{\sigma}(t;w) \to 0 \end{cases}$$

The risk contribution $\mathrm{EpE}_1(t;w)$ and $\mathrm{EpE}_2(t;w)$ are given in the second panel in Figure 4.11. The risk contribution of the second contract is negative when $\mu_2(t)$ is less than -1. This illustrate the diversification effect, implying that some trades can negatively contributes to the CVA risk. This is why the concept of netting sets is important when computing the CVA capital charge. In Figure 4.12, we have done the same exercise when we consider different values of the correlation $\rho_{1,2}$. We observe that the impact of this parameter is not very important except when the correlation is negative. The reason is that the correlation matrix has an impact on the volatility $\sigma_w(t)$ of the portfolio mark-to-market, but not on the expected value $\mu_w(t)$. We now consider that $\mu_1(t) = \mu_2(t) = 1$, $\sigma_1(t) = \sigma_2(t) = 1$ and $\rho_{1,2} = 0$. In Figure 4.13, we analyze the impact of the collateral threshold H. We notice that having a tighter collateral agreement (or a lower threshold H) allows to reduce the counterparty exposure. However, this reduction is not monotonous. It is very important when H is close to zero, but there is no impact when H is large.

[37] In this limit case, we obtain:

$$\mathrm{EpE}(t;w) = \mu_w(t) = \sum_{i=1}^{n} w_i \mu_i(t)$$

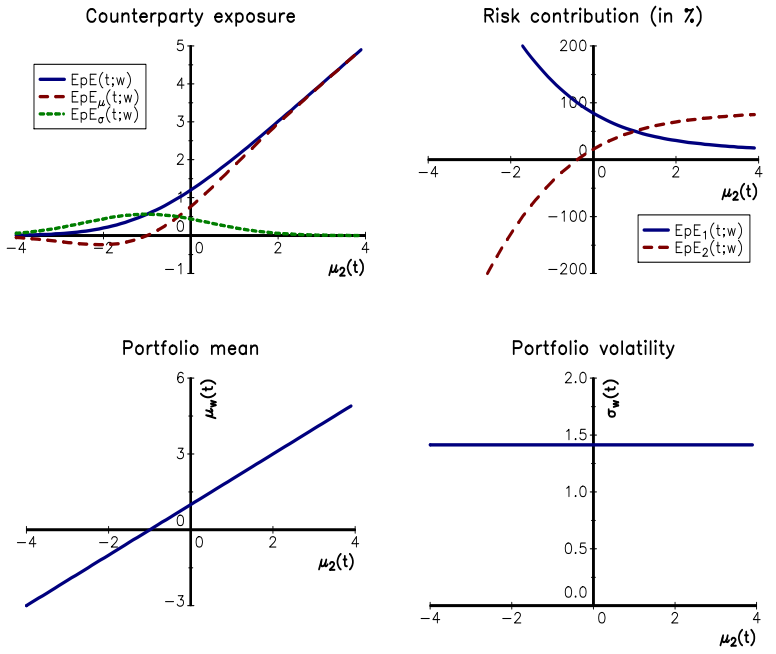


FIGURE 4.11: Impact of $\mu_i(t)/\sigma_i(t)$ on the counterparty exposure

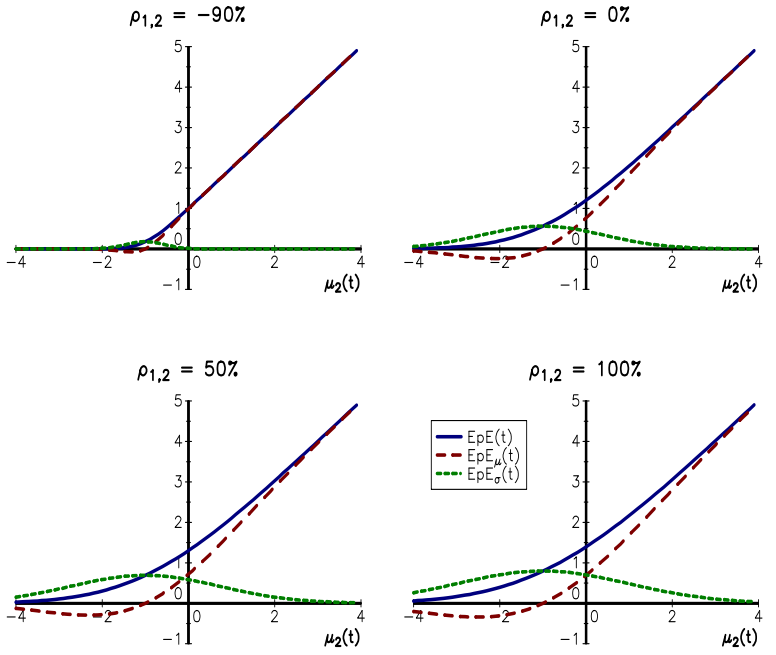


FIGURE 4.12: Impact of the correlation on the counterparty exposure

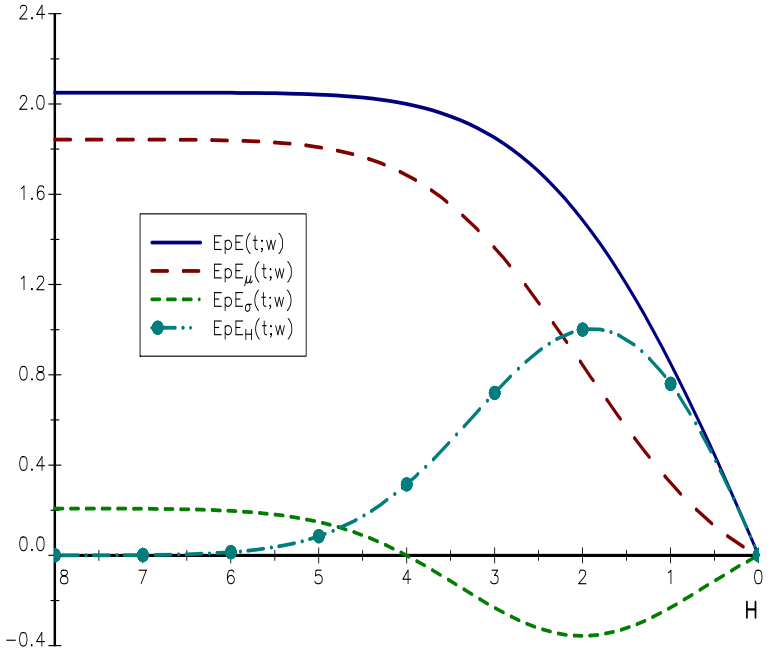


FIGURE 4.13: Decomposition of the counterparty exposure when there is a collateral agreement

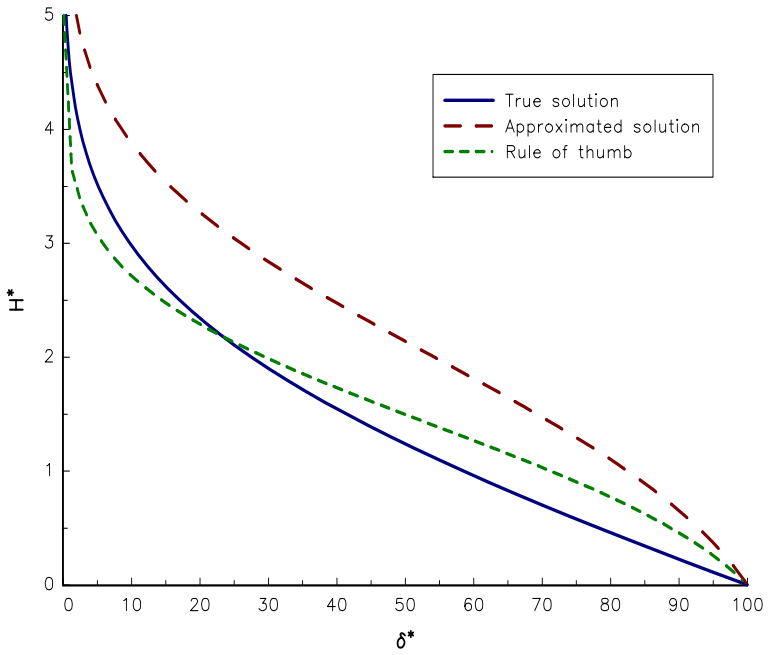


FIGURE 4.14: Optimal collateral threshold

The impact of the threshold can be measured by the ratio:

$$\delta(H) = \frac{\text{EpE}(t;w,\infty) - \text{EpE}(t;w,H)}{\text{EpE}(t;w,\infty)}$$

where $\text{EpE}(t;w,H)$ is the expected positive exposure for a given threshold H. If we would like to reduce the counterparty exposure by $\delta^\star$, we have to solve the non-linear equation $\delta(H) = \delta^\star$ in order to find the optimal value $H^\star$. We can also approximate $\text{EpE}(t;w,H)$ by its mean contribution:

$$\begin{aligned}\delta(H) &\approx \delta_\mu(H) \\ &= \frac{\mu_w(t)\cdot\Phi(\zeta_H)}{\text{EpE}_\mu(t;w,\infty)}\end{aligned}$$

In this case, the solution of the non-linear equation $\delta_\mu(H) = \delta^\star$ is equal to[38]:

$$H^\star = \mu_w(t) - \sigma_w(t)\cdot\Phi^{-1}\left(\frac{\text{EpE}_\mu(t;w,\infty)}{\mu_w(t)}\cdot\delta^\star\right)$$

The computation of $H^\star$ is then straightforward since we have only to calculate $\mu_w(t)$, $\sigma_w(t)$ and the mean contribution $\text{EpE}_\mu(t;w,\infty)$ when there is no collateral agreement. However, the value of $H^\star$ is overestimated because $\text{EpE}_\mu(t;w,H)$ is lower than $\text{EpE}(t;w,H)$. A rule of thumb is then to adjust the solution $H^\star$ by a factor[39], which is generally equal to 0.75. In Figure 4.14, we have represented the optimal collateral threshold $H^\star$ for the previous example.

4.4 Exercises

4.4.1 Impact of netting agreements in counterparty credit risk

The table below gives the current mark-to-market of 7 OTC contracts between Bank A and Bank B:

	Equity			Fixed income		FX	
	$\mathfrak{C}_1$	$\mathfrak{C}_2$	$\mathfrak{C}_3$	$\mathfrak{C}_4$	$\mathfrak{C}_5$	$\mathfrak{C}_6$	$\mathfrak{C}_7$
A	+10	−5	+6	+17	−5	−5	+1
B	−11	+6	−3	−12	+9	+5	+1

The table should be read as follows: Bank A has a mark-to-market equal to +10 for the contract $\mathfrak{C}_1$ whereas Bank B has a mark-to-market equal to −11 for the same contract, Bank A has a mark-to-market equal to −5 for the contract $\mathfrak{C}_2$ whereas Bank B has a mark-to-market equal to +6 for the same contract, etc.

1. (a) Explain why there are differences between the MtM values of a same OTC contract.
   (b) Calculate the exposure at default of Bank A.
   (c) Same question if there is a global netting agreement.
   (d) Same question if the netting agreement only concerns equity products.

[38] The solution $H^\star$ can be viewed as a quantile of the probability distribution of the portfolio mark-to-market: $\text{MtM}(t) \sim \mathcal{N}\left(\mu_w(t), \sigma_w^2(t)\right)$.

[39] The underlying idea is that $\text{EpE}_\mu(t;w,H) \approx 75\% \cdot \text{EpE}(t;w,H)$.

2. In the following, we measure the impact of netting agreements on the exposure at default.

   (a) We consider an OTC contract $\mathfrak{C}$ between Bank A and Bank B. The mark-to-market $\mathrm{MtM}_1(t)$ of Bank A for the contract $\mathfrak{C}$ is defined as follows:

   $$\mathrm{MtM}_1(t) = x_1 + \sigma_1 W_1(t)$$

   where $W_1(t)$ is a Brownian motion. Calculate the potential future exposure of Bank A.

   (b) We consider a second OTC contract between Bank A and Bank B. The mark-to-market is also given by the following expression:

   $$\mathrm{MtM}_2(t) = x_2 + \sigma_2 W_2(t)$$

   where $W_2(t)$ is a second Brownian motion that is correlated with $W_1(t)$. Let ρ be this correlation such that $\mathbb{E}[W_1(t) W_2(t)] = \rho t$. Calculate the expected exposure of bank A if there is no netting agreement.

   (c) Same question when there is a global netting agreement between Bank A and Bank B.

   (d) Comment on these results.

4.4.2 Calculation of the effective expected positive exposure

We denote by $e(t)$ the potential future exposure of an OTC contract with maturity T. The current date is set to $t = 0$.

1. Define the concepts of peak exposure $\mathrm{PE}_\alpha(t)$, maximum peak exposure $\mathrm{MPE}_\alpha(0;t)$, expected exposure $\mathrm{EE}(t)$, expected positive exposure $\mathrm{EPE}(0;t)$, effective expected exposure $\mathrm{EEE}(t)$ and effective expected positive exposure $\mathrm{EEPE}(0;t)$.

2. Calculate these different quantities when the potential future exposure is $e(t) = \sigma \cdot \sqrt{t} \cdot X$ where $X \sim \mathcal{U}_{[0,1]}$.

3. Same question when $e(t) = \exp\left(\sigma \cdot \sqrt{t} \cdot X\right)$ where $X \sim \mathcal{N}(0,1)$.

4. Same question when $e(t) = \sigma \cdot \left(t^3 - \frac{7}{3}Tt^2 + \frac{4}{3}T^2 t\right) \cdot X$ where $X \sim \mathcal{U}_{[0,1]}$.

5. Same question when $e(t) = \sigma \cdot \sqrt{t} \cdot X$ where X is a random variable defined on $[0,1]$ with the following probability density function[40]:

   $$f(x) = \frac{x^a}{a+1}$$

6. Comment on these results.

[40] We assume that $a > 0$.

4.4.3 Calculation of the capital charge for counterparty credit risk

We denote by $e(t)$ the potential future exposure of an OTC contract with maturity T. The current date is set to $t = 0$. Let N and σ be the notional and the volatility of the underlying contract. We assume that $e(t) = N \cdot \sigma \cdot \sqrt{t} \cdot X$ where $0 \leq X \leq 1$, $\Pr\{X \leq x\} = x^\gamma$ and $\gamma > 0$.

1. Calculate the peak exposure $\text{PE}_\alpha(t)$, the expected exposure $\text{EE}(t)$ and the effective expected positive exposure $\text{EEPE}(0;t)$.
2. The bank manages the credit risk with the foundation IRB approach and the counterparty credit risk with an internal model. We consider an OTC contract with the following parameters: N is equal to \$3 mn, the maturity T is one year, the volatility σ is set to 20% and γ is estimated at 2.
   (a) Calculate the exposure at default EAD knowing that the bank uses the regulatory value for the parameter α.
   (b) The default probability of the counterparty is estimated at 1%. Calculate then the capital charge for counterparty credit risk of this OTC contract[41].

4.4.4 Calculation of CVA and DVA measures

We consider an OTC contract with maturity T between Bank A and Bank B. We denote by $\text{MtM}(t)$ the risk-free mark-to-market of Bank A. The current date is set to $t = 0$ and we assume that:

$$\text{MtM}(t) = N \cdot \sigma \cdot \sqrt{t} \cdot X$$

where N is the notional of the OTC contract, σ is the volatility of the underlying asset and X is a random variable, which is defined on the support $[-1, 1]$ and whose density function is:

$$f(x) = \frac{1}{2}$$

1. Define the concept of positive exposure $e^+(t)$. Show that the cumulative distribution function $\mathbf{F}_{[0,t]}$ of $e^+(t)$ has the following expression:

$$\mathbf{F}_{[0,t]}(x) = \mathbb{1}\left\{0 \leq x \leq \sigma\sqrt{t}\right\} \cdot \left(\frac{1}{2} + \frac{x}{2 \cdot N \cdot \sigma \cdot \sqrt{t}}\right)$$

   where $\mathbf{F}_{[0,t]}(x) = 0$ if $x \leq 0$ and $\mathbf{F}_{[0,t]}(x) = 1$ if $x \geq \sigma\sqrt{t}$.
2. Deduce the value of the expected positive exposure $\text{EpE}(t)$.
3. We note $\boldsymbol{\mathcal{R}}_B$ the fixed and constant recovery rate of Bank B. Give the mathematical expression of the CVA.
4. By using the definition of the lower incomplete gamma function $\gamma(s, x)$, show that the CVA is equal to:

$$\text{CVA} = \frac{N \cdot (1 - \boldsymbol{\mathcal{R}}_B) \cdot \sigma \cdot \gamma\left(\frac{3}{2}, \lambda_B T\right)}{4\sqrt{\lambda_B}}$$

   when the default time of Bank B is exponential with parameter λ_B and interest rates are equal to zero.

[41] We will take a value of 70% for the LGD parameter and a value of 20% for the default correlation. We can also use the approximations $-1.06 \approx -1$ and $\Phi(-1) \approx 16\%$.

5. Comment on this result.
6. By assuming that the default time of Bank A is exponential with parameter λ_A, deduce the value of the DVA without additional computations.

4.4.5 Approximation of the CVA for an interest rate swap

This exercise is based on the results of Syrkin and Shirazi (2015).

1. Calculate $\text{EpE}(t) = \mathbb{E}[\max(\text{MtM}(t), 0)]$ when the mark-to-market is equal to $\text{MtM}(t) = Ae^X - B$ and $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
2. We define the mark-to-market of the interest rate swap as follows:

$$\text{MtM}(t) = N \int_t^T f(t,T) B_t(s) \, \mathrm{d}s - N \int_t^T f(0,T) B_t(s) \, \mathrm{d}s$$

   where N and T are the notional and the maturity of the swap, and $f(t,T)$ is the instantaneous forward rate. Comment on this formulation. By assuming that $f(t,T)$ follows a geometric Brownian motion:

$$\mathrm{d}f(t,T) = \mu f(t,T) \, \mathrm{d}t + \sigma f(t,T) \, \mathrm{d}W(t)$$

   and the yield curve is flat – $B_t(s) = e^{-r(s-t)}$, calculate the value of the mark-to-market. Deduce the confidence interval of $\text{MtM}(t)$ with a confidence level α.
3. Calculate the expected mark-to-market and the expected counterparty exposure.
4. Give the expression of the CVA at time t if we assume that the default time is exponentially distributed: $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$.
5. Retrieve the approximation of the CVA found by Syrkin and Shirazi (2015).
6. We consider the following numerical values: $N = 1000$, $f(0,T) = 5\%$, $\mu = 2\%$, $\sigma = 25\%$, $T = 10$ years, $\lambda = 1\%$ and $\boldsymbol{\mathcal{R}} = 50\%$.
   (a) Calculate the 90% confidence interval of $\text{MtM}(t)$.
   (b) Compare the time profile of $\text{EpE}(t)$ and $\mathbb{E}[\text{MtM}(t)]$.
   (c) Compare the time profile of $\text{CVA}(t)$ and its approximation.
   (d) What do you think about the numerical value of μ?

4.4.6 Risk contribution of CVA with collateral

This exercise is based on the results of Pykhtin and Rosen (2010).

1. We consider the portfolio $w = (w_1, \ldots, w_n)$, which is composed of n OTC contracts. We assume that the mark-to-market for the contract $\mathfrak{C}_i$ is given by:

$$\text{MtM}_i(t) = \mu_i(t) + \sigma_i(t) X_i$$

   where $X_i \sim \mathcal{N}(0,1)$. Determine the probability distribution of the portfolio mark-to-market:

$$\text{MtM}(t) = \sum_{i=1}^n w_i \cdot \text{MtM}_i(t)$$

   when $(X_1, \ldots, X_n) \sim \mathcal{N}(0_n, \rho)$ and $\rho = (\rho_{i,j})$ is the correlation matrix.

2. Calculate the correlation $\gamma_i(t)$ between $\mathrm{MtM}_i(t)$ and $\mathrm{MtM}(t)$.

3. Calculate the expected value of the counterparty exposure $e^+(t) = \max(\mathrm{MtM}(t) - C(t), 0)$ when the collateral value is given by $C(t) = \max(\mathrm{MtM}(t) - H, 0)$.

4. We consider the case where there is no collateral: $C(t) = 0$. What is the implicit value of H? Deduce the expression of $\mathrm{EpE}(t; w) = \mathbb{E}[e^+(t)]$. Calculate the risk contribution $\mathcal{RC}_i$ of the contract $\mathfrak{C}_i$. Show that $\mathrm{EpE}(t; w)$ satisfies the Euler allocation principle.

5. We consider the case where there is a collateral: $C(t) \neq 0$. Calculate the risk contribution $\mathcal{RC}_i$ of the contract $\mathfrak{C}_i$. Demonstrate that:

$$\sum_{i=1}^{n} \mathcal{RC}_i = \mathrm{EpE}(t; w) - H \cdot \Phi\left(\frac{\mu_w(t) - H}{\sigma_w(t)}\right)$$

   where $\mu_w(t)$ and $\sigma_w(t)$ are the expected value and volatility of $\mathrm{MtM}(t)$. Comment on this result.

6. Find the risk contribution $\mathcal{RC}_i$ of type A Euler allocation.

7. Find the risk contribution $\mathcal{RC}_i$ of type B Euler allocation.

8. We consider the Merton approach for modeling the default time $\boldsymbol{\tau}$ of the counterparty:

$$X_i = \varrho_i X_B + \sqrt{1 - \varrho_i^2}\,\eta_i$$

   where $X_B \sim \mathcal{N}(0,1)$ and the idiosyncratic risk $\eta_i \sim \mathcal{N}(0,1)$ are independent. Calculate the correlation $\varrho_w(t)$ between $\mathrm{MtM}(t)$ and X_B. Deduce the relationship between $\mathrm{MtM}(t)$ and X_B.

9. Let $B(t) = \Phi^{-1}(1 - \mathbf{S}(t))$ be the default barrier and $\mathbf{S}(t)$ the survival function of the counterparty. How to compute the conditional counterparty exposure $\mathbb{E}[e^+(t) \mid \boldsymbol{\tau} = t]$ and the corresponding risk contribution $\mathcal{RC}_i$? Give their expressions.

Chapter 5

Operational Risk

The integration of operational risk into the Basel II Accord was a long process because of the hostile reaction from the banking sector. At the end of the 1990s, the risk of operational losses was perceived as relatively minor. However, some events had shown that it was not the case. The most famous example was the bankruptcy of the Barings Bank in 1995. The loss of \$1.3 bn was due to a huge position of the trader Nick Leeson in futures contracts without authorization. Other examples included the money laundering in Banco Ambrosiano Vatican Bank (1983), the rogue trading in Sumitomo Bank (1996), the headquarter fire of Crédit Lyonnais (1996), etc. Since the publication of the CP2 in January 2001, the position of banks has significantly changed and operational risk is today perceived as a major risk for the banking industry. Management of operational risk has been strengthened, with the creation of dedicated risk management units, the appointment of compliance officers and the launch of anti-money laundering programs.

5.1 Definition of operational risk

The Basel Committee defines the operational risk in the following way:

> "*Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk*" (BCBS, 2006, page 144).

The operational risk covers then all the losses of the bank that cannot be attributed to market and credit risk. Nevertheless, losses that result from strategic decisions are not taken into account. An example is the purchase of a software or an information system, which is not relevant for the firm. Losses due to reputational risk are also excluded from the definition of operational risk. They are generally caused by an event, which is related to another risk. The difficulty is to measure the indirect loss of such events in terms of business. For instance, if we consider the diesel emissions scandal of Volkswagen, we can estimate the losses due to the recall of cars, class action lawsuits and potential fines. However, it is impossible to know what the impact of this event will be on the future sales and the market share of Volkswagen.

In order to better understand the concept of operational risk, we give here the loss even type classification adopted by the Basel Committee:

1. Internal fraud ("*losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/discrimination events, which involves at least one internal party*")
   (a) Unauthorized activity
   (b) Theft and fraud

2. External fraud ("*losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party*")
   (a) Theft and fraud
   (b) Systems security
3. Employment practices and workplace safety ("*losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity/discrimination events*")
   (a) Employee relations
   (b) Safe environment
   (c) Diversity & discrimination
4. Clients, products & business practices ("*losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements), or from the nature or design of a product*")
   (a) Suitability, disclosure & fiduciary
   (b) Improper business or market practices
   (c) Product flaws
   (d) Selection, sponsorship & exposure
   (e) Advisory activities
5. Damage to physical assets ("*losses arising from loss or damage to physical assets from natural disaster or other events*")
   (a) Disasters and other events
6. Business disruption and system failures ("*losses arising from disruption of business or system failures*")
   (a) Systems
7. Execution, delivery & process management ("*losses from failed transaction processing or process management, from relations with trade counterparties and vendors*")
   (a) Transaction capture, execution & maintenance
   (b) Monitoring and reporting
   (c) Customer intake and documentation
   (d) Customer/client account management
   (e) Trade counterparties
   (f) Vendors & suppliers

This is a long list of loss types, because the banking industry has been a fertile ground for operational risks. We have already cited some well-know operational losses before the crisis.

In 2009, the Basel Committee has published the results of a loss data collection exercise. For this LDCE, 119 banks submitted a total of 10.6 million internal losses with an overall loss amount of €59.6 bn. The largest 20 losses represented a total of €17.6 bn. In Table 5.1, we have reported statistics of the loss data, when the loss is larger than €20 000. For each year, we indicate the number n_L of losses, the total loss amount L and the number n_B of reporting banks. Each bank experienced more than 300 losses larger than €20 000 per year on average. We also notice that these losses represented about 90% of the overall loss amount.

TABLE 5.1: Internal losses larger than €20 000 per year

Year	pre 2002	2002	2003	2004	2005	2006	2007
n_L	14 017	10 216	13 691	22 152	33 216	36 386	36 622
L (in € bn)	3.8	12.1	4.6	7.2	9.7	7.4	7.9
n_B	24	35	55	68	108	115	117

Source: BCBS (2009d).

Since 2008, operational risk has dramatically increased. For instance, rogue trading has impacted many banks and the magnitude of these unauthorized trading losses is much higher than before[1]. The Libor interest rate manipulation scandal led to very large fines ($2.5 bn for Deutsche Bank, $1 bn for Rabobank, $545 mn for UBS, etc.). In May 2015, six banks (Bank of America, Barclays, Citigroup, J.P. Morgan, UBS and RBS) agreed to pay fines totaling $5.6 bn in the case of the forex scandal[2]. The anti-money laundering controls led BNP Paribas to pay a fine of $8.9 bn in June 2014 to the US federal government. In this context, operational risk, and more specifically compliance and legal risk, is a major concern for banks. It is an expansive risk, because of the direct losses, but also because of the indirect costs induced by the proliferation of internal controls and security infrastructure[3].

Remark 60 *Operational risk is not limited to the banking sector. Other financial sectors have been impacted by such risk. The most famous example is the Ponzi scheme organized by Bernard Madoff, which caused a loss of* $50 *bn to his investors.*

5.2 Basel approaches for calculating the regulatory capital

In this section, we present the three approaches described in the Basel II framework in order to calculate the capital charge for operational risk:

1. the basic indicator approach (BIA);
2. the standardized approach (TSA);
3. and advanced measurement approaches (AMA).

[1] We can cite Société Générale in 2008 ($7.2 bn), Morgan Stanley in 2008 ($9.0 bn), BPCE in 2008 ($1.1 bn), UBS in 2011 ($2 bn) and JPMorgan Chase in 2012 ($5.8 bn).

[2] The Libor scandal was a series of fraudulent actions connected to the Libor (London Interbank Offered Rate), while the forex scandal concerns several banks, which have manipulated exchange rates via electronic chatrooms in which traders discussed the trades they planned to do.

[3] A typical example of expansive cost is the risk of cyber attacks.

We also present the Basel III framework of the standardized approach for measuring operational risk capital with effect from January 2022.

5.2.1 The basic indicator approach

The basic indicator approach is the simplest method for calculating the operational risk capital requirement. In this case, the capital charge is a fixed percentage of annual gross income:

$$\mathcal{K} = \alpha \cdot \overline{\mathrm{GI}}$$

where α is set equal to 15% and $\overline{\mathrm{GI}}$ is the average of the positive gross income over the previous three years:

$$\overline{\mathrm{GI}} = \frac{\max(\mathrm{GI}_{t-1}, 0) + \max(\mathrm{GI}_{t-2}, 0) + \max(\mathrm{GI}_{t-3}, 0)}{\sum_{k=1}^{3} \mathbb{1}\{\mathrm{GI}_{t-k} > 0\}}$$

In this approach, the capital charge is related to the financial results of the bank, but not to its risk exposure.

5.2.2 The standardized approach

The standardized approach is an extended version of the previous method. In this case, the bank is divided into eight business lines, which are given in Table 5.2. The bank then calculates the capital charge for each business line:

$$\mathcal{K}_{j,t} = \beta_j \cdot \mathrm{GI}_{j,t}$$

where β_j and $\mathrm{GI}_{j,t}$ are a fixed percentage[4] and the gross income corresponding to the j^{th} business line. The total capital charge is the three-year average of the sum of all the capital charges:

$$\mathcal{K} = \frac{1}{3} \sum_{k=1}^{3} \max\left(\sum_{j=1}^{8} \mathcal{K}_{j,t-k}, 0 \right)$$

We notice that a negative capital charge in one business line may offset positive capital charges in other business lines. If the values of gross income are all positive, the total capital charge becomes:

$$\begin{aligned} \mathcal{K} &= \frac{1}{3} \sum_{k=1}^{3} \sum_{j=1}^{8} \beta_j \cdot \mathrm{GI}_{j,t-k} \\ &= \sum_{j=1}^{8} \beta_j \cdot \overline{\mathrm{GI}}_j \end{aligned}$$

where $\overline{\mathrm{GI}}_j$ is the average gross income over the previous three years of the j^{th} business line.

Example 51 *We consider Bank A, whose activity is mainly driven by retail banking and asset management. We compare it with Bank B, which is more focused on corporate finance. We assume that the two banks are only composed of four business lines: corporate finance,*

[4] The values taken by the beta coefficient are reported in Table 5.2.

TABLE 5.2: Mapping of business lines for operational risk

Level 1	Level 2	β_j
Corporate Finance†	Corporate Finance Municipal/Government Finance Merchant Banking Advisory Services	18%
Trading & Sales‡	Sales Market Making Proprietary Positions Treasury	18%
Retail Banking	Retail Banking Private Banking Card Services	12%
Commercial Banking♮	Commercial Banking	12%
Payment & Settlement	External Clients	18%
Agency Services	Custody Corporate Agency Corporate Trust	15%
Asset Management	Discretionary Fund Management Non-Discretionary Fund Management	12%
Retail Brokerage	Retail Brokerage	12%

†Mergers and acquisitions, underwriting, securitization, syndications, IPO, debt placements.
‡Buying and selling of securities and derivatives, own position securities, lending and repos, brokerage.
♮Project finance, real estate, export finance, trade finance, factoring, leasing, lending, guarantees, bills of exchange.

retail banking, agency services and asset management. The gross income expressed in $ mn for the last three years is given below:

Business line	*Bank A*			*Bank B*		
	$t-1$	$t-2$	$t-3$	$t-1$	$t-2$	$t-3$
Corporate finance	10	15	−30	200	300	150
Retail banking	250	230	205	50	45	−30
Agency services	10	10	12			
Asset management	70	65	72	12	8	−4

For Bank A, we obtain $\text{GI}_{t-1} = 340$, $\text{GI}_{t-2} = 320$ and $\text{GI}_{t-3} = 259$. The average gross income is then equal to 306.33, implying that the BIA capital charge $\boldsymbol{\mathcal{K}}_A^{\text{BIA}}$ is equal to \$45.95 mn. If we consider Bank B, the required capital $\boldsymbol{\mathcal{K}}_B^{\text{BIA}}$ is lower and equal to \$36.55 mn. In the case of the standardized approach, the beta coefficients are respectively equal to 18%, 12%, 15% and 12%. We deduce that:

$$\begin{aligned}\boldsymbol{\mathcal{K}}_A^{\text{TSA}} &= \frac{1}{3} \times \big(\max\left(18\% \times 10 + 12\% \times 250 + 15\% \times 10 + 12\% \times 70, 0\right) + \\ &\quad \max\left(18\% \times 15 + 12\% \times 230 + 15\% \times 10 + 12\% \times 65, 0\right) + \\ &\quad \max\left(-18\% \times 30 + 12\% \times 205 + 15\% \times 12 + 12\% \times 72, 0\right)\big) \\ &= \$36.98 \text{ mn}\end{aligned}$$

We also have $\mathcal{K}_B^{\text{TSA}} = \$42.24$ mn. We notice that $\mathcal{K}_A^{\text{BIA}} > \mathcal{K}_A^{\text{TSA}}$ and $\mathcal{K}_B^{\text{BIA}} < \mathcal{K}_B^{\text{TSA}}$. Bank A has a lower capital charge when using TSA instead of BIA, because it is more exposed to low-risk business lines (retail banking and asset management). For Bank B, it is the contrary because its main exposure concerns high-risk business lines (corporate finance). However, if we assume that the gross income of the corporate finance for Bank B at time $t-3$ is equal to -150 instead of $+150$, we obtain $\mathcal{K}_B^{\text{BIA}} = \$46.13$ mn and $\mathcal{K}_B^{\text{TSA}} = \$34.60$ mn. In this case, the TSA approach is favorable, because the gross income at time $t-3$ is negative implying that the capital contribution at time $t-3$ is equal to zero.

Contrary to the basic indicator approach that requires no criteria to be used, banks must satisfy a list of qualifying criteria for the standardized approach. For instance, the board of directors must be actively involved in the oversight of the operational risk management framework and each business line must have sufficient resources to manage operational risk. International active banks must also collect operational losses and use this information for taking appropriate action.

5.2.3 Advanced measurement approaches

Like the internal model-based approach for market risk, the AMA method is defined by certain criteria without refereing to a specific statistical model:

- The capital charge should cover the one-year operational loss at the 99.9% confidence level. It corresponds to the sum of expected loss (EL) and unexpected loss (UL).
- The model must be estimated using a minimum five-year observation period of internal loss data, and capture tail loss events by considering for example external loss data when it is needed. It must also include scenario analysis and factors reflecting internal control systems.
- The risk measurement system must be sufficiently granular to capture the main operational risk factors. By default, the operational risk of the bank must be divided into the 8 business lines and the 7 event types. For each cell of the matrix, the model must estimate the loss distribution and may use correlations to perform the aggregation.
- The allocation of economic capital across business lines must create incentives to improve operational risk management.
- The model can incorporate the risk mitigation impact of insurance, which is limited to 20% of the total operational risk capital charge.

The validation of the AMA model does not only concern the measurement aspects, but also the soundness of the entire operational risk management system. This concerns governance of operational risk, dedicated resources, management structure, risk maps and key risk indicators (KRI), notification and action procedures, emergency and crisis management, business continuity and disaster recovery plans.

In order to better understand the challenges of an internal model, we have reported in Table 5.3 the distribution of annualized loss amounts by business line and event type obtained with the 2008 loss data collection exercise. We first notice an heterogeneity between business lines. For instance, losses were mainly concentrated in the fourth event type (clients, products & business practices) for the corporate finance business line (93.7%) and the seventh event type (execution, delivery & process management) for the payment & settlement business line (76.4%). On average, these two event types represented more than 75% of the total loss amount. In contrast, fifth and sixth event types (damage to physical

assets, business disruption and system failures) had a small contribution close to 1%. We also notice that operational losses mainly affected retail banking, followed by corporate finance and trading & sales. One of the issues is that this picture of operational risk is no longer valid after 2008 with the increase of losses in trading & sales, but also in payment & settlement. The nature of operational risk changes over time, which is a big challenge to build an internal model to calculate the required capital.

TABLE 5.3: Distribution of annualized operational losses (in %)

Business line	Event type							All
	1	2	3	4	5	6	7	
Corporate Finance	0.2	0.1	0.6	93.7	0.0	0.0	5.4	28.0
Trading & Sales	11.0	0.3	2.3	29.0	0.2	1.8	55.3	13.6
Retail Banking	6.3	19.4	9.8	40.4	1.1	1.5	21.4	32.0
Commercial Banking	11.4	15.2	3.1	35.5	0.4	1.7	32.6	7.6
Payment & Settlement	2.8	7.1	0.9	7.3	3.2	2.3	76.4	2.6
Agency Services	1.0	3.2	0.7	36.0	18.2	6.0	35.0	2.6
Asset Management	11.1	1.0	2.5	30.8	0.3	1.5	52.8	2.5
Retail Brokerage	18.1	1.4	6.3	59.5	0.1	0.2	14.4	5.1
Unallocated	6.5	2.8	28.4	28.3	6.5	1.3	26.2	6.0
All	6.1	8.0	6.0	52.4	1.4	1.2	24.9	100.0

Source: BCBS (2009d).

5.2.4 Basel III approach

From January 2022, the standardized measurement approach (SMA) will replace the three approaches of the Basel II framework. The SMA is based on three components: the business indicator (BI), the business indicator component (BIC) and the internal loss multiplier (ILM). The business indicator is a proxy of the operational risk:

$$\text{BI} = \text{ILDC} + \text{SC} + \text{FC}$$

where ILDC is the interest, leases and dividends component, SC is the services component and FC is the financial component. The underlying idea is to list the main activities that generate operational risk:

$$\begin{cases} \text{ILDC} = \min\left(|\text{INC} - \text{EXP}|, 2.25\% \cdot \text{IRE}\right) + \text{DIV} \\ \text{SC} = \max\left(\text{OI}, \text{OE}\right) + \max\left(\text{FI}, \text{FE}\right) \\ \text{FC} = |\Pi_{\text{TB}}| + |\Pi_{\text{BB}}| \end{cases}$$

where INC represents the interest income, EXP the interest expense, IRE the interest earning assets, DIV the dividend income, OI the other operating income, OE the other operating expense, FI the fee income, FE the fee expense, Π_{TB} the net P&L of the trading book and Π_{BB} the net P&L of the banking book. All these variables are calculated as the average over the last three years. We can draw a parallel between the business indicator components and the TSA components. For example, ILDC concerns corporate finance, retail banking, commercial banking, SC is related to payment & settlement, agency services, asset management, retail brokerage, while FC mainly corresponds to trading & sales. Once the BI is calculated and expressed in \$ bn, we determine the business indicator component, which is given by:

$$\text{BIC} = 12\% \cdot \min\left(\text{BI}, \$1 \text{ bn}\right) + 15\% \cdot \min\left(\text{BI} - 1, \$30 \text{ bn}\right) + 18\% \cdot \min\left(\text{BI} - 30\right)^{+}$$

The BIC formula recalls the basic indicator approach of Basel II, but it introduces a marginal weight by BI tranches. Finally, the bank has to compute the internal loss multiplier, which is defined as:

$$\text{ILM} = \ln\left(e^1 - 1 + \left(\frac{15 \cdot \bar{L}}{\text{BIC}}\right)^{0.8}\right)$$

where $\bar{L}$ is the average annual operational risk losses over the last 10 years. ILM can be lower or greater than one, depending on the value of $\bar{L}$:

$$\begin{cases} \text{ILM} < 1 \Leftrightarrow \bar{L} < \text{BIC}/15 \\ \text{ILM} = 1 \Leftrightarrow \bar{L} = \text{BIC}/15 \\ \text{ILM} > 1 \Leftrightarrow \bar{L} > \text{BIC}/15 \end{cases}$$

The capital charge for the operational risk is then equal to[5]:

$$\boldsymbol{\mathcal{K}} = \text{ILM} \cdot \text{BIC}$$

Remark 61 *The SMA of the Basel III framework may be viewed as a mix of the three approaches of the Basel II framework: BIA, TSA and AMA. Indeed, SMA is clearly a modified version of BIA by considering a basic indicator based on sources of operational risk. In this case, the business indicator can be related to TSA. Finally, the introduction of the* ILM *coefficient is a way to consider a more sensitive approach based on internal losses, which is the basic component of AMA.*

5.3 Loss distribution approach

Although the Basel Committee does not advocate any particular method for the AMA method in the Basel II framework, the loss distribution approach (LDA) is the recognized standard model for calculating the capital charge. This model is not specific to operational risk because it was developed in the case of the collective risk theory at the beginning of 1900s. However, operational risk presents some characteristics that need to be considered.

5.3.1 Definition

The loss distribution approach is described in Klugman *et al.* (2012) and Frachot *et al.* (2001). We assume that the operational loss L of the bank is divided into a matrix of homogenous losses:

$$L = \sum_{k=1}^{K} S_k \tag{5.1}$$

where S_k is the sum of losses of the k^{th} cell and K is the number of cells in the matrix. For instance, if we consider the Basel II classification, the mapping matrix contains 56 cells corresponding to the 8 business lines and 7 event types. The loss distribution approach is a

[5] However, the computation of the ILM coefficient is subject to some standard requirements. For instance, ILM is set to one for banks with a BIC lower than \$1 bn and supervisors can impose the value of the ILM coefficient for banks that do not meet loss data collection criteria.

method to model the random loss S_k of a particular cell. It assumes that S_k is the random sum of homogeneous individual losses:

$$S_k = \sum_{n=1}^{N_k(t)} X_n^{(k)} \tag{5.2}$$

where $N_k(t)$ is the random number of individual losses for the period $[0,t]$ and $X_n^{(k)}$ is the n^{th} individual loss. For example, if we consider internal fraud in corporate finance, we can write the loss for the next year as follows:

$$S = X_1 + X_2 + \ldots + X_{N(1)}$$

where X_1 is the first observed loss, X_2 is the second observed loss, $X_{N(1)}$ is the last observed loss of the year and $N(1)$ is the number of losses for the next year. We notice that we face two sources of uncertainty:

1. we don't know what will be the magnitude of each loss event(severity risk);
2. we don't know how many losses will occur in the next year (frequency risk).

In order to simplify the notations, we omit the index k and rewrite the random sum as follows:

$$S = \sum_{n=1}^{N(t)} X_n \tag{5.3}$$

The loss distribution approach is based on the following assumptions:

- the number $N(t)$ of losses follows the loss frequency distribution $\mathbf{P}$; the probability that the number of loss events is equal to n is denoted by $p(n)$;
- the sequence of individual losses X_n is independent and identically distributed (*iid*); the corresponding probability distribution $\mathbf{F}$ is called the loss severity distribution;
- the number of events is independent from the amount of loss events.

Once the probability distributions $\mathbf{P}$ and $\mathbf{F}$ are chosen, we can determine the probability distribution of the aggregate loss S, which is denoted by $\mathbf{G}$ and is called the compound distribution.

Example 52 *We assume that the number of losses is distributed as follows:*

n	0	1	2	3
$p(n)$	50%	30%	17%	3%

The loss amount can take the values \$100 *and* \$200 *with probabilities* 70% *and* 30%*.*

To calculate the probability distribution $\mathbf{G}$ of the compound loss, we first define the probability distribution of X_1, $X_1 + X_2$ and $X_1 + X_2 + X_3$, because the maximum number of losses is equal to 3. If there is only one loss, we have $\Pr\{X_1 = 100\} = 70\%$ and $\Pr\{X_1 = 200\} = 30\%$. In the case of two losses, we obtain $\Pr\{X_1 + X_2 = 200\} = 49\%$, $\Pr\{X_1 + X_2 = 300\} = 42\%$ and $\Pr\{X_1 + X_2 = 400\} = 9\%$. Finally, the sum of three losses takes the values 300, 400, 500 and 600 with probabilities 34.3%, 44.1%, 18.9% and 2.7%

respectively. We notice that these probabilities are in fact conditional to the number of losses. Using Bayes theorem, we obtain:

$$\Pr\{S = s\} = \sum_n \Pr\left\{\sum_{i=1}^n X_i = s \middle| N(t) = n\right\} \cdot \Pr\{N(t) = n\}$$

We deduce that:

$$\begin{aligned}\Pr\{S = 0\} &= \Pr\{N(t) = 0\} \\ &= 50\%\end{aligned}$$

and:

$$\begin{aligned}\Pr\{S = 100\} &= \Pr\{X_1 = 100\} \times \Pr\{N(t) = 1\} \\ &= 70\% \times 30\% \\ &= 21\%\end{aligned}$$

The compound loss can take the value 200 in two different ways:

$$\begin{aligned}\Pr\{S = 200\} &= \Pr\{X_1 = 200\} \times \Pr\{N(t) = 1\} + \\ &\quad \Pr\{X_1 + X_2 = 200\} \times \Pr\{N(t) = 2\} \\ &= 30\% \times 30\% + 49\% \times 17\% \\ &= 17.33\%\end{aligned}$$

For the other values of S, we obtain $\Pr\{S = 300\} = 8.169\%$, $\Pr\{S = 400\} = 2.853\%$, $\Pr\{S = 500\} = 0.567\%$ and $\Pr\{S = 600\} = 0.081\%$.

The previous example shows that the cumulative distribution function of S can be written as[6]:

$$\mathbf{G}(s) = \begin{cases} \sum_{n=1}^{\infty} p(n)\, \mathbf{F}^{n\star}(s) & \text{for} \quad s > 0 \\ p(0) & \text{for} \quad s = 0 \end{cases} \tag{5.4}$$

where $\mathbf{F}^{n\star}$ is the n-fold convolution of $\mathbf{F}$ with itself:

$$\mathbf{F}^{n\star}(s) = \Pr\left\{\sum_{i=1}^n X_i \leq s\right\} \tag{5.5}$$

In Figure 5.1, we give an example of a continuous compound distribution when the annual number of losses follows the Poisson distribution $\mathcal{P}(50)$ and the individual losses follow the log-normal distribution $\mathcal{LN}(8, 5)$. The capital charge, which is also called the capital-at-risk (CaR), corresponds then to the percentile α:

$$\text{CaR}(\alpha) = \mathbf{G}^{-1}(\alpha) \tag{5.6}$$

The regulatory capital is obtained by setting α to 99.9%: $\mathcal{K} = \text{CaR}(99.9\%)$. This capital-at-risk is valid for one cell of the operational risk matrix. Another issue is to calculate the capital-at-risk for the bank as a whole. This requires defining the dependence function between the different compound losses $(S_1, S_2, \ldots, S_K)$. In summary, here are the different steps to implement the loss distribution approach:

- for each cell of the operational risk matrix, we estimate the loss frequency distribution and the loss severity distribution;
- we then calculate the capital-at-risk;
- we define the dependence function between the different cells of the operational risk matrix, and deduce the aggregate capital-at-risk.

[6] When $\mathbf{F}$ is a discrete probability function, it is easy to calculate $\mathbf{F}^{n\star}(s)$ and then deduce $\mathbf{G}(s)$. However, the determination of $\mathbf{G}(s)$ is more difficult in the general case of continuous probability functions. This issue is discussed in Section 5.3.3 on page 327.

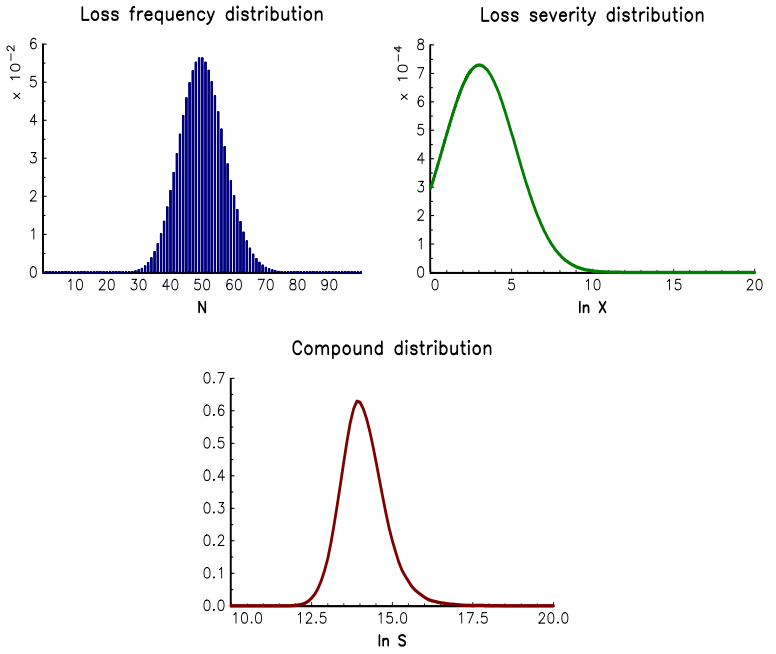


FIGURE 5.1: Compound distribution when $N \sim \mathcal{P}(50)$ and $X \sim \mathcal{LN}(8, 5)$

5.3.2 Parametric estimation

We first consider the estimation of the severity distribution, because we will see that the estimation of the frequency distribution can only be done after this first step.

5.3.2.1 Estimation of the loss severity distribution

We assume that the bank has an internal loss database. We note $\{x_1, \ldots, x_T\}$ the sample collected for a given cell of the operational risk matrix. We consider that the individual losses follow a given parametric distribution $\mathbf{F}$:

$$X \sim \mathbf{F}(x; \theta)$$

where θ is the vector of parameters to estimate.

In order to be a good candidate for modeling the loss severity, the probability distribution $\mathbf{F}$ must satisfy the following properties: the support of $\mathbf{F}$ is the interval $\mathbb{R}_+$, it is sufficiently flexible to accommodate a wide variety of empirical loss data and it can fit large losses. We list here the cumulative distribution functions that are the most used in operational risk models:

- Gamma $X \sim \mathcal{G}(\alpha, \beta)$

$$\mathbf{F}(x; \theta) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

  where $\alpha > 0$ and $\beta > 0$.

- Log-gamma $X \sim \mathcal{LG}(\alpha, \beta)$

$$\mathbf{F}(x; \theta) = \frac{\gamma(\alpha, \beta \ln x)}{\Gamma(\alpha)}$$

  where $\alpha > 0$ and $\beta > 0$.

- Log-logistic $X \sim \mathcal{LL}(\alpha, \beta)$

$$\begin{aligned} \mathbf{F}(x;\theta) &= \frac{1}{1+(x/\alpha)^{-\beta}} \\ &= \frac{x^\beta}{\alpha^\beta + x^\beta} \end{aligned}$$

  where $\alpha > 0$ and $\beta > 0$.

- Log-normal $X \sim \mathcal{LN}(\mu, \sigma^2)$

$$\mathbf{F}(x;\theta) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

  where $x > 0$ and $\sigma > 0$.

- Generalized extreme value $X \sim \mathcal{GEV}(\mu, \sigma, \xi)$

$$\mathbf{F}(x;\theta) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

  where $x > \mu - \sigma/\xi$, $\sigma > 0$ and $\xi > 0$.

- Pareto $X \sim \mathcal{P}(\alpha, x_-)$

$$\mathbf{F}(x;\theta) = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

  where $x \geq x_-$, $\alpha > 0$ and $x_- > 0$.

The vector of parameters θ can be estimated by the method of maximum likelihood (ML) or the generalized method of moments (GMM). In Chapter 10, we show that the log-likelihood function associated to the sample is:

$$\boldsymbol{\ell}(\theta) = \sum_{i=1}^{T} \ln f(x_i;\theta) \tag{5.7}$$

where $f(x;\theta)$ is the density function. In the case of the GMM, the empirical moments are:

$$\begin{cases} h_{i,1}(\theta) = x_i - \mathbb{E}[X] \\ h_{i,2}(\theta) = (x_i - \mathbb{E}[X])^2 - \text{var}(X) \end{cases} \tag{5.8}$$

In Table 5.4, we report the density function $f(x;\theta)$, the mean $\mathbb{E}[X]$ and the variance $\text{var}(X)$ when X follows one of the probability distributions described previously. For instance, if we consider that $X \sim \mathcal{LN}(\mu, \sigma^2)$, it follows that the log-likelihood function is:

$$\boldsymbol{\ell}(\theta) = -\sum_{i=1}^{T} \ln x_i - \frac{T}{2}\ln\sigma^2 - \frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{T}\left(\frac{\ln x_i - \mu}{\sigma}\right)^2$$

whereas the empirical moments are:

$$\begin{cases} h_{i,1}(\theta) = x_i - e^{\mu+\frac{1}{2}\sigma^2} \\ h_{i,2}(\theta) = \left(x_i - e^{\mu+\frac{1}{2}\sigma^2}\right)^2 - e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) \end{cases}$$

TABLE 5.4: Density function, mean and variance of parametric probability distribution

Distribution	$f(x;\theta)$	$\mathbb{E}[X]$	$\text{var}(X)$
$\mathcal{G}(\alpha,\beta)$	$\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$\mathcal{LG}(\alpha,\beta)$	$\frac{\beta^\alpha(\ln x)^{\alpha-1}}{x^{\beta+1}\Gamma(\alpha)}$	$\left(\frac{\beta}{\beta-1}\right)^\alpha$ if $\beta>1$	$\left(\frac{\beta}{\beta-2}\right)^\alpha-\left(\frac{\beta}{\beta-1}\right)^{2\alpha}$ if $\beta>2$
$\mathcal{LL}(\alpha,\beta)$	$\frac{\beta(x/\alpha)^{\beta-1}}{\alpha\left(1+(x/\alpha)^\beta\right)^2}$	$\frac{\alpha\pi}{\beta\sin(\pi/\beta)}$ if $\beta>1$	$\alpha^2\left(\frac{2\pi}{\beta\sin(2\pi/\beta)}-\frac{\pi^2}{\beta^2\sin^2(\pi/\beta)}\right)$ if $\beta>2$
$\mathcal{LN}\left(\mu,\sigma^2\right)$	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	$\exp\left(\mu+\frac{1}{2}\sigma^2\right)$	$\exp\left(2\mu+\sigma^2\right)\left(\exp\left(\sigma^2\right)-1\right)$
$\mathcal{GEV}(\mu,\sigma,\xi)$	$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1+1/\xi)}$ $\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$	$\mu+\frac{\sigma}{\xi}(\Gamma(1-\xi)-1)$ if $\xi<1$	$\frac{\sigma^2}{\xi^2}\left(\Gamma(1-2\xi)-\Gamma^2(1-\xi)\right)$ if $\xi<\frac{1}{2}$
$\mathcal{P}(\alpha,x_-)$	$\frac{\alpha x_-^\alpha}{x^{\alpha+1}}$	$\frac{\alpha x_-}{\alpha-1}$ if $\alpha>1$	$\frac{\alpha x_-^2}{(\alpha-1)^2(\alpha-2)}$ if $\alpha>2$

In the case of the log-normal distribution, the vector θ is composed of two parameters μ and σ, implying that two moments are sufficient to define the GMM estimator. This is also the case of other probability distributions, except the GEV distribution that requires specification of three empirical moments[7].

Example 53 *We assume that the individual losses take the following values expressed in thousand dollars:* 10.1, 12.5, 14, 25, 317.3, 353, 1 200, 1 254, 52 000 *and* 251 000.

Using the method of maximum likelihood, we find that $\hat{\alpha}_{\text{ML}}$ and $\hat{\beta}_{\text{ML}}$ are equal to 15.70 and 1.22 for the log-gamma distribution and 293 721 and 0.51 for the log-logistic distribution. In the case of the log-normal distribution[8], we obtain $\hat{\mu}_{\text{ML}}=12.89$ and $\hat{\sigma}_{\text{ML}}=3.35$.

The previous analysis assumes that the sample of operational losses for estimating θ represents a comprehensive and homogenous information of the underlying probability distribution $\mathbf{F}$. In practice, loss data are plagued by various sources of bias. The first issue lies in the data generating process which underlies the way data have been collected. In almost all cases, loss data have gone through a truncation process by which data are recorded only when their amounts are higher than some thresholds. In practice, banks' internal thresholds are set in order to balance two conflicting wishes: collecting as many data as possible while reducing costs by collecting only significant losses. These thresholds, which are defined by the global risk management policy of the bank, must satisfy some criteria:

[7] We can use the moment of order 3, which corresponds to:

$$\mathbb{E}\left[(X-\mathbb{E}[X])^3\right]=\frac{\sigma^3}{\xi^3}\left(\Gamma(1-3\xi)-3\Gamma(1-2\xi)\Gamma(1-\xi)+2\Gamma^3(1-\xi)\right)$$

[8] If we consider the generalized method of moments, the estimates are $\hat{\mu}_{\text{GMM}}=16.26$ and $\hat{\sigma}_{\text{GMM}}=1.40$.

> "*A bank must have an appropriate de minimis gross loss threshold for internal loss data collection, for example* €10 000. *The appropriate threshold may vary somewhat between banks, and within a bank across business lines and/or event types. However, particular thresholds should be broadly consistent with those used by peer banks*" (BCBS, 2006, page 153).

The second issue concerns the use of relevant external data, especially when there is reason to believe that the bank is exposed to infrequent, yet potentially severe losses. Typical examples are rogue trading or cyber attacks. If the bank has not yet experienced a large amount of loss due to these events in the past, this does not mean that it will never experience such problems in the future. Therefore, internal loss data must be supplemented by external data from public and/or pooled industry databases. Unfortunately, incorporating external data is rather dangerous and requires careful methodology to avoid the pitfalls regarding data heterogeneity, scaling problems and lack of comparability between too heterogeneous data. Unfortunately, there is no satisfactory solution to deal with these scaling issues, implying that banks use external data by taking into account only reporting biases and a fixed and known threshold[9].

The previous issues imply that operational risk loss data cannot be reduced to the sample of individual losses, but also requires specifying the threshold H_i for each individual loss x_i. The form of operational loss data is then $\{(x_i, H_i), i = 1, \ldots, T\}$, where x_i is the observed value of X knowing that X is larger than the threshold H_i. Reporting thresholds affect severity estimation in the sense that the sample severity distribution (i.e. the severity distribution of reported losses) is different from the '*true*' one (i.e. the severity distribution one would obtain if all the losses were reported). Unfortunately, the true distribution is the most relevant for calculating capital charge. As a consequence, linking the sample distribution to the true one is a necessary task. From a mathematical point of view, the true distribution is the probability distribution of X whereas the sample distribution is the probability distribution of $X \mid X \geq H_i$. We deduce that the sample distribution for a given threshold H is the conditional probability distribution defined as follows:

$$\begin{aligned}
\mathbf{F}^{\star}(x;\theta \mid H) &= \Pr\{X \leq x \mid X \geq H\} \\
&= \frac{\Pr\{X \leq x, X \geq H\}}{\Pr\{X \geq H\}} \\
&= \frac{\Pr\{X \leq x\} - \Pr\{X \leq \min(x, H)\}}{\Pr\{X \geq H\}} \\
&= \mathbb{1}\{x \geq H\} \cdot \frac{\mathbf{F}(x;\theta) - \mathbf{F}(H;\theta)}{1 - \mathbf{F}(H;\theta)}
\end{aligned} \tag{5.9}$$

It follows that the density function is:

$$f^{\star}(x;\theta \mid H) = \mathbb{1}\{x \geq H\} \cdot \frac{f(x;\theta)}{1 - \mathbf{F}(H;\theta)}$$

To estimate the vector of parameters θ, we continue to use the method of maximum likelihood or the generalized method of moments by considering the correction due to the

[9] See Baud *et al.* (2002, 2003) for more advanced techniques based on unknown and stochastic thresholds.

truncation of data. For the ML estimator, we have then:

$$\begin{aligned}\boldsymbol{\ell}(\theta) &= \sum_{i=1}^{T} \ln f^{\star}(x_i;\theta \mid H_i) \\ &= \sum_{i=1}^{T} \ln f(x_i;\theta) + \sum_{i=1}^{T} \ln \mathbb{1}\{x_i \geq H_i\} - \sum_{i=1}^{T} \ln\left(1 - \mathbf{F}(H_i;\theta)\right)\end{aligned} \tag{5.10}$$

where H_i is the threshold associated to the i^{th} observation. The correction term $-\sum_{i=1}^{T} \ln\left(1 - \mathbf{F}(H_i;\theta)\right)$ shows that maximizing a conventional log-likelihood function which ignores data truncation is totally misleading. We also notice that this term vanishes when H_i is equal to zero[10]. For the GMM estimator, the empirical moments become:

$$\begin{cases} h_{i,1}(\theta) = x_i - \mathbb{E}[X \mid X \geq H_i] \\ h_{i,2}(\theta) = (x_i - \mathbb{E}[X \mid X \geq H_i])^2 - \text{var}(X \mid X \geq H_i) \end{cases} \tag{5.11}$$

There is no reason that the conditional moment $\mathbb{E}[X^m \mid X \geq H_i]$ is equal to the unconditional moment $\mathbb{E}[X^m]$. Therefore, the conventional GMM estimator is biased and this is why we have to apply again the threshold correction.

If we consider again the log-normal distribution, the expression of the log-likelihood function (5.10) is[11]:

$$\begin{aligned}\boldsymbol{\ell}(\theta) &= -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \sum_{i=1}^{T}\ln x_i - \frac{1}{2}\sum_{i=1}^{T}\left(\frac{\ln x_i - \mu}{\sigma}\right)^2 - \\ &\quad \sum_{i=1}^{T}\ln\left(1 - \Phi\left(\frac{\ln H_i - \mu}{\sigma}\right)\right)\end{aligned}$$

Let us now calculate the conditional moment $\mu'_m(X) = \mathbb{E}[X^m \mid X \geq H]$. By using the notation $\Phi_c(x) = 1 - \Phi((x-\mu)/\sigma)$, we have:

$$\begin{aligned}\mu'_m(X) &= \frac{1}{\Phi_c(\ln H)}\int_H^{\infty} \frac{x^m}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right) \mathrm{d}x \\ &= \frac{1}{\Phi_c(\ln H)}\int_{\ln H}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 + my\right) \mathrm{d}y \\ &= \frac{\exp\left(m\mu + m^2\sigma^2/2\right)}{\Phi_c(\ln H)}\int_{\ln H}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y - \left(\mu + m\sigma^2\right)}{\sigma}\right)^2\right) \mathrm{d}y \\ &= \frac{\Phi_c\left(\ln H - m\sigma^2\right)}{\Phi_c(\ln H)} \exp\left(m\mu + m^2\sigma^2/2\right)\end{aligned}$$

We deduce that:

$$\mathbb{E}[X \mid X \geq H] = a(\theta, H) = \frac{1 - \Phi\left(\frac{\ln H - \mu - \sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{\mu + \frac{1}{2}\sigma^2}$$

[10] Indeed, we have $\mathbf{F}(0;\theta) = 0$ and $\ln\left(1 - \mathbf{F}(0;\theta)\right) = 0$.

[11] By construction, the observed value x_i is larger than the threshold H_i, meaning that $\ln \mathbb{1}\{x_i \geq H_i\}$ is equal to 0.

and:

$$\mathbb{E}\left[X^2 \mid X \geq H\right] = b(\theta, H) = \frac{1 - \Phi\left(\frac{\ln H - \mu - 2\sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{2\mu + 2\sigma^2}$$

We finally obtain:

$$\begin{cases} h_{i,1}(\theta) = x_i - a(\theta, H_i) \\ h_{i,2}(\theta) = x_i^2 - 2x_i a(\theta, H_i) + 2a^2(\theta, H_i) - b(\theta, H_i) \end{cases}$$

In order to illustrate the impact of the truncation, we report in Figure 5.2 the cumulative distribution function and the probability density function of $X \mid X > H$ when X follows the log-normal distribution $\mathcal{LN}(8, 5)$. The threshold H is set at \$10 000, meaning that the bank collects operational losses when the amount is larger than this threshold. In the bottom panels of the figure, we indicate the mean and the variance with respect to the threshold H. We notice that data truncation increases the magnitude of the mean and the variance. For instance, when H is set at \$10 000, the conditional mean and variance are multiplied by a factor equal to 3.25 with respect to the unconditional mean and variance.

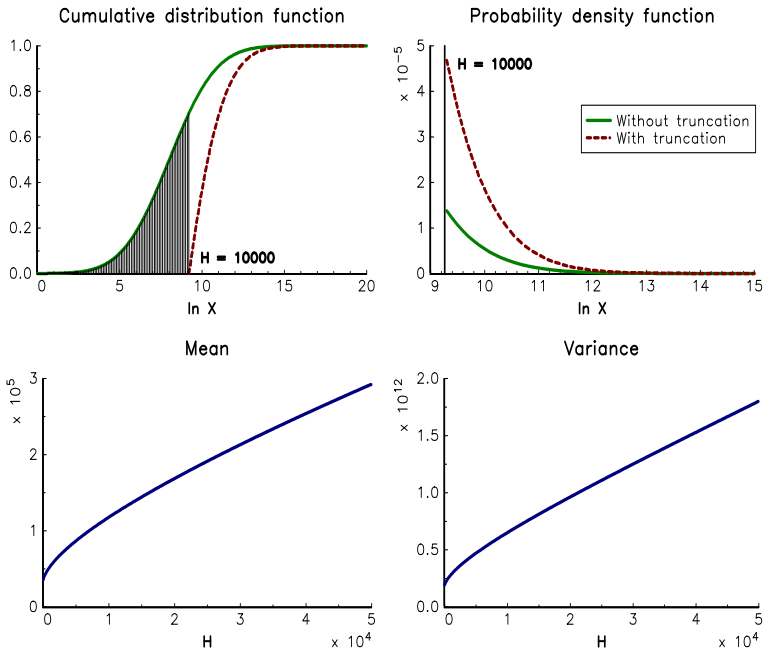


FIGURE 5.2: Impact of the threshold H on the severity distribution

Example 54 *We consider Example 53 and assume that the losses have been collected using a unique threshold that is equal to* \$5 000.

By using the truncation correction, the ML estimates become $\hat{\mu}_{\text{ML}} = 8.00$ and $\hat{\sigma}_{\text{ML}} = 5.71$ for the log-normal model. In Figure 5.3, we compare the log-normal cumulative distribution function without and with the truncation correction. We notice that the results are very different.

The previous example shows that estimating the parameters of the probability distribution is not sufficient to define the severity distribution. Indeed, ML and GMM give

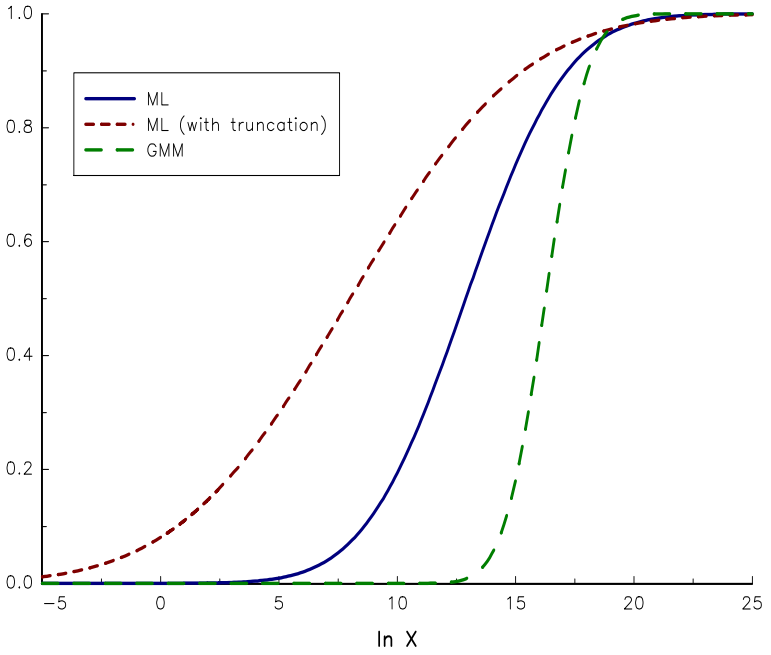


FIGURE 5.3: Comparison of the estimated severity distributions

two different log-normal probability distributions. The issue is to decide which is the best parametrization. In a similar way, the choice between the several probability families (log-normal, log-gamma, GEV, Pareto, etc.) is an open question. This is why fitting the severity distribution does not reduce to estimate the parameters of a given probability distribution. It must be completed by a second step that consists in selecting the best estimated probability distribution. However, traditional goodness-of-fit tests (Kolmogorov-Smirnov, Anderson-Darling, etc.) are not useful, because they concern the entire probability distribution. In operational risk, extreme events are more relevant. This explains why QQ plots and order statistics are generally used to assess the fitting of the upper tail. A QQ plot represents the quantiles of the empirical distribution against those of the theoretical model. If the statistical model describes perfectly the data, we obtain the diagonal line $y = x$. In Figure 5.4, we show an example of QQ plot. We notice that the theoretical quantiles obtained from the statistical model are in line with those calculated with the empirical data when the quantile is lower than 80%. Otherwise, the theoretical quantiles are above the empirical quantiles, meaning that extreme events are underestimated by the statistical model. We deduce that the body of the distribution is well estimated, but not the upper tail of the distribution. However, medium losses are less important than high losses in operational risk.

5.3.2.2 Estimation of the loss frequency distribution

In order to model the frequency distribution, we have to specify the counting process $N(t)$, which defines the number of losses occurring during the time period $[0, t]$. The number of losses for the time period $[t_1, t_2]$ is then equal to:

$$N(t_1; t_2) = N(t_2) - N(t_1)$$

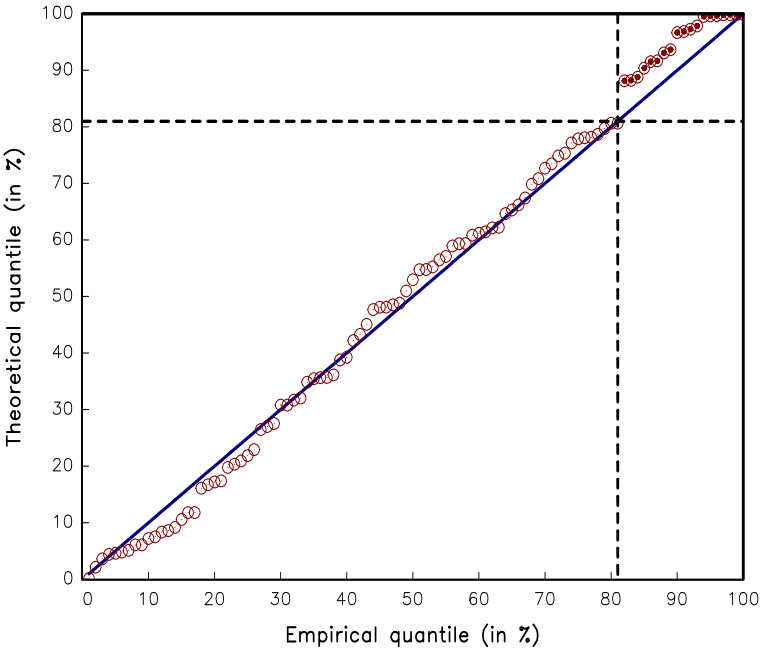


FIGURE 5.4: An example of QQ plot where extreme events are underestimated

We generally made the following statements about the stochastic process $N(t)$:

- the distribution of the number of losses $N(t; t+h)$ for each $h > 0$ is independent of t; moreover, $N(t; t+h)$ is stationary and depends only on the time interval h;
- the random variables $N(t_1; t_2)$ and $N(t_3; t_4)$ are independent if the time intervals $[t_1, t_2]$ and $[t_3, t_4]$ are disjoint;
- no more than one loss may occur at time t.

These simple assumptions define a Poisson process, which satisfies the following properties:

1. there exists a scalar $\lambda > 0$ such that the distribution of $N(t)$ has a Poisson distribution with parameter λt;
2. the duration between two successive losses is *iid* and follows the exponential distribution $\mathcal{E}(\lambda)$.

Let $p(n)$ be the probability to have n losses. We deduce that:

$$\begin{aligned} p(n) &= \Pr\{N(t) = n\} \\ &= \frac{e^{-\lambda t} \cdot (\lambda t)^n}{n!} \end{aligned} \tag{5.12}$$

Without loss of generality, we can fix $t = 1$ because it corresponds to the required one-year time period for calculating the capital charge. In this case, $N(1)$ is simply a Poisson distribution with parameter λ. This probability distribution has a useful property for time aggregation. Indeed, the sum of two independent Poisson variables N_1 and N_2 with parameters λ_1 and λ_2 is also a Poisson variable with parameter $\lambda_1 + \lambda_2$. This property is a direct result of the definition of the Poisson process. In particular, we have:

$$\sum_{k=1}^{K} N\left(\frac{k-1}{K}; \frac{k}{K}\right) = N(1)$$

where $N((k-1)/K; k/K) \sim \mathcal{P}(\lambda/K)$. This means that we can estimate the frequency distribution at a quarterly or monthly period and convert it to an annual period by simply multiplying the quarterly or monthly intensity parameter by 4 or 12.

The estimation of the annual intensity λ can be done using the method of maximum likelihood. In this case, $\hat{\lambda}$ is the mean of the annual number of losses:

$$\hat{\lambda} = \frac{1}{n_y} \sum_{y=1}^{n_y} N_y \tag{5.13}$$

where N_y is the number of losses occurring at year y and n_y is the number of observations. One of the key features of the Poisson distribution is that the variance equals the mean:

$$\lambda = \mathbb{E}[N(1)] = \text{var}(N(1)) \tag{5.14}$$

We can use this property to estimate λ by the method of moments. If we consider the first moment, we obtain the ML estimator, whereas we have with the second moment:

$$\hat{\lambda} = \frac{1}{n_y} \sum_{y=1}^{n_y} \left(N_y - \bar{N}\right)^2$$

where $\bar{N}$ is the average number of losses.

Example 55 *We assume that the annual number of losses from 2006 to 2015 is the following:* 57, 62, 45, 24, 82, 36, 98, 75, 76 *and* 45.

The mean is equal to 60 whereas the variance is equal to 474.40. In Figure 5.5, we show the probability mass function of the Poisson distribution with parameter 60. We notice that the parameter λ is not enough large to reproduce the variance and the range of the sample. However, using the moment estimator based on the variance is completely unrealistic.

When the variance exceeds the mean, we use the negative binomial distribution $\mathcal{NB}(r,p)$, which is defined as follows:

$$\begin{aligned} p(n) &= \binom{r+n-1}{n} (1-p)^r p^n \\ &= \frac{\Gamma(r+n)}{n!\,\Gamma(r)} (1-p)^r p^n \end{aligned}$$

where $r > 0$ and $p \in [0,1]$. The negative binomial distribution can be viewed as the probability distribution of the number of successes in a sequence of *iid* Bernoulli random variables $\mathcal{B}(p)$ until we get r failures. The negative binomial distribution is then a generalization of the geometric distribution. Concerning the first two moments, we have:

$$\mathbb{E}[\mathcal{NB}(r,p)] = \frac{p \cdot r}{1-p}$$

and:

$$\text{var}(\mathcal{NB}(r,p)) = \frac{p \cdot r}{(1-p)^2}$$

We verify that:

$$\begin{aligned} \text{var}(\mathcal{NB}(r,p)) &= \frac{1}{1-p} \cdot \mathbb{E}[\mathcal{NB}(r,p)] \\ &> \mathbb{E}[\mathcal{NB}(r,p)] \end{aligned}$$

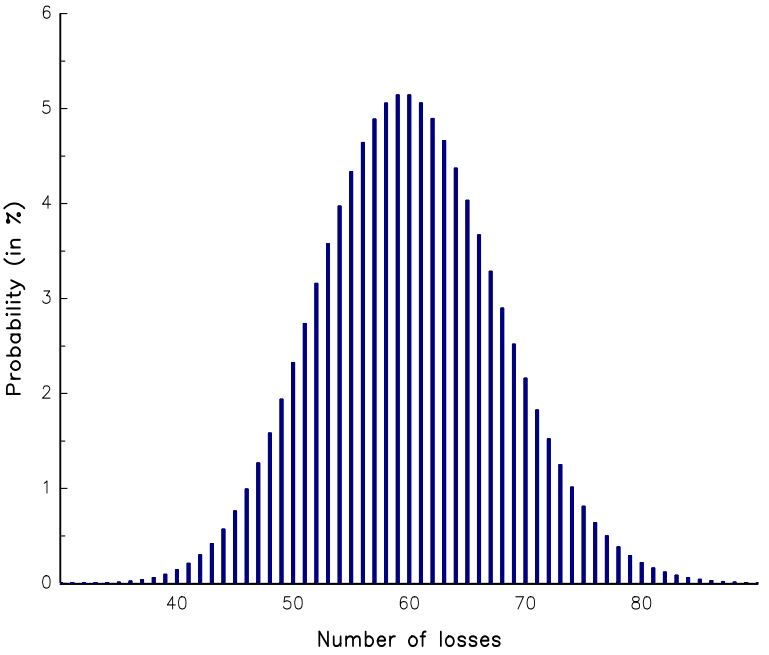


FIGURE 5.5: PMF of the Poisson distribution $\mathcal{P}(60)$

Remark 62 *The negative binomial distribution corresponds to a Poisson process where the intensity parameter is random and follows a gamma distribution*[12]*:*

$$\mathcal{NB}(r,p) \sim \mathcal{P}(\Lambda) \quad \textit{and} \quad \Lambda \sim \mathcal{G}(\alpha,\beta)$$

where $\alpha = r$ *and* $\beta = (1-p)/p$.

We consider again Example 55 and assume that the number of losses is described by the negative binomial distribution. Using the method of moments, we obtain the following estimates:

$$\hat{r} = \frac{m^2}{v-m} = \frac{60^2}{474.40-60} = 8.6873$$

and

$$\hat{p} = \frac{v-m}{v} = \frac{474.40-60}{474.40} = 0.8735$$

where m is the mean and v is the variance of the sample. Using these estimates as the starting values of the numerical optimization procedure, the ML estimates are $\hat{r} = 7.7788$ and $\hat{p} = 0.8852$. We report the corresponding probability mass function $p(n)$ in Figure 5.6. We notice that this distribution better describes the sample that the Poisson distribution, because it has a larger support. In fact, we show in Figure 5.7 the probability density function of λ for the two estimated counting processes. For the Poisson distribution, λ is constant and equal to 60, whereas λ has a gamma distribution $\mathcal{G}(7.7788, 0.1296)$ in the case of the negative binomial distribution. The variance of the gamma distribution explains the larger variance of the negative binomial distribution with respect to the Poisson distribution, while we notice that the two distributions have the same mean.

[12] See Exercise 5.4.6 on page 346.

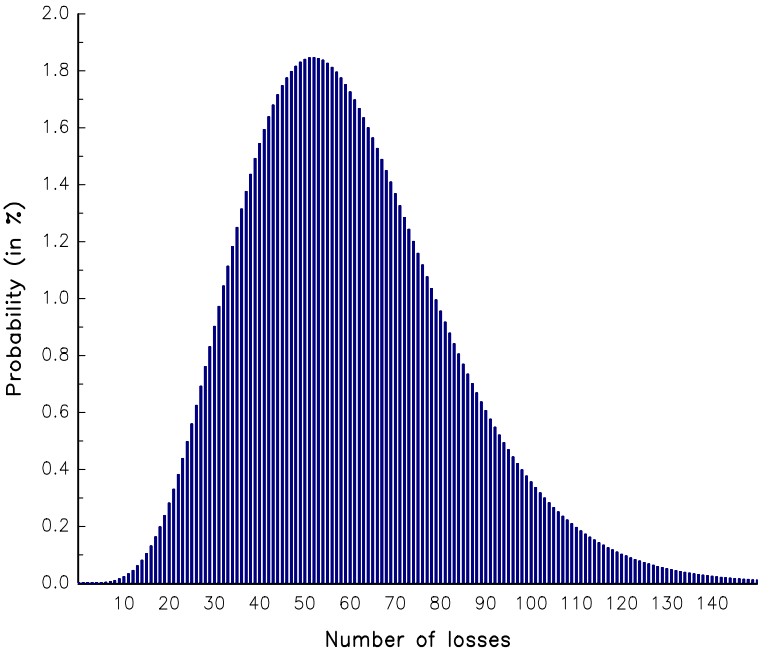


FIGURE 5.6: PMF of the negative binomial distribution

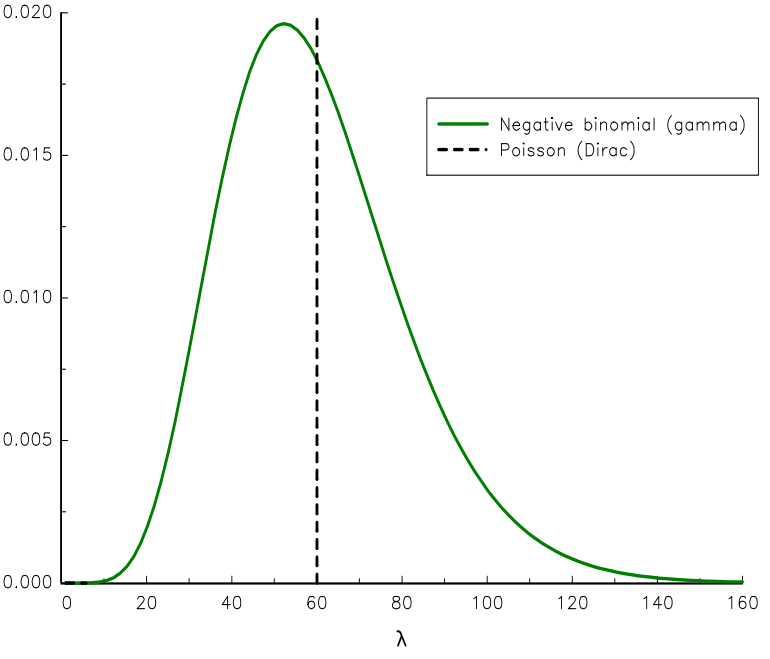


FIGURE 5.7: Probability density function of the parameter λ

As in the case of the severity distribution, data truncation and reporting bias have an impact of the frequency distribution (Frachot *et al.*, 2006). For instance, if one bank's reporting threshold H is set at a high level, then the average number of reported losses will be low. It does not imply that the bank is allowed to have a lower capital charge than another bank that uses a lower threshold and is otherwise identical to the first one. It simply means that the average number of losses must be corrected for reporting bias as well. It appears that the calibration of the frequency distribution comes as a second step (after having calibrated the severity distribution) because the aforementioned correction needs an estimate of the exceedance probability $\Pr\{X > H\}$ for its calculation. This is rather straightforward: the difference (more precisely the ratio) between the number of reported events and the '*true*' number of events (which would be obtained if all the losses were reported, *i.e.* with a zero-threshold) corresponds exactly to the probability of one loss being higher than the threshold. This probability is a direct by-product of the severity distribution.

Let $N_H(t)$ be the number of events that are larger than the threshold H. By definition, $N_H(t)$ is the counting process of exceedance events:

$$N_H(t) = \sum_{i=1}^{N(t)} \mathbb{1}\{X_i > H\}$$

It follows that:

$$\begin{aligned}
\mathbb{E}[N_H(t)] &= \mathbb{E}\left[\sum_{i=1}^{N(t)} \mathbb{1}\{X_i > H\}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}\{X_i > H\} \middle| N(t) = n\right] \\
&= \mathbb{E}[N(t)] \cdot \mathbb{E}[\mathbb{1}\{X_i > H\}]
\end{aligned}$$

because the random variables $X_1, \ldots, X_n$ are *iid* and independent from the random number of events $N(t)$. We deduce that:

$$\begin{aligned}
\mathbb{E}[N_H(t)] &= \mathbb{E}[N(t)] \cdot \Pr\{X_i > H\} \\
&= \mathbb{E}[N(t)] \cdot (1 - \mathbf{F}(H;\theta)) \qquad (5.15)
\end{aligned}$$

This latter equation provides information about the transformation of the counting process $N(t)$ into the exceedance process. However, it only concerns the mean and not the distribution itself. One interesting feature of data truncation is when the distribution of the threshold exceedance process belongs to the same distribution class of the counting process. It is the case of the Poisson distribution:

$$\mathbf{P}_H(\lambda) = \mathbf{P}(\lambda_H)$$

Using Equation (5.15), it follows that the Poisson parameter λ_H of the exceedance process is simply the product of the Poisson parameter λ by the exceedance probability $\Pr\{X > H\}$:

$$\lambda_H = \lambda \cdot (1 - \mathbf{F}(H;\theta))$$

We deduce that the estimator $\hat{\lambda}$ has the following expression:

$$\hat{\lambda} = \frac{\hat{\lambda}_H}{1 - \mathbf{F}\left(H;\hat{\theta}\right)}$$

where $\hat{\lambda}_H$ is the average number of losses that are collected above the threshold H and $\mathbf{F}\left(x;\hat{\theta}\right)$ is the parametric estimate of the severity distribution.

Example 56 *We consider that the bank has collected the loss data from 2006 to 2015 with a threshold of* \$20 000. *For a given event type, the calibrated severity distribution corresponds to a log-normal distribution with parameters* $\hat{\mu} = 7.3$ *and* $\hat{\sigma} = 2.1$, *whereas the annual number of losses is the following:* 23, 13, 50, 12, 25, 36, 48, 27, 18 *and* 35.

Using the Poisson distribution, we obtain $\hat{\lambda}_H = 28.70$. The probability that the loss exceeds the threshold H is equal to:

$$\Pr\{X > 20\,000\} = 1 - \Phi\left(\frac{\ln(20\,000) - 7.3}{2.1}\right) = 10.75\%$$

This means that only 10.75% of losses can be observed when we apply a threshold of \$20 000. We then deduce that the estimate of the Poisson parameter is equal to:

$$\hat{\lambda} = \frac{28.70}{10.75\%} = 266.90$$

On average, there are in fact about 270 loss events per year.

We could discuss whether the previous result remains valid in the case of the negative binomial distribution. If it is the case, then we have:

$$\mathbf{P}_H(r, p) = \mathbf{P}(r_H, p_H)$$

Using Equation (5.15), we deduce that:

$$\frac{p_H \cdot r_H}{1 - p_H} = \frac{p \cdot r}{1 - p} \cdot (1 - \mathbf{F}(H; \theta))$$

If we assume that r_H is equal to r, we obtain:

$$p_H = \frac{p \cdot (1 - \mathbf{F}(H; \theta))}{1 - p \cdot \mathbf{F}(H; \theta)}$$

We verify the following inequality $p \leq p_H \leq 1$. However, this solution is not completely satisfactory.

5.3.3 Calculating the capital charge

Once the frequency and severity distributions are calibrated, the computation of the capital charge is straightforward. For that, we can use the Monte Carlo method or different analytical methods. The Monte Carlo method is much more used, because it is more flexible and gives better results in the case of low frequency/high severity events. Analytical approaches, which are very popular in insurance, can be used for high frequency/low severity events. One remaining challenge, however, is aggregating the capital charge of the different cells of the mapping matrix. By construction, the loss distribution approach assumes that aggregate losses are independent. Nevertheless, regulation are forcing banks to take into account positive correlation between risk events. The solution is then to consider copula functions.

5.3.3.1 Monte Carlo approach

We reiterate that the one-year compound loss of a given cell is defined as follows:

$$S = \sum_{i=1}^{N(1)} X_i$$

where $X_i \sim \mathbf{F}$ and $N(1) \sim \mathbf{P}$. The capital-at-risk is then the 99% quantile of the compound loss distribution. To estimate the capital charge by Monte Carlo, we first simulate the annual number of losses from the frequency distribution and then simulate individual losses in order to calculate the compound loss. Finally, the quantile is estimated by order statistics. The algorithm is described below.

Algorithm 1 Compute the capital-at-risk for an operational risk cell

```
Initialize the number of simulations n_S
for j = 1 : n_S do
  Simulate an annual number n of losses from the frequency distribution P
  S_j ← 0
  for i = 1 : n do
    Simulate a loss X_i from the severity distribution F
    S_j = S_j + X_i
  end for
end for
Calculate the order statistics S_{1:n_S}, ..., S_{n_S:n_S}
Deduce the capital-at-risk CaR = S_{αn_S:n_S} with α = 99.9%
return CaR
```

Let us illustrate this algorithm when $N(1) \sim \mathcal{P}(4)$ and $X_i \sim \mathcal{LN}(8, 4)$. Using a linear congruential method, the simulated values of $N(1)$ are 3, 4, 1, 2, 3, etc. while the simulated values of X_i are 3388.6, 259.8, 13328.3, 39.7, 1220.8, 1486.4, 15197.1, 3205.3, 5070.4, 84704.1, 64.9, 1237.5, 187073.6, 4757.8, 50.3, 2805.7, etc. For the first simulation, we have three losses and we obtain:

$$S_1 = 3388.6 + 259.8 + 13328.3 = \$16\,976.7$$

For the second simulation, the number of losses is equal to four and the compound loss is equal to:

$$S_2 = 39.7 + 1220.8 + 1486.4 + 15197.1 = \$17\,944.0$$

For the third simulation, we obtain $S_3 = \$3\,205.3$, and so on. Using n_S simulations, the value of the capital charge is estimated with the 99.9% empirical quantile based on order statistics. For instance, Figure 5.8 shows the histogram of 2 000 simulated values of the capital-at-risk estimated with one million simulations. The true value is equal to \$3.24 mn. However, we notice that the variance of the estimator is large. Indeed, the range of the MC estimator is between \$3.10 mn and \$3.40 mn in our experiments with one million simulation runs.

The estimation of the capital-at-risk with a high accuracy is therefore difficult. The convergence of the Monte Carlo algorithm is low and the estimated quantile can be very far from the true quantile especially when the severity loss distribution is heavy tailed and the confidence level α is high. That's why it is important to control the accuracy of $\mathbf{G}^{-1}(\alpha)$. This can be done by verifying that the estimated moments are close to the theoretical ones. For the first two central moments, we have:

$$\mathbb{E}[S] = \mathbb{E}[N(1)] \cdot \mathbb{E}[X_i]$$

and:

$$\text{var}(S) = \mathbb{E}[N(1)] \cdot \text{var}(X_i) + \text{var}(N(1)) \cdot \mathbb{E}^2[X_i]$$

To illustrate the convergence problem, we consider the example of the compound Poisson distribution where $N(1) \sim P(10)$ and $X_i \sim \mathcal{LN}(5, \sigma^2)$. We compute the aggregate loss

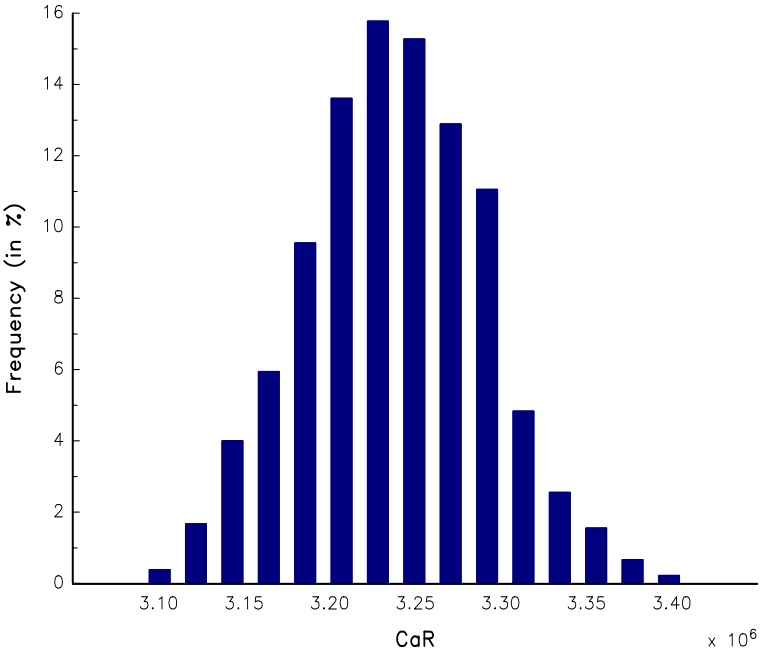


FIGURE 5.8: Histogram of the MC estimator $\widehat{\text{CaR}}$

distribution by the Monte Carlo method for different number n_S of simulations and different runs. To measure the accuracy, we calculate the ratio between the MC standard deviation $\hat{\sigma}_{n_S}(S)$ and the true value $\sigma(S)$:

$$R(n_s) = \frac{\hat{\sigma}_{n_S}(S)}{\sigma(S)}$$

We notice that the convergence is much more erratic when σ takes a high value (Figure 5.10) than when σ is low (Figure 5.9). When σ takes the value 1, the convergence of the Monte Carlo method is verified with 100 000 simulations. When σ takes the value 2.5, 100 million simulations are not sufficient to estimate the second moment, and then the capital-at-risk. Indeed, the occurrence probability of extreme events is generally underestimated. Sometimes, a severe loss is simulated implying a jump in the empirical standard deviation (see Figure 5.10). This is why we need a large number of simulations in order to be confident when estimating the 99.9% capital-at-risk with high severity distributions.

Remark 63 *With the Monte Carlo approach, we can easily integrate mitigation factors such as insurance coverage. An insurance contract is generally defined by a deductive*[13] A *and the maximum amount* B *of a loss, which is covered by the insurer. The effective loss* $\tilde{X}_i$ *suffered by the bank is then the difference between the loss of the event and the amount paid by the insurer:*

$$\tilde{X}_i = X_i - \max(\min(X_i, B) - A, 0)$$

The relationship between X_i *and* $\tilde{X}_i$ *is shown in Figure 5.11. In this case, the annual loss of the bank becomes:*

$$S = \sum_{i=1}^{N(1)} \tilde{X}_i$$

[13]It corresponds to the loss amount the bank has to cover by itself.

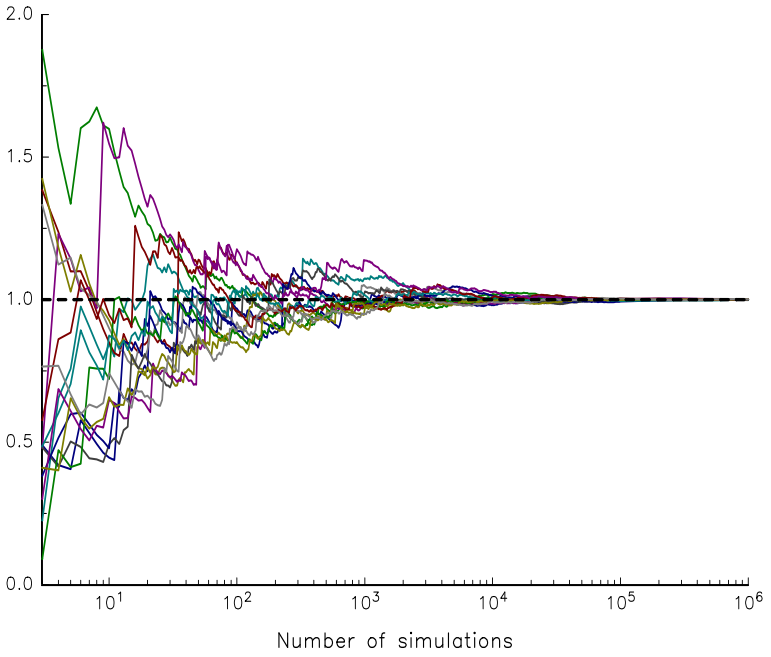


FIGURE 5.9: Convergence of the accuracy ratio $R(n_s)$ when $\sigma = 1$

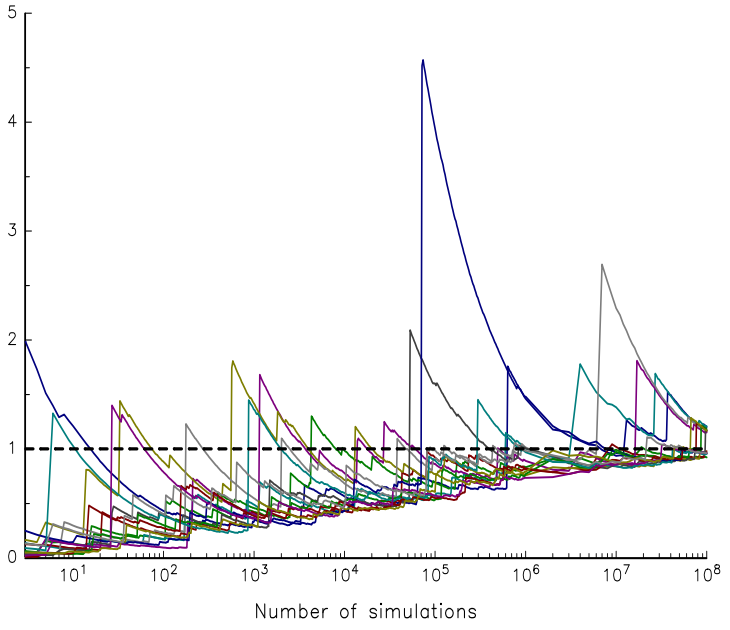


FIGURE 5.10: Convergence of the accuracy ratio $R(n_s)$ when $\sigma = 2.5$

Taking into account an insurance contract is therefore equivalent to replace X_i by $\tilde{X}_i$ in the Monte Carlo simulations.

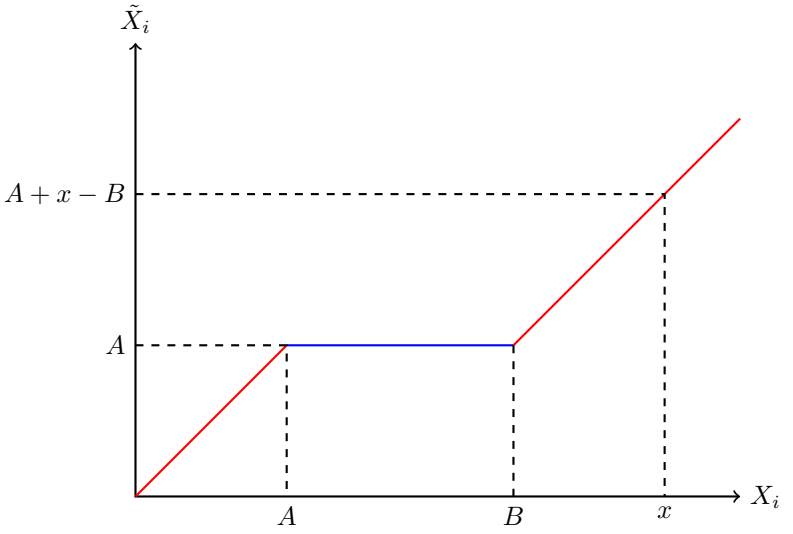


FIGURE 5.11: Impact of the insurance contract on the operational risk loss

5.3.3.2 Analytical approaches

There are three analytical (or semi-analytical) methods to compute the aggregate loss distribution: the solution based on characteristic functions, Panjer recursion and the single loss approximation.

Method of characteristic functions Formally, the characteristic function of the random variable X is defined by:

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$$

If X has a continuous probability distribution $\mathbf{F}$, we obtain:

$$\varphi_X(t) = \int_0^\infty e^{itx} \, \mathrm{d}\mathbf{F}(x)$$

We notice that the characteristic function of the sum of n independent random variables is the product of their characteristic functions:

$$\begin{aligned}
\varphi_{X_1+\ldots+X_n}(t) &= \mathbb{E}\left[e^{it(X_1+X_2+\cdots+X_n)}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[e^{itX_i}\right] \\
&= \prod_{i=1}^{n} \varphi_{X_i}(t)
\end{aligned}$$

It comes that the characteristic function of the compound distribution $\mathbf{G}$ is given by:

$$\varphi_S(t) = \sum_{n=0}^{\infty} p(n) \left(\varphi_X(t)\right)^n = \varphi_{N(1)}\left(\varphi_X(t)\right)$$

where $\varphi_{N(1)}(t)$ is the probability generating function of $N(1)$. For example, if $N(1) \sim \mathcal{P}(\lambda)$, we have:

$$\varphi_{N(1)}(t) = e^{\lambda(t-1)}$$

and:

$$\varphi_S(t) = e^{\lambda(\varphi_X(t)-1)}$$

We finally deduce that S has the probability density function given by the Laplace transform of $\varphi_S(t)$:

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_S(t)\,\mathrm{d}t$$

Using this expression, we can easily compute the cumulative distribution function and its inverse with the fast fourier transform.

Panjer recursive approach Panjer (1981) introduces recursive approaches to compute high-order convolutions. He showed that if the probability mass function of the counting process $N(t)$ satisfies:

$$p(n) = \left(a + \frac{b}{n}\right) p(n-1)$$

where a and b are two scalars, then the following recursion holds:

$$g(x) = p(1) f(x) + \int_0^x \left(a + b\frac{y}{x}\right) f(y)\, g(x-y)\,\mathrm{d}y$$

where $x > 0$. For discrete severity distributions satisfying $f_n = \Pr\{X_i = n\delta\}$ where δ is the monetary unit (e.g. \$10 000), the Panjer recursion becomes:

$$\begin{aligned} g_n &= \Pr\{S = n\delta\} \\ &= \frac{1}{1 - af_0}\sum_{j=1}^{n}\left(a + \frac{bj}{n}\right) f_j g_{n-j} \end{aligned}$$

where:

$$\begin{aligned} g_0 &= \sum_{n=0}^{\infty} p(n)\,(f_0)^n \\ &= \begin{cases} p(0)\, e^{bf_0} & \text{if } a = 0 \\ p(0)\,(1 - af_0)^{-1-b/a} & \text{otherwise} \end{cases} \end{aligned}$$

The capital-at-risk is then equal to:

$$\mathrm{CaR}(\alpha) = n^\star \delta$$

where:

$$n^\star = \inf\left\{ n : \sum_{j=0}^{n} g_j \geq \alpha \right\}$$

Like the method of characteristic functions, the Panjer recursion is very popular among academics, but produces significant numerical errors in practice when applied to operational risk losses. The issue is the support of the compound distribution, whose range can be from zero to several billions.

Example 57 *We consider the compound Poisson distribution with log-normal losses and different sets of parameters:*

(a) $\lambda = 5$, $\mu = 5$, $\sigma = 1.0$;

(b) $\lambda = 5$, $\mu = 5$, $\sigma = 1.5$;

(c) $\lambda = 5$, $\mu = 5$, $\sigma = 2.0$;

(d) $\lambda = 50$, $\mu = 5$, $\sigma = 2.0$.

In order to implement the Panjer recursion, we have to perform a discretization of the severity distribution. Using the central difference approximations, we have:

$$\begin{aligned} f_n &= \Pr\left\{n\delta - \frac{\delta}{2} \leq X_i \leq n\delta + \frac{\delta}{2}\right\} \\ &= \mathbf{F}\left(n\delta + \frac{\delta}{2}\right) - \mathbf{F}\left(n\delta - \frac{\delta}{2}\right) \end{aligned}$$

To initialize the algorithm, we use the convention $f_0 = \mathbf{F}(\delta/2)$. In Figure 5.12, we compare the cumulative distribution function of the aggregate loss obtained with the Panjer recursion and Monte Carlo simulations[14]. We deduce the capital-at-risk for different values of α in Table 5.5. In our case, the Panjer algorithm gives a good approximation, because the support of the distribution is '*bounded* '. When the aggregate loss can take very large values, we need a lot of iterations to achieve the convergence[15]. Moreover, we may have underflow in computations because $g_0 \approx 0$.

TABLE 5.5: Comparison of the capital-at-risk calculated with Panjer recursion and Monte Carlo simulations

α	Panjer recursion				Monte Carlo simulations			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
90%	2400	4500	11000	91000	2350	4908	11648	93677
95%	2900	6500	19000	120000	2896	6913	19063	123569
99%	4300	13500	52000	231000	4274	13711	51908	233567
99.5%	4900	18000	77000	308000	4958	17844	77754	310172
99.9%	6800	32500	182000	604000	6773	32574	185950	604756

Single loss approximation If the severity belongs to the family of subexponential distributions, then Böcker and Klüppelberg (2005) and Böcker and Sprittulla (2006) show that the percentile of the compound distribution can be approximated by the following expression:

$$\mathbf{G}^{-1}(\alpha) \approx (\mathbb{E}[N(1)] - 1) \cdot \mathbb{E}[X_i] + \mathbf{F}^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N(1)]}\right) \tag{5.16}$$

It follows that the capital-at-risk is the sum of the expected loss and the unexpected loss defined as follows:

$$\begin{aligned} \text{EL} &= \mathbb{E}[N(1)] \cdot \mathbb{E}[X_i] \\ \text{UL}(\alpha) &= \mathbf{F}^{-1}\left(1 - \frac{1-\alpha}{N(1)}\right) - \mathbb{E}[X_i] \end{aligned}$$

[14] We use one million simulations.

[15] In this case, it is not obvious that the Panjer recursion is faster than Monte Carlo simulations.

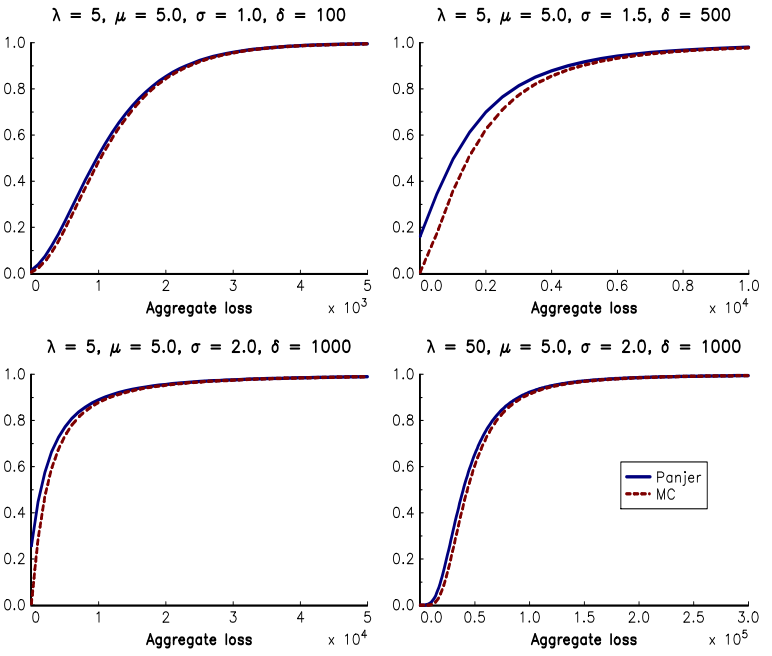


FIGURE 5.12: Comparison between the Panjer and MC compound distributions

To understand Formula (5.16), we recall that subexponential distributions are a special case of heavy-tailed distributions and satisfy the following property:

$$\lim_{x\to\infty} \frac{\Pr\{X_1 + \cdots + X_n > x\}}{\Pr\{\max(X_1, \ldots, X_n) > x\}} = 1$$

This means that large values of the aggregate loss are dominated by the maximum loss of one event. If we decompose the capital-at-risk as a sum of risk contributions, we obtain:

$$\mathbf{G}^{-1}(\alpha) = \sum_{i=1}^{\mathbb{E}[N(1)]} \mathcal{RC}_i$$

where:

$$\mathcal{RC}_i = \mathbb{E}[X_i] \qquad \text{for } i \neq i^\star$$

and:

$$\mathcal{RC}_{i^\star} = \mathbf{F}^{-1}\left(1 - \frac{1-\alpha}{N(1)}\right)$$

In this model, the capital-at-risk is mainly explained by the single largest loss $i^\star$. If we neglect the small losses, the capital-at-risk at the confidence level α_{CaR} is related to the quantile α_{Severity} of the loss severity:

$$\alpha_{\text{Severity}} = 1 - \frac{1-\alpha_{\text{CaR}}}{N(1)}$$

This relationship[16] is shown in Figure 5.13 and explains why this framework is called the single loss approximation (SLA). For instance, if the annual number of losses is equal to 100 on average, computing the capital-at-risk with a 99.9% confidence level is equivalent to estimate the quantile 99.999% of the loss severity.

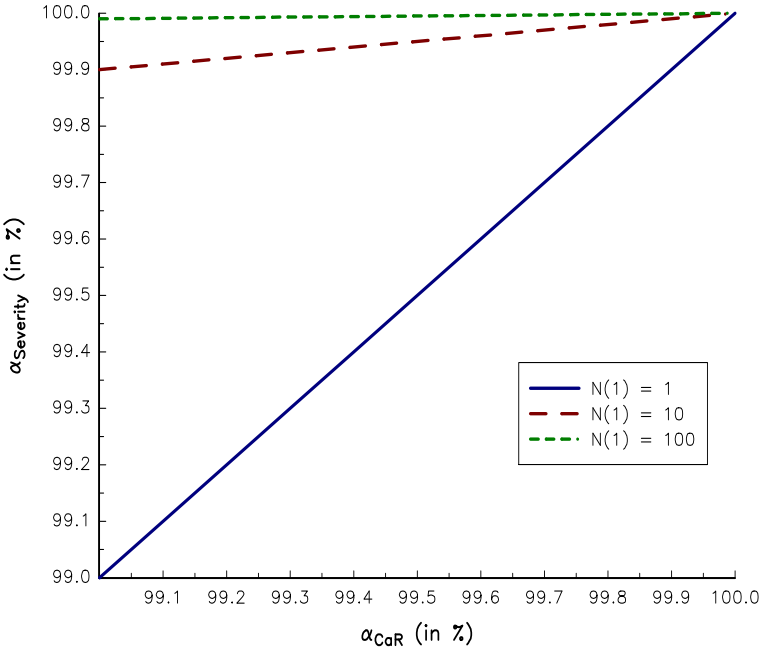


FIGURE 5.13: Relationship between α_{CaR} and α_{Severity}

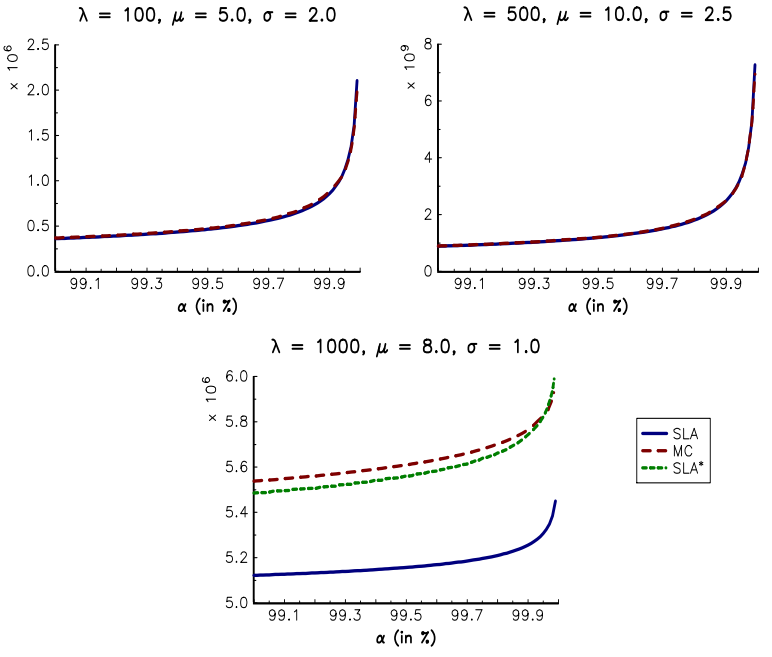


FIGURE 5.14: Numerical illustration of the single loss approximation

The most popular subexponential distributions used in operational risk modeling are the log-gamma, log-logistic, log-normal and Pareto probability distributions (BCBS, 2014f). For instance, if $N(1) \sim \mathcal{P}(\lambda)$ and $X_i \sim \mathcal{LN}(\mu, \sigma^2)$, we obtain:

$$\text{EL} = \lambda \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

and:

$$\text{UL}(\alpha) = \exp\left(\mu + \sigma\Phi^{-1}\left(1 - \frac{1-\alpha}{\lambda}\right)\right) - \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

In Figure 5.14, we report the results of some experiments for different values of parameters. In the top panels, we assume that $\lambda = 100$, $\mu = 5.0$ and $\sigma = 2.0$ (left panel), and $\lambda = 500$, $\mu = 10.0$ and $\sigma = 2.5$ (right panel). These two examples correspond to medium severity/low frequency and high severity/low frequency events. In these cases, we obtain a good approximation. In the bottom panel, the parameters are $\lambda = 1000$, $\mu = 8.0$ and $\sigma = 1.0$. The approximation does not work very well, because we have a low severity/high frequency events and the risk can then not be explained by an extreme single loss. The underestimation of the capital-at-risk is due to the underestimation of the number of losses. In fact, with low severity/high frequency events, the risk is not to face a large single loss, but to have a high number of losses in the year. This is why it is better to approximate the capital-at-risk with the following formula:

$$\mathbf{G}^{-1}(\alpha) \approx \left(\mathbf{P}^{-1}(\alpha) - 1\right)\mathbb{E}[X_i] + \mathbf{F}^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N(1)]}\right)$$

where $\mathbf{P}$ is the cumulative distribution function of the counting process $N(1)$. In Figure 5.14, we have also reported this approximation SLA* for the third example. We verify that it gives better results for high frequency events than the classic approximation.

5.3.3.3 Aggregation issues

We recall that the loss at the bank level is equal to:

$$L = \sum_{k=1}^{K} S_k$$

where S_k is the aggregate loss of the k^{th} cell of the mapping matrix. For instance, if the matrix is composed of the eight business lines (BL) and seven even types (ET) of the Basel II classification, we have $L = \sum_{k \in \mathcal{K}} S_k$ where $\mathcal{K} = \{(\text{BL}_{k_1}, \text{ET}_{k_2}), k_1 = 1, \ldots, 8; k_2 = 1, \ldots, 7\}$. Let $\text{CaR}_{k_1,k_2}(\alpha)$ be the capital charge calculated for the business line k_1 and the event type k_2. We have:

$$\text{CaR}_{k_1,k_2}(\alpha) = \mathbf{G}^{-1}_{k_1,k_2}(\alpha)$$

One solution to calculate the capital charge at the bank level is to sum up all the capital charges:

$$\begin{aligned}\text{CaR}(\alpha) &= \sum_{k_1=1}^{8}\sum_{k_2=1}^{7}\text{CaR}_{k_1,k_2}(\alpha) \\ &= \sum_{k_1=1}^{8}\sum_{k_2=1}^{7}\mathbf{G}^{-1}_{k_1,k_2}(\alpha)\end{aligned}$$

[16] In Chapter 12, we will see that such transformation is common in extreme value theory.

From a theoretical point of view, this is equivalent to assume that all the aggregate losses S_k are perfectly correlated. This approach is highly conservative and ignores diversification effects between business lines and event types.

Let us consider the two-dimensional case:

$$\begin{aligned} L &= S_1 + S_2 \\ &= \sum_{i=1}^{N_1} X_i + \sum_{j=1}^{N_2} Y_j \end{aligned}$$

In order to take into account the dependence between the two aggregate losses S_1 and S_2, we can assume that frequencies N_1 and N_2 are correlated or severities X_i and Y_j are correlated. Thus, the aggregate loss correlation $\rho(S_1, S_2)$ depends on two key parameters:

- the frequency correlation $\rho(N_1, N_2)$;
- the severity correlation $\rho(X_i, Y_j)$.

For example, we should observe that historically, the number of external fraud events is high (respectively low) when the number of internal fraud events is also high (respectively low). Severity correlation is more difficult to justify. In effect, a basic feature of the LDA model requires assuming that individual losses are jointly independent. Therefore it is conceptually difficult to assume simultaneously severity independence within each class of risk and severity correlation between two classes. By assuming that $\rho(X_i, Y_j) = 0$, Frachot *et al.* (2004) find an upper bound of the aggregate loss correlation. We have:

$$\begin{aligned} \operatorname{cov}(S_1, S_2) &= \mathbb{E}[S_1 S_2] - \mathbb{E}[S_1] \cdot \mathbb{E}[S_2] \\ &= \mathbb{E}\left[\sum\nolimits_{i=1}^{N_1} X_i \cdot \sum\nolimits_{j=1}^{N_2} Y_j\right] - \mathbb{E}\left[\sum\nolimits_{i=1}^{N_1} X_i\right] \cdot \mathbb{E}\left[\sum\nolimits_{j=1}^{N_2} Y_j\right] \\ &= \mathbb{E}[N_1 N_2] \cdot \mathbb{E}[X_i] \cdot \mathbb{E}[Y_j] - \mathbb{E}[N_1] \cdot \mathbb{E}[X_i] \cdot \mathbb{E}[N_2] \cdot \mathbb{E}[Y_j] \\ &= (\mathbb{E}[N_1 N_2] - \mathbb{E}[N_1] \cdot \mathbb{E}[N_2]) \cdot \mathbb{E}[X_i] \cdot \mathbb{E}[Y_j] \end{aligned}$$

and:

$$\rho(S_1, S_2) = \frac{(\mathbb{E}[N_1 N_2] - \mathbb{E}[N_1] \cdot \mathbb{E}[N_2]) \cdot \mathbb{E}[X_i] \cdot \mathbb{E}[Y_j]}{\sqrt{\operatorname{var}(S_1) \cdot \operatorname{var}(S_2)}}$$

If we assume that the counting processes N_1 and N_2 are Poisson processes with parameters λ_1 and λ_2, we obtain:

$$\rho(S_1, S_2) = \rho(N_1, N_2) \cdot \eta(X_i) \cdot \eta(Y_j)$$

where:

$$\begin{aligned} \eta(X) &= \frac{\mathbb{E}[X]}{\sqrt{\mathbb{E}[X^2]}} \\ &= \frac{1}{\sqrt{1 + \operatorname{CV}^2(X)}} \leq 1 \end{aligned}$$

Here $\operatorname{CV}(X) = \sigma(X) / \mathbb{E}[X]$ denotes the coefficient of variation of the random variable X. As a result, aggregate loss correlation is always lower than frequency correlation:

$$0 \leq \rho(S_1, S_2) \leq \rho(N_1, N_2) \leq 1$$

We deduce that an upper bound of the aggregate loss correlation is equal to:

$$\rho^+ = \eta(X_i) \cdot \eta(Y_j)$$

For high severity events, severity independence likely dominates frequency correlation and we obtain $\rho^+ \simeq 0$ because $\eta(X_i) \simeq 0$.

Let us consider the example of log-normal severity distributions. We have:

$$\rho^+ = \exp\left(-\frac{1}{2}\sigma_X^2 - \frac{1}{2}\sigma_Y^2\right)$$

We notice that this function is decreasing with respect to σ_X and σ_Y. Figure 5.15 shows the relationship between σ_X, σ_Y and ρ^+. We verify that ρ^+ is small when σ_X and σ_Y take large values. For instance, if $\sigma_X = \sigma_Y = 2$, the aggregate loss correlation is lower than 2%.

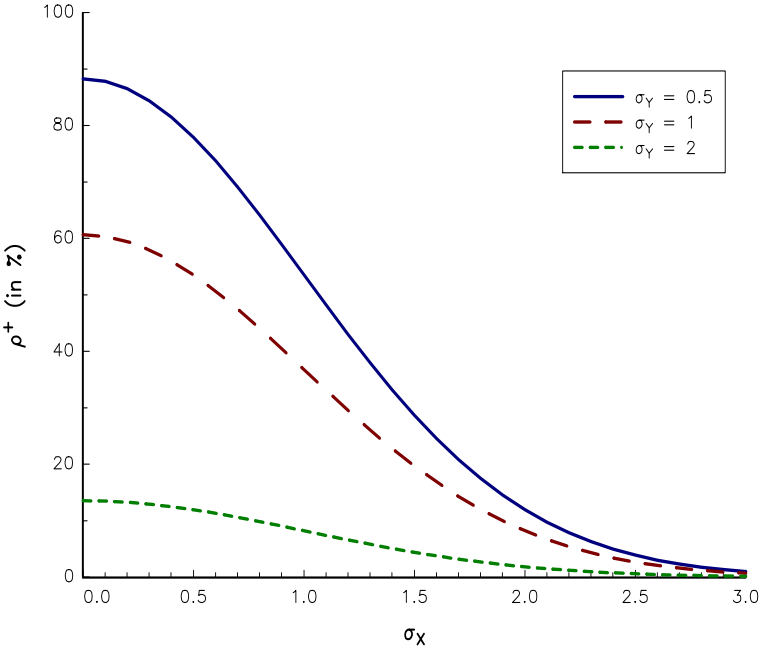


FIGURE 5.15: Upper bound ρ^+ of the aggregate loss correlation

There are two ways to take into account correlations for computing the capital charge of the bank. The first approach is to consider the normal approximation:

$$\mathrm{CaR}(\alpha) = \sum_k \mathrm{EL}_k + \sqrt{\sum_{k,k'} \rho_{k,k'} \cdot (\mathrm{CaR}_k(\alpha) - \mathrm{EL}_k) \cdot (\mathrm{CaR}_{k'}(\alpha) - \mathrm{EL}_{k'})}$$

where $\rho_{k,k'}$ is the correlation between the cells k and k' of the mapping matrix. The second approach consists in introducing the dependence between the aggregate losses using a copula function $\mathbf{C}$. The joint distribution of $(S_1, \ldots, S_K)$ has the following form:

$$\Pr\{S_1 \leq s_1, \ldots, S_K \leq s_K\} = \mathbf{C}(\mathbf{G}_1(s_1), \ldots, \mathbf{G}_K(s_K))$$

where $\mathbf{G}_k$ is the cumulative distribution function of the k^{th} aggregate loss S_k. In this case, the quantile of the random variable $L = \sum_{k=1}^K S_k$ is estimated using Monte Carlo simulations. The difficulty comes from the fact that the distributions $\mathbf{G}_k$ have no analytical expression. The solution is then to use the method of empirical distributions, which is presented on page 806.

5.3.4 Incorporating scenario analysis

The concept of scenario analysis should deserve further clarification. Roughly speaking, when we refer to scenario analysis, we want to express the idea that banks' experts and experienced managers have some reliable intuitions on the riskiness of their business and that these intuitions are not entirely reflected in the bank's historical internal data. As a first requirement, we expect that experts should have the opportunity to give their approval to capital charge results. In a second step, one can imagine that experts' intuitions are directly plugged into severity and frequency estimations. Experts' intuition can be captured through scenario building. More precisely, a scenario is given by a potential loss amount and the corresponding probability of occurrence. As an example, an expert may assert that a loss of one million dollars or higher is expected to occur once every (say) 5 years. This is a valuable information in many cases, either when loss data are rare and do not allow for statistically sound results or when historical loss data are not sufficiently forward-looking. In this last case, scenario analysis allows to incorporate external loss data.

In what follows, we show how scenarios can be translated into restrictions on the parameters of frequency and severity distributions. Once these restrictions have been identified, a calibration strategy can be designed where parameters are calibrated by maximizing some standard criterion subject to these constraints. As a result, parameter estimators can be seen as a mixture of the internal data-based estimator and the scenario-based implied estimator.

5.3.4.1 Probability distribution of a given scenario

We assume that the number of losses $N(t)$ is a Poisson process with intensity λ. Let τ_n be the arrival time of the n^{th} loss:

$$\tau_n = \inf\{t \geq 0 : N(t) = n\}$$

We know that the durations $T_n = \tau_n - \tau_{n-1}$ between two consecutive losses are *iid* exponential random variables with parameter λ. We recall that the losses X_n are also *iid* with distribution $\mathbf{F}$. We note now $T_n(x)$ the duration between two losses exceeding x. It is obvious that the durations are *iid*. It suffice now to characterize $T_1(x)$. By using the fact that a finite sum of exponential times is an Erlang distribution, we have:

$$\begin{aligned}
\Pr\{T_1(x) > t\} &= \sum_{n\geq 1} \Pr\{\tau_n > t; X_1 < x, \ldots, X_{n-1} < x; X_n \geq x\} \\
&= \sum_{n\geq 1} \Pr\{\tau_n > t\} \cdot \mathbf{F}(x)^{n-1} \cdot (1 - \mathbf{F}(x)) \\
&= \sum_{n\geq 1} \mathbf{F}(x)^{n-1} \cdot (1 - \mathbf{F}(x)) \cdot \left(\sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right) \\
&= (1 - \mathbf{F}(x)) \cdot \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \left(\sum_{n=k}^{\infty} \mathbf{F}(x)^n \right) \\
&= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \mathbf{F}(x)^k \\
&= e^{-\lambda(1-\mathbf{F}(x))t}
\end{aligned}$$

We deduce that $T_n(x)$ follows an exponential distribution with parameter $\lambda(x) = \lambda(1 - \mathbf{F}(x))$. The average duration between two losses exceeding x is also the mean of $T_n(x)$:

$$\mathbb{E}[T_n(x)] = \frac{1}{\lambda(1 - \mathbf{F}(x))}$$

Example 58 *We assume that the annual number of losses follows a Poisson distribution where* $\lambda = 5$ *and the severity of losses are log-normal* $\mathcal{LN}(9, 4)$.

In Figure 5.16, we simulate the corresponding Poisson process $N(t)$ and also the events whose loss is larger than \$20 000 and \$50 000. We then show the exponential distribution[17] of $T_n(x)$.

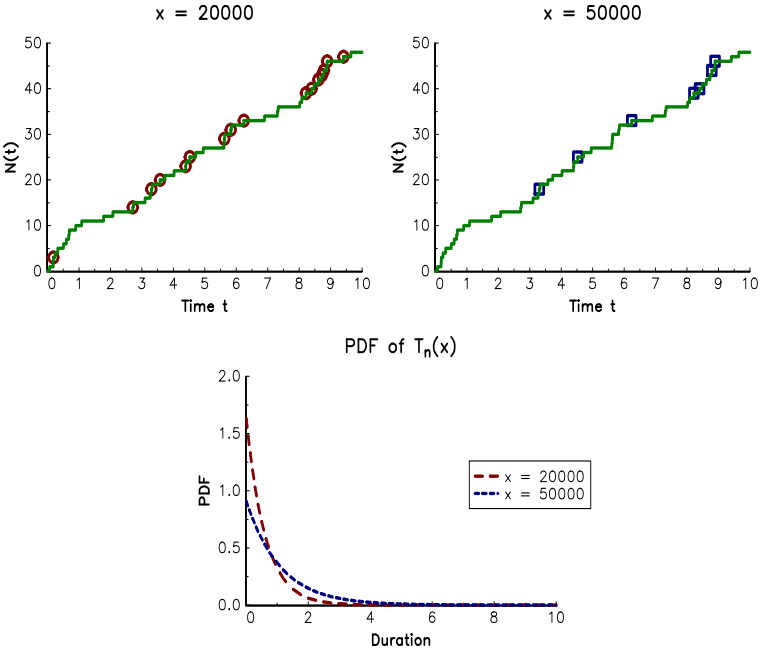


FIGURE 5.16: Simulation of the Poisson process $N(t)$ and peak over threshold events

5.3.4.2 Calibration of a set of scenarios

Let us consider a scenario defined as "a loss of x or higher occurs once every d years". By assuming a compound Poisson distribution with a parametric severity distribution $\mathbf{F}(x;\theta)$, λ is the average number of losses per year, $\lambda(x) = \lambda(1 - \mathbf{F}(x;\theta))$ is the average number of losses higher than x and $1/\lambda(x)$ is the average duration between two losses exceeding x. As a result, for a given scenario (x, d), parameters (λ, θ) must satisfy:

$$d = \frac{1}{\lambda(1 - \mathbf{F}(x;\theta))}$$

Suppose that we face different scenarios $\{(x_s, d_s), s = 1, \ldots, n_S\}$. We may estimate the implied parameters underlying the expert judgements using the quadratic criterion:

$$\left(\hat{\lambda}, \hat{\theta}\right) = \arg\min \sum_{s=1}^{n_S} w_s \cdot \left(d_s - \frac{1}{\lambda(1 - \mathbf{F}(x_s;\theta))}\right)^2$$

[17] For the parameter $\lambda(x)$, we have:

$$\lambda\left(2 \times 10^4\right) = 5 \times \left(1 - \Phi\left(\frac{\ln\left(2 \times 10^4\right) - 9}{2}\right)\right) = 1.629$$

and $\lambda\left(5 \times 10^4\right) = 0.907$.

where w_s is the weight of the s^{th} scenario. The previous approach belongs to the method of moments. As a result, we can show that the optimal weights w_s correspond to the inverse of the variance of d_s:

$$\begin{aligned} w_s &= \frac{1}{\text{var}\left(d_s\right)} \\ &= \lambda\left(1-\mathbf{F}\left(x_s;\theta\right)\right) \end{aligned}$$

To solve the previous optimization program, we proceed by iterations. Let $\left(\hat{\lambda}_m, \hat{\theta}_m\right)$ be the solution of the minimization program:

$$\left(\hat{\lambda}_m, \hat{\theta}_m\right) = \arg\min \sum_{j=1}^{p} \hat{\lambda}_{m-1} \cdot \left(1-\mathbf{F}\left(x_s;\hat{\theta}_{m-1}\right)\right) \cdot \left(d_s - \frac{1}{\lambda\left(1-\mathbf{F}\left(x_s;\theta\right)\right)}\right)^2$$

Under some conditions, the estimator $\left(\hat{\lambda}_m, \hat{\theta}_m\right)$ converge to the optimal solution. We also notice that we can simplify the optimization program by using the following approximation:

$$w_s = \frac{1}{\text{var}\left(d_s\right)} = \frac{1}{\mathbb{E}\left[d_s\right]} \simeq \frac{1}{d_s}$$

Example 59 *We assume that the severity distribution is log-normal and consider the following set of expert's scenarios:*

x_s *(in \$ mn)*	1	2.5	5	7.5	10	20
d_s *(in years)*	1/4	1	3	6	10	40

If $w_s = 1$, we obtain $\hat{\lambda} = 43.400$, $\hat{\mu} = 11.389$ and $\hat{\sigma} = 1.668$ (#1). Using the approximation $w_s \simeq 1/d_s$, the estimates become $\hat{\lambda} = 154.988$, $\hat{\mu} = 10.141$ and $\hat{\sigma} = 1.855$ (#2). Finally, the optimal estimates are $\hat{\lambda} = 148.756$, $\hat{\mu} = 10.181$ and $\hat{\sigma} = 1.849$ (#3). In the table below, we report the estimated values of the duration. We notice that they are close to the expert's scenarios.

x_s (in \$ mn)	1	2.5	5	7.5	10	20
#1	0.316	1.022	2.964	5.941	10.054	39.997
#2	0.271	0.968	2.939	5.973	10.149	39.943
#3	0.272	0.970	2.941	5.974	10.149	39.944

Remark 64 *We can combine internal loss data, expert's scenarios and external loss data*[18] *by maximizing the penalized likelihood:*

$$\begin{aligned} \hat{\theta} = \arg\max \quad & \varpi_{\text{internal}} \cdot \boldsymbol{\ell}\left(\theta\right) - \varpi_{\text{expert}} \cdot \sum_{s=1}^{n_S} w_s \left(d_s - \frac{1}{\lambda\left(1-\mathbf{F}\left(x_s;\theta\right)\right)}\right)^2 - \\ & \varpi_{\text{external}} \cdot \sum_{s=1}^{n_S^\star} w_s^\star \left(d_s^\star - \frac{1}{\lambda\left(1-\mathbf{F}\left(x_s^\star;\theta\right)\right)}\right)^2 \end{aligned}$$

where ϖ_{internal}, ϖ_{expert} *and* ϖ_{external} *are the weights reflecting the confidence placed on internal loss data, expert's scenarios and external loss data.*

[18] In this case, each external loss is treated as an expert's scenario.

5.3.5 Stability issue of the LDA model

One of the big issues of AMA (and LDA) models is their stability. It is obvious that the occurrence of a large loss changes dramatically the estimated capital-at-risk as explained by Ames *et al.* (2015):

> "*Operational risk is fundamentally different from all other risks taken on by a bank. It is embedded in every activity and product of an institution, and in contrast to the conventional financial risks (e.g. market, credit) is harder to measure and model, and not straight forwardly eliminated through simple adjustments like selling off a position. While it varies considerably, operational risk tends to represent about 10-30% of the total risk pie, and has grown rapidly since the 2008-09 crisis. It tends to be more fat-tailed than other risks, and the data are poorer. As a result, models are fragile – small changes in the data have dramatic impacts on modeled output – and thus required operational risk capital is unstable*".

In this context, the Basel Committee has decided to review the different measurement approaches to calculate the operational risk capital. In Basel III, advanced measurement approaches have been dropped. This decision marks a serious setback for operational risk modeling. The LDA model continues to be used by Basel II jurisdictions, and will continue to be used by large international banks, because it is the only way to assess an economic capital using internal loss data. Moreover, internal losses continue to be collected by banks in order to implement the SMA of Basel III. Finally, the LDA model will certainly become the standard model for satisfying Pillar 2 requirements. However, solutions for stabilizing the LDA model can only be partial and even hazardous or counter-intuitive, because it ignores the nature of operational risk.

5.4 Exercises

5.4.1 Estimation of the loss severity distribution

We consider a sample of n individual losses $\{x_1, \ldots, x_n\}$. We assume that they can be described by different probability distributions:

(i) X follows a log-normal distribution $\mathcal{LN}\left(\mu, \sigma^2\right)$.

(ii) X follows a Pareto distribution $\mathcal{P}\left(\alpha, x_-\right)$ defined by:

$$\Pr\{X \leq x\} = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

where $x \geq x_-$ and $\alpha > 0$.

(iii) X follows a gamma distribution $\mathcal{G}\left(\alpha, \beta\right)$ defined by:

$$\Pr\{X \leq x\} = \int_0^x \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma\left(\alpha\right)} \mathrm{d}t$$

where $x \geq 0$, $\alpha > 0$ and $\beta > 0$.

(iv) The natural logarithm of the loss X follows a gamma distribution: $\ln X \sim \mathcal{G}\left(\alpha; \beta\right)$.

1. We consider the case *(i)*.

   (a) Show that the probability density function is:

   $$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)$$

   (b) Calculate the first two moments of X. Deduce the orthogonal conditions of the generalized method of moments.

   (c) Find the maximum likelihood estimators $\hat{\mu}$ and $\hat{\sigma}$.

2. We consider the case *(ii)*.

   (a) Calculate the first two moments of X. Deduce the GMM conditions for estimating the parameter α.

   (b) Find the maximum likelihood estimator $\hat{\alpha}$.

3. We consider the case *(iii)*. Write the log-likelihood function associated to the sample of individual losses $\{x_1, \ldots, x_n\}$. Deduce the first-order conditions of the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$.

4. We consider the case *(iv)*. Show that the probability density function of X is:

   $$f(x) = \frac{\beta^\alpha (\ln x)^{\alpha-1}}{\Gamma(\alpha)\, x^{\beta+1}}$$

   What is the support of this probability density function? Write the log-likelihood function associated to the sample of individual losses $\{x_1, \ldots, x_n\}$.

5. We now assume that the losses $\{x_1, \ldots, x_n\}$ have been collected beyond a threshold H meaning that $X \geq H$.

   (a) What does the generalized method of moments become in the case *(i)*?

   (b) Calculate the maximum likelihood estimator $\hat{\alpha}$ in the case *(ii)*.

   (c) Write the log-likelihood function in the case *(iii)*.

5.4.2 Estimation of the loss frequency distribution

We consider a dataset of individual losses $\{x_1, \ldots, x_n\}$ corresponding to a sample of T annual loss numbers $\{N_{Y_1}, \ldots, N_{Y_T}\}$. This implies that:

$$\sum_{t=1}^{T} N_{Y_t} = n$$

If we measure the number of losses per quarter $\{N_{Q_1}, \ldots, N_{Q_{4T}}\}$, the previous equation becomes:

$$\sum_{t=1}^{4T} N_{Q_t} = n$$

1. We assume that the annual number of losses follows a Poisson distribution $\mathcal{P}(\lambda_Y)$. Calculate the maximum likelihood estimator $\hat{\lambda}_Y$ associated to the sample $\{N_{Y_1}, \ldots, N_{Y_T}\}$.

2. We assume that the quarterly number of losses follows a Poisson distribution $\mathcal{P}(\lambda_Q)$. Calculate the maximum likelihood estimator $\hat{\lambda}_Q$ associated to the sample $\{N_{Q_1}, \ldots, N_{Q_{4T}}\}$.

3. What is the impact of considering a quarterly or annual basis on the computation of the capital charge?

4. What does this result become if we consider a method of moments based on the first moment?

5. Same question if we consider a method of moments based on the second moment.

5.4.3 Using the method of moments in operational risk models

1. Let $N(t)$ be the number of losses for the time interval $[0, t]$. We note $\{N_1, \ldots, N_T\}$ a sample of $N(t)$ and we assume that $N(t)$ follows a Poisson distribution $\mathcal{P}(\lambda)$. We recall that:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

   (a) Calculate the first moment $\mathbb{E}[N(t)]$.

   (b) Show the following result:

$$\mathbb{E}\left[\prod_{i=0}^{m} (N(t) - i)\right] = \lambda^{m+1}$$

   Then deduce the variance of $N(t)$.

   (c) Propose two estimators based on the method of moments.

2. Let S be the random sum:

$$S = \sum_{i=0}^{N(t)} X_i$$

   where $X_i \sim \mathcal{LN}(\mu, \sigma^2)$, $X_i \perp X_j$ and $N(t) \sim \mathcal{P}(\lambda)$.

   (a) Calculate the mathematical expectation $\mathbb{E}[S]$.

   (b) We recall that:

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i \neq j} x_i x_j$$

   Show that:

$$\text{var}(S) = \lambda \exp\left(2\mu + 2\sigma^2\right)$$

   (c) How can we estimate μ and σ if we have already calibrated λ?

3. We assume that the annual number of losses follows a Poisson distribution $\mathcal{P}(\lambda)$. We also assume that the individual losses are independent and follow a Pareto distribution $\mathcal{P}(\alpha, x_-)$ defined by:

$$\Pr\{X \leq x\} = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

   where $x \geq x_-$ and $\alpha > 1$.

(a) Show that the duration between two consecutive losses that are larger than ℓ is an exponential distribution with parameter $\lambda x_-^\alpha \ell^{-\alpha}$.

(b) How can we use this result to calibrate experts' scenarios?

5.4.4 Calculation of the Basel II required capital

We consider the simplified balance sheet of a bank, which is described below.

1. In the Excel file, we provide the price evolution of stocks A and B. The trading portfolio consists of 10 000 shares A and 25 000 shares B. Calculate the daily historical VaR of this portfolio by assuming that the current stock prices are equal to \$105.5 and \$353. Deduce the capital charge for market risk assuming that the VaR has not fundamentally changed during the last 3 months[19].

2. We consider that the credit portfolio of the bank can be summarized by 4 meta-credits whose characteristics are the following:

	Sales	EAD	PD	LGD	M
Bank		\$80 mn	1%	75%	1.0
Corporate	\$500 mn	\$200 mn	5%	60%	2.0
SME	\$30 mn	\$50 mn	2%	40%	4.5
Mortgage		\$50 mn	9%	45%	
Retail		\$100 mn	4%	85%	

   Calculate the IRB capital charge for the credit risk.

3. We assume that the bank is exposed to a single operational risk. The severity distribution is a log-normal probability distribution $\mathcal{LN}(8,4)$, whereas the frequency distribution is the following discrete probability distribution:

$$\begin{aligned} \Pr\{N=5\} &= 60\% \\ \Pr\{N=10\} &= 40\% \end{aligned}$$

   Calculate the AMA capital charge for the operational risk.

4. Deduce the capital charge of the bank and the capital ratio knowing that the capital of the bank is equal to \$70 mn.

5.4.5 Parametric estimation of the loss severity distribution

1. We assume that the severity losses are log-logistic distributed $X_i \sim \mathcal{LL}(\alpha,\beta)$ where:

$$\mathbf{F}(x;\alpha,\beta) = \frac{(x/\alpha)^\beta}{1+(x/\alpha)^\beta}$$

   (a) Find the density function.

   (b) Deduce the log-likelihood function of the sample $\{x_1,\ldots,x_n\}$.

   (c) Show that the ML estimators satisfy the following first-order conditions:

$$\begin{cases} \sum_{i=1}^n \mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right) = n/2 \\ \sum_{i=1}^n \left(2\mathbf{F}\left(x_i;\hat{\alpha},\hat{\beta}\right)-1\right)\ln x_i = n/\hat{\beta} \end{cases}$$

[19] The multiplication coefficient ξ is set equal to 0.5.

(d) The sample of loss data is 2 918, 740, 3 985, 2 827, 2 839, 6 897, 7 665, 3 766, 3 107 and 3 304. Verify that $\hat{\alpha} = 3\,430.050$ and $\hat{\beta} = 3.315$ are the ML estimates.

(e) What does the log-likelihood function of the sample $\{x_1, \ldots, x_n\}$ become if we assume that the losses were collected beyond a threshold H?

5.4.6 Mixed Poisson process

1. We consider the mixed poisson process where $N(t) \sim \mathcal{P}(\Lambda)$ and Λ is a random variable. Show that:
$$\text{var}(N(t)) = \mathbb{E}[N(t)] + \text{var}(\Lambda)$$

2. Deduce that $\text{var}(N(t)) \geq \mathbb{E}[N(t)]$. Determine the probability distribution Λ such that the equality holds. Let $\varphi(n)$ be the following ratio:
$$\varphi(n) = \frac{(n+1) \cdot p(n+1)}{p(n)}$$
Show that $\varphi(n)$ is constant.

3. We assume that $\Lambda \sim \mathcal{G}(\alpha, \beta)$.
   (a) Calculate $\mathbb{E}[N(t)]$ and $\text{var}(N(t))$.
   (b) Show that $N(t)$ has a negative binomial distribution $\mathcal{NB}(r, p)$. Calculate the parameters r and p with respect to α and β.
   (c) Show that $\varphi(n)$ is an affine function.

4. We assume that $\Lambda \sim \mathcal{E}(\lambda)$.
   (a) Calculate $\mathbb{E}[N(t)]$ and $\text{var}(N(t))$.
   (b) Show that $N(t)$ has a geometric distribution $\mathcal{G}(p)$. Determine the parameter p.

Chapter 6

Liquidity Risk

Liquidity is a long-standing issue and also an elusive concept (Grossman and Miller, 1988). It cannot be observed directly, because it measures the ease of trading an asset. More precisely, it measures the asset's ability to be sold as soon as possible without causing a significant price movement. This is why it is difficult to capture liquidity in a single measure (bid-ask spread, trading volume, etc.). Moreover, liquidity risk generally refers to two related notions: market liquidity and funding liquidity. Market liquidity concerns assets. For instance, the most liquid asset is cash because it can always be used easily and immediately. Many stocks and sovereign bonds are considered fairly liquid, because they can be sold in the day. On the contrary, private equity and real estate are less liquid assets, because it can take months to sell them. Funding liquidity concerns asset liability mismatch due to liquidity and maturity transformation activities. According to Drehmann and Nikolaou (2013), funding liquidity is defined "*as the ability to settle obligations with immediacy. It follows that, a bank is illiquid if it is unable to settle obligations in time*". The concept of funding liquidity is of course important for banks, but also for other financial entities (insurance companies, asset managers, hedge funds, etc.).

This chapter is organized as follows. The first section is dedicated to the measurement of asset liquidity. In the second section, we consider how funding liquidity affects the risk of financial institutions. The last section presents the regulatory framework for managing liquidity risk in a bank. This chapter may be viewed as an introduction of liquidity risk, which is developed in Chapter 7, which focuses on asset liability management risk and is complemented by Chapter 8, which is dedicated to the systemic risk, because liquidity and systemic risks are highly connected.

6.1 Market liquidity

Sarr and Lybek (2002) propose to classify market liquidity measures into four categories: (1) transaction cost measures, (2) volume-based measures, (3) equilibrium price-based measures, and (4) market-impact measures. The choice of one measure depends on the objective of the liquidity measurement. The first category is useful for investors, who would like to know the cost of selling or buying immediately a security (stocks, bonds, futures, etc.). The second category is related to the breadth of the market and measures the trading activity of a security. The last two categories (price-based and market-impact measures) concern more the resilience and the efficiency of the market. The underlying idea is to understand how trading prices can move away from fundamental prices. By construction, these last two categories are more developed by academics whereas investors are more concerned by the first two categories.

6.1.1 Transaction cost versus volume-based measures

6.1.1.1 Bid-ask spread

The traditional liquidity measure is the bid-ask quoted spread $\mathbf{S}_t$, which is defined by:

$$\mathbf{S}_t = \frac{P_t^{\text{ask}} - P_t^{\text{bid}}}{P_t^{\text{mid}}}$$

where P_t^{ask}, P_t^{bid} and P_t^{mid} are the ask, bid and mid[1] quotes for a given security at time t. By construction, the bid-ask spread can only be computed in an organized exchange with order books. Here, the ask price corresponds to the lowest price of sell orders, whereas the bid price is the highest price of buy orders. In this context, $\mathbf{S}_t$ may be viewed as a transaction cost measure and is the standard liquidity measure in equity markets.

TABLE 6.1: An example of a limit order book

i^{th} limit	Buy orders		Sell orders	
	$Q_t^{\text{bid,i}}$	$P_t^{\text{bid,i}}$	$Q_t^{\text{ask,i}}$	$P_t^{\text{ask,i}}$
1	65 201	26.325	70 201	26.340
2	85 201	26.320	116 201	26.345
3	105 201	26.315	107 365	26.350
4	76 500	26.310	35 000	26.355
5	20 000	26.305	35 178	26.360

Example 60 *In Table 6.1, we provide a snapshot of the limit order book of the Lyxor Euro Stoxx 50 ETF recorded at NYSE Euronext Paris*[2]*. $Q_{t_j}^{\text{bid,i}}$ and $P_{t_j}^{\text{bid,i}}$ (resp. $Q_{t_j}^{\text{ask,i}}$ and $P_{t_j}^{\text{ask,i}}$) indicate the quantity and the price of the buyer (resp. the seller) for the i^{th} limit.*

This limit order book is represented in Figure 6.1, where the x-axis represents the quoted prices and the y-axis represented the buy and sell quantities. The bid and ask prices correspond to the prices of the best limit. We have $P_t^{\text{bid}} = 26.325$ and $P_t^{\text{ask}} = 26.340$, implying that the mid price is equal to:

$$P_t^{\text{mid}} = \frac{26.325 + 26.340}{2} = 26.3325$$

We deduce that the bid-ask spread is:

$$\mathbf{S}_t = \frac{26.340 - 26.325}{26.3325} = 5.696 \text{ bps}$$

There are other variants of the bid-ask spread, which do not use quoted prices, but traded prices. For instance, the *effective spread* is equal to:

$$\mathbf{S}_\tau^e = 2 \left| \frac{P_\tau - P_t^{\text{mid}}}{P_t^{\text{mid}}} \right|$$

[1] We have:

$$P_t^{\text{mid}} = \frac{P_t^{\text{ask}} + P_t^{\text{bid}}}{2}$$

[2] The corresponding date is 14:00:00 and 56,566 micro seconds on 28 December 2012.

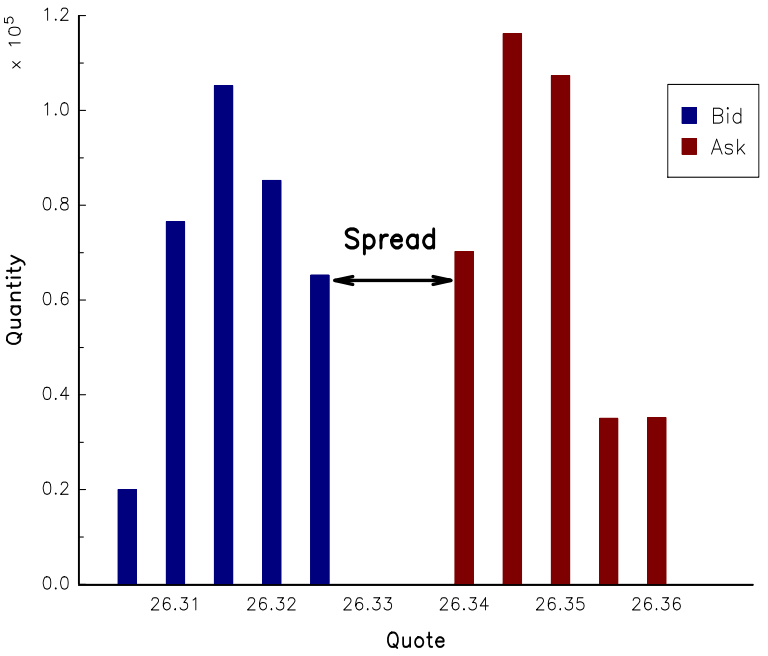


FIGURE 6.1: An example of a limit order book

where τ is the trade index, P_τ is the price of the τ^{th} trade and P_τ^{mid} is the midpoint of market quote calculated at the time t of the τ^{th} trade. In a similar way, the *realized spread* uses the same formula than the effective spread, but replaces P_t^{mid} by the mid quote of the security at time $t + \Delta$:

$$\mathbf{S}_\tau^r = 2 \left| \frac{P_\tau - P_{t+\Delta}^{\text{mid}}}{P_{t+\Delta}^{\text{mid}}} \right|$$

Generally, Δ is set to five minutes. The realized spread $\mathbf{S}_\tau^r$ represents the temporary component of the effective spread (Goyenko *et al.*, 2009). In this case, $P_{t+\Delta}$ may be viewed as the equilibrium price of the security after the trade[3]. In particular, if the trade has a price impact, we have $P_{t+\Delta}^{\text{mid}} \neq P_t^{\text{mid}}$ and $\mathbf{S}_\tau^r \not\approx \mathbf{S}_\tau^e$.

6.1.1.2 Trading volume

The second popular measure is the trading volume $\mathbf{V}_t$, which indicates the dollar value of the security exchanged during the period t:

$$\mathbf{V}_t = \sum_{\tau \in t} Q_\tau P_\tau$$

[3]Another variant of the realized spread is the signed spread:

$$\mathbf{S}_\tau^r = 2 s_\tau \left(\frac{P_\tau - P_{t+\Delta}}{P_{t+\Delta}} \right)$$

where:

$$s_\tau = \begin{cases} +1 & \text{if the trade is a buy} \\ -1 & \text{if the trade is a sell} \end{cases}$$

where Q_τ and P_τ are the τ^{th} quantity and price traded during the period. Generally, we consider a one-day period and use the following approximation:

$$\mathbf{V}_t \approx Q_t P_t$$

where Q_t is the number of securities traded during the day t and P_t is the closing price of the security.

A related measure is the turnover which is the ratio between the trading volume and the free float market capitalization M_t of the asset:

$$\mathbf{T}_t = \frac{\mathbf{V}_t}{M_t} = \frac{\mathbf{V}_t}{N_t P_t}$$

where N_t is the number of outstanding '*floating*' shares. The asset turnover ratio indicates how many times each share changes hands in a given period[4]. For instance, if the annual turnover is two, this means that the shares have changed hands, on average, two times during the year.

Example 61 *We consider a stock, whose average daily volume Q_t is equal to* 1 200 *shares whereas the total number of shares N_t is equal to* 500 000*. We assume that the price is equal to* \$13 500*.*

We deduce that the daily volume is:

$$\mathbf{V}_t = 1\,200 \times 13\,500 = \$16.2 \text{ mn}$$

Because the market capitalization M_t is equal to \$6.75 bn, the daily turnover represents 0.24%. It follows that the annualized turnover[5] is about 62%.

6.1.1.3 Liquidation ratio

Another popular measure is the liquidation ratio $\mathcal{LR}(m)$, which measures the proportion of a given position that can be liquidated after m trading days. This statistic depends on the size of the position and the liquidation policy. A simple rule is to define a maximum number of shares that can be sold every day. The market convention is to consider a proportion of the three-month average daily volume (ADV). This serves as a proxy to bound liquidation costs: the higher the proportion of the ADV, the larger the trading costs. Another interesting statistic is the liquidation time $\mathcal{LR}^{-1}(p)$, which is the inverse function of the liquidity ratio. It indicates the number of required trading days in order to liquidate a proportion p of the position.

Example 62 *We consider again Example 61. Suppose that we have a position of* \$30 *mn in this stock. In order to minimize trading impacts, the liquidation policy is set to* 25% *of the average daily volume.*

The liquidity policy implies that we can sell $25\% \times 16.2 = \$4.05$ mn every day. We deduce that:

$$\begin{aligned}
\mathcal{LR}(1) &= \frac{4.05}{30} = 13.5\% \\
\mathcal{LR}(2) &= \frac{2 \times 4.05}{30} = 27\%
\end{aligned}$$

[4] This is why this ratio is generally expressed in an annual basis.
[5] We multiply the daily turnover by a factor of 260.

Finally, we have:

$$\mathcal{LR}(m) = \begin{cases} m \times 13.5\% & \text{if } m \leq 7 \\ 100\% & \text{if } m \geq 8 \end{cases}$$

The liquidation of the position requires 8 trading days.

We now consider a portfolio invested into n assets. We denote $(x_1, \ldots, x_n)$ the number of shares held in the portfolio. Let $P_{i,t}$ be the current price of asset i. The value of the portfolio is equal to $\sum_{i=1}^{n} x_i P_{i,t}$. For each asset that composes the portfolio, we denote x_i^+ the maximum number of shares for asset i that can be sold during a trading day. The number of shares $x_i(m)$ liquidated after m trading days is defined as follows:

$$x_i(m) = \min\left(\left(x_i - \sum_{k=0}^{m-1} x_i(k)\right)^+, x_i^+\right)$$

with $x_i(0) = 0$. The liquidation ratio $\mathcal{LR}(m)$ is then the proportion of the portfolio liquidated after m trading days:

$$\mathcal{LR}(m) = \frac{\sum_{i=1}^{n} \sum_{k=0}^{m} x_i(k) \cdot P_{i,t}}{\sum_{i=1}^{n} x_i \cdot P_{i,t}}$$

TABLE 6.2: Statistics of the liquidation ratio (size = \$10 bn, liquidation policy = 10% of ADV)

Statistics	SPX	SX5E	DAX	NDX	MSCI EM	MSCI INDIA	MSCI EMU SC
m (in days)	Liquidation ratio $\mathcal{LR}(t)$ in %						
1	88.4	12.3	4.8	40.1	22.1	1.5	3.0
2	99.5	24.7	9.6	72.6	40.6	3.0	6.0
5	100.0	58.8	24.1	99.7	75.9	7.6	14.9
10	100.0	90.1	47.6	99.9	93.9	15.1	29.0
α (in %)	Liquidation time $\mathcal{LR}^{-1}(\alpha)$ in days						
50	1	5	11	2	3	37	21
75	1	7	17	3	5	71	43
90	2	10	23	3	9	110	74
99	2	15	29	5	17	156	455

Source: Roncalli and Weisang (2015).

In Table 6.2, we report the liquidation ratio and the liquidation time for several equity index portfolios using a size of \$10 bn and assuming we can sell 10% of the ADV every day[6]. The indices are the S&P 500 index (SPX), Euro Stoxx 50 index (SX5E), DAX index, NASDAQ 100 index (NDX), MSCI EM index, MSCI INDIA index and MSCI EMU Small Cap index. We read the results as follows: $\mathcal{LR}(1)$ is equal to 88.4% for the S&P 500 index meaning that we can liquidate 88.4% (or \$8.84 bn) of the portfolio on the first trading day; $\mathcal{LR}(5)$ is equal to 24.1% for the DAX index meaning that we can liquidate 24.1% of the assets after five trading days; $\mathcal{LR}^{-1}(75\%)$ is equal to 43 for the MSCI EMU Small Cap index meaning that we need 43 trading days to liquidate \$7.5 bn for this portfolio. We observe that the liquidation risk profile is different from one equity index portfolio to another.

TABLE 6.3: Statistics of the liquidation ratio (size = \$10 bn, liquidation policy = 30% of ADV)

Statistics	SPX	SX5E	DAX	NDX	MSCI EM	MSCI INDIA	MSCI EMU SC
t (in days)	Liquidation ratio $\mathcal{LR}(t)$ in %						
1	100.0	37.0	14.5	91.0	55.5	4.5	9.0
2	100.0	67.7	28.9	99.8	81.8	9.1	17.8
5	100.0	99.2	68.6	100.0	98.5	22.6	40.4
10	100.0	100.0	99.6	100.0	100.0	43.1	63.2
α (in %)	Liquidation time $\mathcal{LR}^{-1}(\alpha)$ in days						
50	1	2	4	1	1	13	7
75	1	3	6	1	2	24	15
90	1	4	8	1	3	37	25
99	1	5	10	2	6	52	152

Source: Roncalli and Weisang (2015).

These figures depend on the liquidation policy and the liquidation size. For instance, if we use an average daily volume of 30%, we obtain the results given in Table 6.3. In this case, liquidity ratios are improved. Nevertheless, we continue to observe that all these indices do not present the same liquidity profile. In Figure 6.2, we report the liquidation ratio for different indices. We notice that the liquidity profile is better for the S&P 500 index for a size of \$50 bn than for the Euro Stoxx 50 index for a size of \$10 bn. We also observe that liquidating \$1 bn of MSCI INDIA index is approximately equivalent to liquidating \$10 bn of Euro Stoxx 50 index. These results depend on the free-float market capitalization of each index. For instance, the capitalization of the S&P 500 is equal to \$18 tn at the end of April 2015. This contrasts with the capitalization of the MSCI EMU Small Cap, which is equal to \$448 bn.

6.1.1.4 Liquidity ordering

The bid-ask spread and the daily trading volume are easily available in financial information systems (Bloomberg, Reuters, etc.). They represent two aspects of the liquidity. $\mathbf{S}_t$ is an estimate of the trading cost in the case of small orders. When we consider an order of big size, $\mathbf{S}_t$ is not valid because the order may have an impact on the price and it may also be not possible to trade immediately. In this case, it is better to consider $\mathbf{V}_t$, which gives the average trading activity of the security. Indeed, the investor may compare the size of his order and the depth of the market.

These two statistics may be used to compare the liquidity $\mathcal{L}$ of securities i and j. We will say that the liquidity of security i is better than the liquidity of security j if security i has a lower bid-ask spread:

$$\mathbf{S}_{i,t} \leq \mathbf{S}_{j,t} \Rightarrow \mathcal{L}(i) \succ \mathcal{L}(j)$$

or if security i has a higher trading volume:

$$\mathbf{V}_{i,t} \geq \mathbf{V}_{j,t} \Rightarrow \mathcal{L}(i) \succ \mathcal{L}(j)$$

[6]For the composition of the portfolio and the ADV statistics, Roncalli and Weisang (2015) use the data of 30 April 2015.

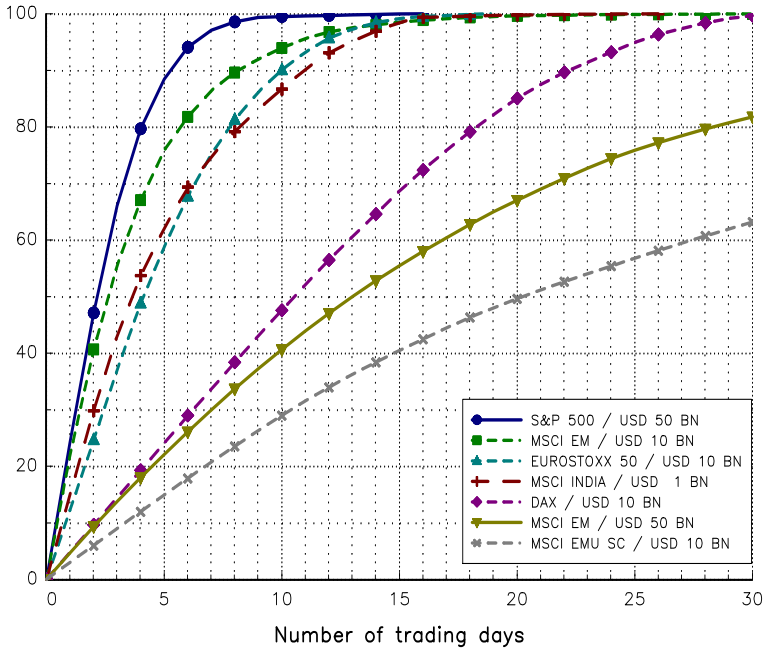


FIGURE 6.2: Comparing the liquidation ratio (in %) between index fund portfolios

Source: Roncalli and Weisang (2015).

It would be wrong to think that the two measures $\mathbf{S}_t$ and $\mathbf{V}_t$ gives the same liquidity ordering. For instance, Roncalli and Zheng (2015) show that it is far from being the case in European ETF markets. In fact, liquidity is a multi-faceted concept and recovers various dimensions. This explains that there is a multiplication of other liquidity measures.

6.1.2 Other liquidity measures

The Hui-Heubel liquidity ratio is a measure of the resilience and the depth. It combines turnover and price impact:

$$\mathbf{H}_t^2 = \frac{1}{\mathbf{T}_t}\left(\frac{P_t^{\text{high}} - P_t^{\text{low}}}{P_t^{\text{low}}}\right)$$

where P_t^{high} and P_t^{low} are the highest and lowest prices during the period t, and $\mathbf{T}_t$ is the turnover observed for the same period. $\mathbf{H}_t^2$ can be calculated on a daily basis or with a higher frequency period. For example, Sarr and Lybek (2002) propose to consider a 5-day period in order to capture medium-term price impacts.

Among price-based measures, Sarr and Lybek (2002) include the variance ratio of Hasbrouck and Schwartz (1988), also called the market efficiency coefficient (MEC), which is the ratio between the annualized variance of long-period returns $R_{t,t+h}$ ($h \gg 1$) and the annualized variance of short-period returns $R_{t,t+1}$:

$$\mathbf{VR} = \frac{\text{var}\,(R_{t,t+h})}{\text{var}\,(R_{t,t+1})}$$

However, this ratio may be not pertinent because it is related to the reversal alternative risk premium or the auto-correlation trading strategy. In fact, this ratio is another measure

of the auto-correlation of asset returns. Goyenko *et al.* (2009) define the price impact as the "*cost of demanding additional instantaneous liquidity*". In this case, it corresponds to the derivative of the spread with respect to the order size:

$$\mathbf{PI} = \frac{\bar{S}^e\,(\text{big}) - \bar{S}^e\,(\text{small})}{\bar{Q}\,(\text{big}) - \bar{Q}\,(\text{small})}$$

where $\bar{S}^e\,(\text{big})$ and $\bar{Q}\,(\text{big})$ (resp. $\bar{S}^e\,(\text{small})$ and $\bar{Q}\,(\text{small})$) are the average of the effective spread and the average of order size for big trades (resp. small trades). This measure is however difficult to implement, because we need to split all the trades into big and small orders. This is why this measure is very sensitive to the choice of the size threshold. A more interesting and popular market-impact measure is the Amihud measure defined by:

$$\mathbf{ILLIQ} = \frac{1}{n_t}\sum_t \frac{|R_{t,t+1}|}{\mathbf{V}_t}$$

where $R_{t,t+1}$ is the daily return, $\mathbf{V}_t$ is the daily trading volume and n_t is the number of days used to calculate the sum. Amihud (2002) uses this ratio to measure the relationship between the (absolute) price slope and the order flow.

This liquidity ratio is one of the most popular academic measures with the implicit spread of Roll (1984), who assumes that the fundamental price $P_t^\star$ follows a random walk:

$$P_t^\star = P_{t-1}^\star + \varepsilon_t$$

whereas the observed price depends on the trade direction:

$$P_t = P_t^\star + s_t \cdot \left(\frac{\mathbf{S}}{2}\right)$$

where $\mathbf{S}$ is the bid-ask spread and:

$$s_t = \begin{cases} +1 & \text{if the trade is a buy} \\ -1 & \text{if the trade is a sell} \end{cases}$$

We deduce that:

$$\Delta P_t = \Delta s_t \cdot \left(\frac{\mathbf{S}}{2}\right) + \varepsilon_t$$

By assuming that buy and sell orders have the same probability, Roll shows that the first-order auto-covariance of price changes is equal to:

$$\begin{aligned} \operatorname{cov}(\Delta P_t, \Delta P_{t-1}) &= \operatorname{cov}(\Delta s_t, \Delta s_{t-1}) \cdot \left(\frac{\mathbf{S}}{2}\right)^2 \\ &= -\left(\frac{\mathbf{S}}{2}\right)^2 \end{aligned}$$

We can therefore deduce the implied spread by the following expression:

$$\tilde{\mathbf{S}} = 2\sqrt{-\operatorname{cov}(\Delta P_t, \Delta P_{t-1})}$$

To estimate $\tilde{\mathbf{S}}$, we can use the empirical covariance of price changes or the Gibbs sampler proposed by Hasbrouck (2009).

Remark 65 *The seminal paper of Roll has been extended in several directions: asymmetric information, serial dependence of the trades, etc*[7].

[7] See Huang and Stoll (1996, 1997) for a survey.

6.1.3 The liquidity-adjusted CAPM

The liquidity-adjusted CAPM is an extension of the capital asset pricing model of Sharpe (1964). This model, which has been proposed by Acharya and Pedersen (2005), analyzes the relationship between liquidity and asset prices. It goes beyond the traditional approach, which consists in considering a static liquidity premium that affects asset returns. In this approach, the level of liquidity[8] is the most important factor to take into account. For instance, when we consider real assets, we generally consider that their returns must incorporate a risk premium. A typical example concerns private equity. However, most of the time, when we think about the liquidity premium, we think that it is related to the level of illiquidity of the asset, and not to the dynamics of the liquidity. However, the issue is more complex:

> "*[...] there is also broad belief among users of financial liquidity – traders, investors and central bankers – that the principal challenge is not the average level of financial liquidity... but its variability and uncertainty*" (Persaud, 2003).

The liquidity-adjusted CAPM (or L-CAPM) considers a framework where both the level of liquidity and the variability have an impact on asset prices.

We note $R_{i,t}$ and $L_{i,t}$ the gross return and the relative (stochastic) illiquidity cost of Asset i. At the equilibrium, Acharya and Pedersen (2005) show that "*the CAPM in the imagined frictionless economy translates into a CAPM in net returns for the original economy with illiquidity costs*":

$$\mathbb{E}\left[R_{i,t} - L_{i,t}\right] - r = \tilde{\beta}_i \cdot \left(\mathbb{E}\left[R_{m,t} - L_{m,t}\right] - r\right) \tag{6.1}$$

where r is the return of the risk-free asset, $R_{m,t}$ and $L_{m,t}$ are the gross return and the illiquidity cost of the market portfolio, and $\tilde{\beta}_i$ is the liquidity-adjusted beta of Asset i:

$$\tilde{\beta}_i = \frac{\operatorname{cov}\left(R_{i,t} - L_{i,t}, R_{m,t} - L_{m,t}\right)}{\operatorname{var}\left(R_{m,t} - L_{m,t}\right)}$$

Equation (6.1) shows that the net risk premium of an asset, that is the risk premium minus the liquidity cost, is equal to its beta times the net market risk premium. However, the beta in this formula is different than the formula in the CAPM, because the beta depends on the liquidity of the asset and the liquidity of the market. Indeed, the liquidity-adjusted beta can be decomposed into four betas[9]:

$$\tilde{\beta}_i = \beta_i + \beta\left(L_{i,t}, L_{m,t}\right) - \beta\left(R_{i,t}, L_{m,t}\right) - \beta\left(L_{i,t}, R_{m,t}\right)$$

where $\beta_i = \beta\left(R_{i,t} R_{m,t}\right)$ is the standard market beta, $\beta\left(L_{i,t}, L_{m,t}\right)$ is the beta associated to the commonality in liquidity with the market liquidity, $\beta\left(R_{i,t}, L_{m,t}\right)$ is the beta associated to the return sensitivity to market liquidity and $\beta\left(L_{i,t}, R_{m,t}\right)$ is the beta associated to the liquidity sensitivity to market returns. Therefore, some assets have a low (or high) beta with

[8] Or more precisely the level of illiquidity.

[9] We have:

$$\begin{aligned}
\operatorname{cov}\left(R_{i,t} - L_{i,t}, R_{m,t} - L_{m,t}\right) &= \mathbb{E}\left[\left(R_{i,t} - L_{i,t}\right) \cdot \left(R_{m,t} - L_{m,t}\right)\right] - \\
&\quad \mathbb{E}\left[R_{i,t} - L_{i,t}\right] \cdot \mathbb{E}\left[R_{m,t} - L_{m,t}\right] \\
&= \mathbb{E}\left[R_{i,t} R_{m,t} + L_{i,t} L_{m,t} - R_{i,t} L_{m,t} - L_{i,t} R_{m,t}\right] - \\
&\quad \left(\mathbb{E}\left[R_{i,t}\right] - \mathbb{E}\left[L_{i,t}\right]\right) \cdot \left(\mathbb{E}\left[R_{m,t}\right] - \mathbb{E}\left[L_{m,t}\right]\right) \\
&= \operatorname{cov}\left(R_{i,t}, R_{m,t}\right) + \operatorname{cov}\left(L_{i,t}, L_{m,t}\right) - \\
&\quad \operatorname{cov}\left(R_{i,t}, L_{m,t}\right) - \operatorname{cov}\left(L_{i,t}, R_{m,t}\right)
\end{aligned}$$

respect to the market portfolio, not because their returns do not covary (or highly covary) with the returns of the market portfolio, not because of their liquidity level, but because of the time-variability of the liquidity and the impact of the liquidity on market returns.

Acharya and Pedersen (2005) propose to rewrite Equation (6.1) into the CAPM equation:

$$\mathbb{E}\left[R_{i,t}\right] - r = \alpha_i + \beta_i \cdot \left(\mathbb{E}\left[R_{m,t}\right] - r\right) \tag{6.2}$$

where α_i is a function of the relative liquidity of Asset i with respect to the market portfolio and the liquidity betas:

$$\begin{aligned}\alpha_i &= \left(\mathbb{E}\left[L_{i,t}\right] - \tilde{\beta}_i \cdot \mathbb{E}\left[L_{m,t}\right]\right) + \\ &\quad \left(\beta\left(L_{i,t}, L_{m,t}\right) - \beta\left(R_{i,t}, L_{m,t}\right) - \beta\left(L_{i,t}, R_{m,t}\right)\right) \cdot \pi_m\end{aligned}$$

where $\pi_m = \mathbb{E}\left[R_M\right] - r$. It follows that the asset return can be written as:

$$R_{i,t} = \alpha_{i,t} + \beta_i \cdot R_{m,t} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t} \sim \mathcal{N}\left(0, \tilde{\sigma}_i^2\right)$ and $\tilde{\sigma}_i$ is the specific volatility of the asset. We retrieve the classical one-factor model, but with a time-varying alpha component. Contrary to the common wisdom, the alpha of the asset is not only equal to the illiquidity level. Indeed, Acharya and Pedersen (2005) show that the level of liquidity explains 75% of the alpha, whereas 25% of the alpha is explained by the liquidity sensitivity of the asset to market returns.

The previous model can also be written as follows:

$$R_{i,t} - r = \mu\left(L_{i,t}\right) + \left(\mathcal{R}\left(L_{i,t}\right) + \beta_i\right) \cdot \left(R_{m,t} - r\right) + \varepsilon_{i,t}$$

where $\mu\left(L_{i,t}\right)$ is the relative liquidity level:

$$\mu\left(L_{i,t}\right) = \mathbb{E}\left[L_{i,t}\right] - \tilde{\beta}_i \cdot \mathbb{E}\left[L_{m,t}\right]$$

and $\mathcal{R}\left(L_{i,t}\right)$ is the aggregated liquidity risk:

$$\mathcal{R}\left(L_{i,t}\right) = \beta\left(L_{i,t}, L_{m,t}\right) - \beta\left(R_{i,t}, L_{m,t}\right) - \beta\left(L_{i,t}, R_{m,t}\right)$$

$\mathcal{R}\left(L_{i,t}\right)$ is composed of three liquidity covariance risks, and Acharya and Pedersen (2005) interpret each of them as follows:

1. the first covariance risk $\beta\left(L_{i,t}, L_{m,t}\right)$ shows that an asset that becomes illiquid when the market becomes illiquid should have a higher risk premium; this risk is related to the substitution effects we observe when the market becomes illiquid;

2. the second covariance risk $\beta\left(R_{i,t}, L_{m,t}\right)$ indicates that assets that perform well in times of market illiquidity should have a lower risk premium because of the solvency constraints faced by investors;

3. the third covariance risk $\beta\left(L_{i,t}, R_{m,t}\right)$ means that investors accept a lower risk premium on assets that are liquid in a bear market, because they have the property to be sold in illiquid markets.

It is obvious that these three liquidity risks are correlated and Acharya and Pedersen (2005) estimate the following correlation figures:

	$\beta\left(L_{i,t}, L_{m,t}\right)$	$\beta\left(R_{i,t}, L_{m,t}\right)$	$\beta\left(L_{i,t}, R_{m,t}\right)$
$\beta\left(L_{i,t}, L_{m,t}\right)$	100%		
$\beta\left(R_{i,t}, L_{m,t}\right)$	−57%	100%	
$\beta\left(L_{i,t}, R_{m,t}\right)$	−94%	73%	100%

The Acharya-Pedersen model illustrates perfectly well some stylized facts concerning some asset classes such as corporate bonds, small cap stocks or private equities. In particular, it shows that liquidity has an impact on the risk premium of securities, but it has also an impact on the price dynamics because of the liquidity risk or uncertainty. This implies that liquidity has an impact on the systematic return component. We will see later how this interconnectedness between asset liquidity and market liquidity is important when regulators would like to manage the systemic risk of the financial system.

6.2 Funding liquidity

According to Nikolaou (2009), we must distinguish three liquidity types: market liquidity, funding liquidity and central bank liquidity. Funding liquidity is the ability of banks to meet their liabilities, whereas central bank liquidity is the ability of central banks to supply the liquidity needed by the financial system. As noticed by Nikolaou (2009), central bank liquidity is not an issue as long as there is a demand for the domestic currency. In this section, we focus on funding liquidity, which is in fact the main issue of liquidity risk. Indeed, the 2008 Global Financial Crisis has demonstrated that it is the problematic linkage layer even when central bank liquidity is infinite.

6.2.1 Asset liability mismatch

Whereas market liquidity is asset specific, funding liquidity is agent specific (Brunnermeier and Pedersen, 2009). For instance, we can measure the market liquidity of a stock, a bond or a futures contract. In a similar way, we can measure the funding liquidity of a bank, an insurer or a corporate firm. We can extend these measures to a portfolio of securities or a group of entities. Therefore, we can define the global market liquidity of an asset class, for example the liquidity of US large cap stocks or the liquidity of EUR-denominated convertible bonds. We can also define the global funding liquidity of a financial system, for example the liquidity of Italian banks or the liquidity of the Japanese financial system. At first sight, funding liquidity seems to be the mirror image of market liquidity when we consider banks instead of securities. In fact, it is a false view for several reasons. The first reason concerns the distinction between funding liquidity and funding liquidity risk:

> "*We define funding liquidity as the ability to settle obligations with immediacy. Consequently, a bank is illiquid if it is unable to settle obligations. Legally, a bank is then in default. Given this definition we define funding liquidity risk as the possibility that over a specific horizon the bank will become unable to settle obligations with immediacy*" (Drehmann and Nikolaou, 2013, page 2174).

In the previous section, we have seen several measures of the market liquidity, and these measures can be used to calculate the market liquidity risk, that is the market liquidity at some time horizon. Funding liquidity is more a binary concept and is related to credit risk. Therefore, funding liquidity risk may be viewed as the probability that the bank will face a funding liquidity problem in the future. The difficulty is then to make the distinction between funding liquidity and credit risks, since their definitions are very close. Said differently, the issue is to measure the probability of funding liquidity risk and not the probability of default risk (Drehmann and Nikolaou, 2013).

Drehmann and Nikolaou (2013) considers a stock-flow measure. Let D_i be the indicator function, which takes the value 0 if the bank faces no funding liquidity risk, or 1 otherwise. We have the following relationship:

$$D_i = 0 \Leftrightarrow O_t \leq I_t + M_t \tag{6.3}$$

where O_t are the outflows, I_t are the inflows and M_t is the stock of money at time t. The general components of O_t and I_t are:

$$O_t = L_{\text{new},t} + A_{\text{due},t} + IP_t$$

and:

$$I_t = L_{\text{due},t} + A_{\text{new},t} + IR_t$$

where $L_{\text{new},t}$ and $L_{\text{due},t}$ are liabilities which are newly issued or due, $A_{\text{new},t}$ and $A_{\text{due},t}$ are assets which are newly issued or due, and IP_t and IR_t are interest payments paid or received by the bank. These outflows/inflows concerns 5 categories: (DP) depositors, (IB) interbank, (AM) asset market, (OB) off-balance sheet items and (CB) central banks. The authors define then the net liquidity demand $\text{NLD}_t = O_t - I_t - M_t$ as the net amount of central bank money the bank needs to remain liquid and show that this variable must satisfy the following inequality:

$$\text{NLD}_t \leq P_t^{\text{DP}} L_{\text{new},t}^{\text{DP}} + P_t^{\text{IB}} L_{\text{new},t}^{\text{IB}} + P_t^{\text{AM}} A_{\text{sold},t} + P_t^{\text{CB}} CB_{\text{new},t} \tag{6.4}$$

where P_t^k is the price of the category k, $L_{\text{new},t}^{\text{DP}}$ and $L_{\text{new},t}^{\text{IB}}$ correspond to the new liabilities from depositors and the interbank market, $A_{\text{sold},t}$ is the amount of selling assets and $CB_{\text{new},t}$ is the new central bank money. Equation (6.4) gives the different components that the bank can access when $O_t > I_t + M_t$.

This simple model shows that three dimensions are important when measuring the funding liquidity risk. First, the time horizon is a key parameter. Second, the projection of assets and liabilities is not an easy task. This is particularly true if the bank is highly leveraged or operates a larger maturity transformation between assets and liabilities. Third, we have to take into account spillover effects. Indeed, the bank does not know the reaction function of the other financial agents if it faces asset/liability liquidity mismatch[10].

6.2.2 Relationship between market and funding liquidity risks

Brunnermeier and Pedersen (2009) highlights the interconnectedness of market liquidity and funding liquidity:

> "*Traders provide market liquidity, and their ability to do so depends on their availability of funding. Conversely, traders' funding, i.e., their capital and margin requirements, depends on the assets' market liquidity. We show that, under certain conditions, margins are destabilizing and market liquidity and funding liquidity are mutually reinforcing, leading to liquidity spirals*" (Brunnermeier and Pedersen, 2009, page 2201).

The model allows the authors to show that market liquidity can suddenly dry up. This analysis has been extended by Nikolaou (2009), who includes the central bank liquidity for analyzing the liquidity linkages. In normal times, we observe a virtuous liquidity circle that reinforces the financial system stability (Figure 6.3). Nevertheless, the liquidity linkages can

[10] This problem is discussed in Chapter 8, which is dedicated to the systemic risk.

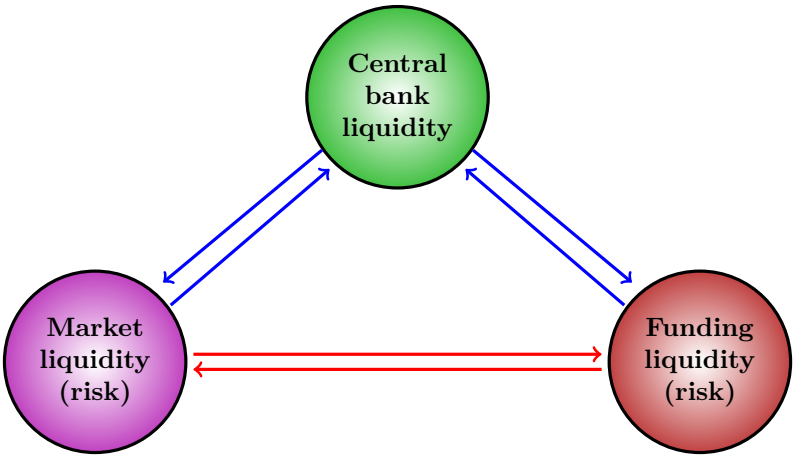


FIGURE 6.3: The liquidity nodes of the financial system

Source: Nikolaou (2009).

be broken in bad times. It is commonly accepted that funding liquidity is the central node, because central banks have no interest to break the linkages and market illiquidity can only be temporary. Indeed, market illiquidity is generally associated to periods of selling (or bear) markets. However, there is an intrinsic imbalance between supply and demand in financial markets. Investors, financial institutions, households and corporate firms have a common objective to buy financial assets, not to sell them, in order to finance retirement pensions, future consumptions and economic growth. Therefore, the weak link is the funding liquidity, because funding liquidity risks can easily create systemic risks.

The 2008 Global Financial Crisis is the typical example of a systemic risk crisis, which is mainly due to the liquidity risk, especially the funding liquidity risk. We can represent the different sequences of the crisis with the scheme given in Figure 6.4. The starting point was the subprime debt crisis that has impacted banks. Therefore, it was first a credit risk crisis. However, its strength weakened the banks, leading to a reduction of the funding liquidity. At the same time, the banking system dramatically reduces the funding to corporate firms, asset managers and hedge funds. In order to obtain cash, investors sold liquid assets, and more especially stocks. The drop of stock prices affected banks because the value of collateral portfolios has decreased. It followed a feedback loop between credit risk, liquidity risk, market risk and collateral risk.

Remark 66 *Many people compare the GFC to the dot-com crisis, certainly because the performance of the stock market is similar. Indeed, during the dot-com crisis, the S&P 500 index experienced a maximum drawdown about* 49% *between March 2000 and March 2003, whereas it was equal to* 56% *during the subprime crisis. However, the behavior of stocks was different during these two periods. During the internet crisis,* 55% *of stocks posted a negative performance, while* 45% *of stocks posted a positive performance. The dot-com crisis is then a crisis of valuation. During the GFC,* 95% *of stocks posted a negative performance. In fact, the 2008 crisis of the stock market is mainly a liquidity crisis. This explains that almost all stocks had a negative return. This example perfectly illustrates the interconnectedness of funding liquidity and market risks.*

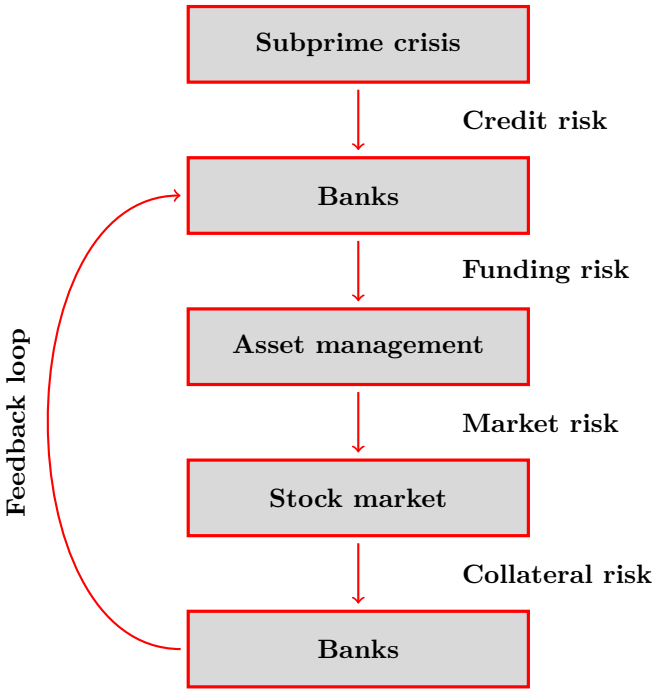


FIGURE 6.4: Spillover effects during the 2008 global financial crisis

6.3 Regulation of the liquidity risk

In Basel III, the liquidity risk is managed using two layers. First, the market liquidity issue is included in the market risk framework by considering five liquidity horizons: 10, 20, 40, 60 and 120 days. For example, the liquidity horizon for large cap equity prices is set to 10 days whereas the liquidity horizon for credit spread volatilities is set to 120 days. Therefore, there is a differentiation in terms of asset classes and instruments. Second, the Basel Committee has developed two minimum standards for funding liquidity: the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR). The objective of the LCR is to promote short-term resilience of the bank's liquidity risk profile, whereas the objective of the NSFR is to promote resilience over a longer time horizon. Moreover, these tools are completed by the leverage ratio. Indeed, although its first objective is not to manage the funding liquidity risk, the leverage risk is an important component of the funding liquidity risk[11].

6.3.1 Liquidity coverage ratio

6.3.1.1 Definition

The liquidity coverage ratio is defined as:

$$\text{LCR} = \frac{\text{HQLA}}{\text{Total net cash outflows}} \geq 100\%$$

[11] See Section 8.1.2 on page 456.

TABLE 6.4: Stock of HQLA

Level	Description	Haircut
Level 1 assets		
	Coins and bank notes Sovereign, central bank, PSE, and MDB assets qualifying for 0% risk weighting Central bank reserves Domestic sovereign or central bank debt for non-0% risk weighting	0%
Level 2 assets (maximum of 40% of HQLA)		
Level 2A assets		
	Sovereign, central bank, PSE and MDB assets qualifying for 20% risk weighting Corporate debt securities rated AA− or higher Covered bonds rated AA− or higher	15%
Level 2B assets (maximum of 15% of HQLA)		
	RMBS rated AA or higher	25%
	Corporate debt securities rated between A+ and BBB−	50%
	Common equity shares	50%

Source: BCBS (2013a).

where the numerator is the stock of high quality liquid assets (HQLA) in stressed conditions, and the denominator is the total net cash outflows over the next 30 calendar days. The underlying idea of the LCR is that the bank has sufficient liquid assets to meet its liquidity needs for the next month.

An asset is considered to be a high quality liquid asset if it can be easily converted into cash. Therefore, the concept of HQLA is related to asset quality and asset liquidity. Here are the comprehensive list of characteristics used by the Basel Committee for defining HQLA:

- fundamental characteristics (low risk, ease and certainty of valuation, low correlation with risky assets, listed on a developed and recognized exchange);
- market-related characteristics (active and sizable market, low volatility, flight to quality).

BCBS (2013a) divides the stock of HQLA into two buckets (see Table 6.4). The first bucket (called level 1 assets) has a 0% haircut. It consists of coins and banknotes, central bank reserves, and qualifying marketable securities from sovereigns, central banks, public sector entities (PSE), and multilateral development banks (MDB), whose risk weight is 0% under the Basel II SA framework for credit risk. The level 1 assets also include sovereign or central bank debt securities issued in the domestic currency of the bank's home country. In the second bucket (also called level 2 assets), assets have a haircut higher than 0%. For instance, a 15% haircut is applied to sovereign, central bank, PSE and MBD assets that have a 20% risk weight under the Basel II SA framework for credit risk. A 15% haircut is also valid for corporate debt securities that are rated at least AA−. Three other types of assets are included in the second bucket (level 2B assets). They concern RMBS rated AA or higher, corporate debt securities with a rating between A+ and BBB−, and common equity shares that belong to a major stock index. Moreover, the HQLA portfolio must be well

diversified in order to avoid concentration (except for sovereign debt of the bank's home country).

We notice that level 2 assets are subject to two caps. Let x_{HQLA}, x_1 and x_2 be the value of HQLA, level 1 assets and level 2 assets. We have:

$$\begin{aligned} x_{\text{HQLA}} &= x_1 + x_2 \\ \text{s.t.} &\begin{cases} x_2 = x_{2A} + x_{2B} \\ x_{2A} \leq 0.40 \cdot x_{\text{HQLA}} \\ x_{2B} \leq 0.15 \cdot x_{\text{HQLA}} \end{cases} \end{aligned}$$

We deduce that one trivial solution is:

$$\begin{cases} x^\star_{\text{HQLA}} = \min\left(\frac{5}{3}x_1, x_1 + x_2\right) \\ x^\star_1 = x_1 \\ x^\star_2 = x^\star_{\text{HQLA}} - x^\star_1 \\ x^\star_{2A} = \min\left(x^\star_2, x_{2A}\right) \\ x^\star_{2B} = x^\star_2 - x^\star_{2A} \end{cases} \tag{6.5}$$

Example 63 *We consider the following assets: (1) coins and bank notes = \$200 mn, (2) central bank reserves = \$100 mn, (3) 20% risk-weighted sovereign debt securities = \$200 mn, (4) AA corporate debt securities = \$300 mn, (5) qualifying RMBS = \$200 mn and (6) BB+ corporate debt securities = \$500 mn.*

Results are given in the table below. We notice that the gross value of assets is equal to \$1.5 bn. However, level 2 assets represent 80% of this amount, implying that the 40% cap is exceeded. Therefore, we have to perform the correction given by Equation (6.5). Finally, the stock of HQLA is equal to \$500 mn.

	Assets	Gross Value	Haircut	Net Value	Capped Value
Level 1 assets	(1) + (2)	300	0%	300	300
Level 2 assets		1 200		825	200
2A	(3) + (4)	500	15%	425	200
2B	(5) + (6)	700		400	0
	(5)	200	25%	150	0
	(6)	500	50%	250	0
Total		1 500		1 125	500

Remark 67 *The previous example shows that the bank may use secured funding transactions (repos) to circumvent the caps on level 2 assets. This is why the LCR requires adjusting the amount of HQLA by taking into account the unwind of repos maturing within 30 calendar days that involve the exchange of HQLA*[12].

The value of total net cash outflows is defined as follows:

$$\begin{aligned} \text{Total net cash outflows} = \text{Total expected cash outflows} - \\ \min\left(\begin{array}{c}\text{Total expected cash inflows,} \\ 75\% \text{ of total expected cash outflows}\end{array}\right) \end{aligned}$$

[12] See §48 and Annex 1 in BCBS (2013a).

TABLE 6.5: Cash outflows of the LCR

Liabilities Description	Rate
Retail deposits	
Demand and term deposits (less than 30 days)	
Stable deposits covered by deposit insurance	3%
Stable deposits	5%
Less stable deposits	10%
Term deposits (with residual maturity greater than 30 days)	0%
Unsecured wholesale funding	
Demand and term deposits (less than 30 days) provided by small business customers	
Stable deposits	5%
Less stable deposits	10%
Deposits generated by clearing, custody and cash management	25%
Portion covered by deposit insurance	5%
Cooperative banks in an institutional network	25%
Corporates, sovereigns, central banks, PSEs and MDBs	40%
Portion covered by deposit insurance	20%
Secured funding transactions	
With a central bank counterparty	0%
Backed by level 1 assets	0%
Backed by level 2A assets	15%
Backed by non-level 1 or non-level 2A assets with domestic sovereigns, PSEs or MDBs as a counterparty	25%
Backed by level 2B RMBS assets	25%
Backed by other level 2B assets	50%
All other secured funding transactions	100%
Additional requirements	
Margin/collateral calls	$\geq 20\%$
ABCP, SIVs, conduits, SPVs, etc.	100%
Net derivative cash outflows	100%
Other credit/liquidity facilities	$\geq 5\%$

Source: BCBS (2013a).

Therefore, it is the difference between cash outflows and cash inflows, but with a floor of 25% of cash outflows. Cash outflows/inflows are estimated by applying a run-off/flow-in rate to each category of liabilities/receivables.

In Table 6.5, we report the main categories of cash outflows and their corresponding run-off rates. These outflow rates are calibrated according to expected stability or '*stickiness*'. Run-off rates range from 3% to 100%, depending on the nature of the funding. The less stable funding is perceived to be, the higher the outflow rate. For example, a 3% rate is assigned to stable retail deposits that benefit of deposit insurance (protection offered by government or public guarantee schemes). On the contrary, the rate of deposits from corporates is equal to 40%.

The categories of cash inflows are given in Table 6.6. Maturing secured lending transactions (reverse repos and securities borrowing) have inflow rates from 0% to 100%. Amounts receivable from retail and corporate counterparties have an inflow rate of 50%, whereas amounts receivable from financial institutions have an inflow rate of 100%.

TABLE 6.6: Cash inflows of the LCR

Receivables Description	Rate
Maturing secured lending transactions	
Backed by level 1 assets	0%
Backed by level 2A assets	15%
Backed by level 2B RMBS assets	25%
Backed by other level 2B assets	50%
Backed by non-HQLAs	100%
Other cash inflows	
Credit/liquidity facilities provided to the bank	0%
Inflows to be received from retail counterparties	50%
Inflows to be received from non-financial wholesale counterparties	50%
Inflows to be received from financial institutions and central banks	100%
Net derivative receivables	100%

Source: BCBS (2013a).

Example 64 *The bank has* \$500 *mn of HQLA. Its main liabilities are: (1) retail stable deposit* = \$17.8 *bn (*\$15 *bn have a government guarantee), (2) retail term deposit (with a maturity of* 6 *months)* = \$5 *bn, (3) stable deposit provided by small business customers* = \$1 *bn, and (4) deposit of corporates* = \$200 *mn. In the next thirty days, the bank also expects to receive* \$100 *mn of loan repayments, and* \$10 *mn due to a maturing derivative.*

We first calculate the expected cash outflows for the next thirty days:

$$\begin{aligned} \text{Cash outflows} &= 3\% \times 15\,000 + 5\% \times 2\,800 + 0\% \times 5\,000 + \\ &\quad 5\% \times 1\,000 + 40\% \times 200 \\ &= \$720 \text{ mn} \end{aligned}$$

We then estimate the cash inflows expected by the bank for the next month:

$$\text{Cash inflows} = 50\% \times 100 + 100\% \times 10 = \$60 \text{ mn}$$

Finally, we deduce that the liquidity coverage ratio of the bank is equal to:

$$\text{LCR} = \frac{500}{720 - 60} = 75.76\%$$

6.3.1.2 Monitoring tools

In addition to the LCR, the Basel Committee has defined five monitoring tools in order to analyze the bank's liquidity risk management:

1. Contractual maturity mismatch
2. Concentration of funding
3. Available unencumbered assets
4. LCR by significant currency
5. Market-related monitoring tools

The contractual maturity mismatch defines "*the gaps between the contractual inflows and outflows of liquidity for defined time bands*". The Basel Committee suggests the following time buckets: overnight, 7 days, 14 days, 1, 2, 3, 6 and 9 months, 1, 3, 5 and 5+ years. The goal of the second metric is to identify the main sources of liquidity problems. Thus, the concentration of funding groups three types of information:

1. funding liabilities sourced from each significant counterparty as a % of total liabilities;
2. funding liabilities sourced from each significant product/instrument as a % of total liabilities;
3. list of asset and liability amounts by significant currency.

The third metric concerns available unencumbered assets that are marketable as collateral in secondary markets, and available unencumbered assets that are eligible for central banks' standing facilities. The bank must then report the amount, type and location of available unencumbered assets that can be used as collateral assets. For the fourth metric, the bank must calculate a '*foreign currency LCR*' for all the significant currencies. A currency is said to be significant if the liabilities denominated in that currency is larger than 5% of the total liabilities. Supervisors are in charge of producing the last metric, which corresponds to market data that can serve as warning indicators of liquidity risks (CDS spread of the bank, trading volume in equity markets, bid/ask spreads of sovereign bonds, etc.). These indicators can be specific to the bank, the financial sector, or a financial market.

6.3.2 Net stable funding ratio

While the LCR measures the funding liquidity risk for the next month, NSFR is designed in order to promote resilience of the bank's liquidity profile for the next year. Like the LCR, NSFR is based on the asset liability approach, but it is more comprehensive than the LCR because of the long-term horizon. In some sense, it is closer to the framework that has been proposed by Drehmann and Nikolaou (2013).

6.3.2.1 Definition

It is defined as the amount of available stable funding (ASF) relative to the amount of required stable funding (RSF):

$$\text{NSFR} = \frac{\text{Available amount of stable funding}}{\text{Required amount of stable funding}} \geq 100\%$$

The available amount of stable funding (ASF) corresponds to the regulatory capital plus some other liabilities, whereas the required amount of stable funding (RSF) is the sum of weighted assets and off-balance sheet exposures. We have:

$$\text{ASF} = \sum_i f_i^{\text{ASF}} \cdot L_i$$

and:

$$\text{RSF} = \sum_j f_j^{\text{RSF}} \cdot A_j$$

where f_i^{ASF} is the ASF factor for liability i, L_i is the amount of liability i, f_j^{RSF} is the RSF factor for asset j, and A_j is the amount of asset j.

6.3.2.2 ASF and RSF factors

The ASF factor can take 5 values: 100%, 95%, 90%, 50% and 0%. Here are the main components of the available amount of stable funding:

- Liabilities receiving a 100% ASF factor
  This concerns (1) regulatory capital (excluding tier 2 instruments with residual maturity of less than one year) and (2) other capital instruments with effective residual maturity of one year or more.
- Liabilities receiving a 95% ASF factor
  This includes (3) stable non-maturity/term deposits of retail and small business customers (with residual maturity of less than one year).
- Liabilities receiving a 90% ASF factor
  This corresponds to (4) less stable non-maturity/term deposits of retail and small business customers (with residual maturity of less than one year).
- Liabilities receiving a 50% ASF factor
  In this category, we find (5) funding provided by sovereigns, corporates, MDBs and PSEs (with residual maturity of less than one year), (6) funding provided by central banks and financial institutions (with residual maturity between 6 months and one year) and (7) operational deposits.
- Liabilities receiving a 0% ASF factor
  This category corresponds to (8) all the other liabilities.

The RSF factor takes values between 0% and 100%. The main components of the required amount of stable funding are the following:

- Assets receiving a 0% RSF
  This concerns (1) coins and banknotes and (2) all central bank reserves and all claims on central banks with residual maturities of less than six months.
- Assets receiving a 5% RSF
  In this category, we find (3) other unencumbered level 1 assets.
- Assets receiving a 10% RSF
  This includes (4) unencumbered secured loans to financial institutions with residual maturities of less than six months.
- Assets receiving a 15% RSF
  This category if composed of (5) all other unencumbered loans to financial institutions with residual maturities of less than six months and (6) unencumbered level 2A assets.
- Assets receiving a 50% RSF
  This corresponds to (7) unencumbered level 2B assets and (8) all other assets with residual maturity of less than one year.
- Assets receiving a 65% RSF
  This concerns (9) unencumbered residential mortgages and loans (excluding loans to financial institutions) with a residual maturity of one year or more and with a risk weight of less than or equal to 35% under the Standardized Approach.

- Assets receiving a 85% RSF
  In this category, we have (10) cash, securities or other assets posted as initial margin for derivative contracts and provided to contribute to the default fund of a CCP, (11) other unencumbered performing loans (excluding loans to financial institutions) with a residual maturity of one year or more and with a risk weight of less greater than 35% under the Standardized Approach, (12) exchange-traded equities and (13) physical traded commodities, including gold.

- Assets receiving a 100% RSF
  This category is defined by (14) all assets that are encumbered for a period of one year or more and (15) all other assets (non-performing loans, loans to financial institutions with a residual maturity of one year or more, non-exchange-traded equities, etc.).

Example 65 *We assume that the bank has the following simplified balance sheet:*

Assets		*Amount*	*Liabilities*		*Amount*
Loans	*Residential*	150	*Deposits*	*Stable*	100
	Corporate	60		*Less stable*	150
Level	*1A*	70	*Short-term*	*borrowing*	50
	2B	40	*Capital*		20

We deduce that:

$$\text{ASF} = 95\% \times 100 + 90\% \times 150 + 50\% \times 50 + 100\% \times 20 = 275$$

and:

$$\text{RSF} = 85\% \times 150 + 85\% \times 60 + 5\% \times 70 + 50\% \times 40 = 202$$

The NSFR is then equal to:

$$\text{NSFR} = \frac{275}{202} = 136\%$$

6.3.3 Leverage ratio

As said previously, the leverage ratio completes the framework of market and liquidity risks. It is defined as the capital measure divided by the exposure measure. Since January 2018, this ratio must be below 3%. The capital measure corresponds to the tier 1 capital, while the exposure measure is composed of four main exposures: on-balance sheet exposures, derivative exposures, securities financing transaction (SFT) exposures and off-balance sheet items. The big issue is the definition of derivative exposures, because we can adopt either a notional or a mark-to-market approach. Finally, the Basel Committee has decided to define them as the sum of replacement cost and potential future exposure, meaning that derivative exposures correspond to a CCR exposure measure.

Remark 68 *We could have discussed the leverage risk ratio in other chapters, in particular when considering systemic risk and shadow banking system. In fact, liquidity, leverage and systemic risks are so connected that it is difficult to distinguish them.*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 7

Asset Liability Management Risk

Asset liability management (ALM) corresponds to the processes that address the mismatch risk between assets and liabilities. These methods concern financial institutions, which are mainly defined by a balance sheet. For example, this is the case of pension funds and insurance companies. In this chapter, we focus on ALM risks in banks, and more precisely ALM risks of the banking book. Previously, we have already seen some risks that impact the banking book such as credit or operational risk. In what follows, we consider the four specific ALM risks: liquidity risk[1], interest rate risk, option risk and currency risk.

Generally, ALM risks are little taught in university faculties because they are less known by academics. In fact, asset liability management is a mix of actuarial science, accounting and statistical modeling, and seems at first sight less mathematical than risk management. Another difference is that the ALM function is generally within the finance department and not within the risk management department. This is because ALM implies to take decisions that are not purely related to risk management considerations, but also concerns commercial choices and business models.

7.1 General principles of the banking book risk management

Before presenting the tools to manage the ALM risks, we define the outlines of the asset and liability management. In particular, we show why ALM risks are so specific if we compare them to market and credit risks. In fact, asset and liability management has two components. The first component is well-identified and corresponds to the risk measurement of ALM operations. The second component is much more vague, because it concerns both risk management and business development. Indeed, banking business is mainly a financial intermediation business, since banks typically tend to borrow short term and lend long term. The mismatch between assets and liabilities is then inherent to banking activities. Similarly, the balance sheet of a bank and its income statement are highly related, implying that future income may be explained by the current balance sheet. The debate on whether the ALM department is a profit center summarizes this duality between risk and business management.

[1] Liquidity risk was the subject of the previous chapter. However, we have discussed this topic from a risk management point of view by focusing on the regulatory ratios (LCR and NSFR). In this chapter, we tackle the issue of liquidity risk from an ALM perspective.

7.1.1 Definition

7.1.1.1 Balance sheet and income statement

The ALM core function is to measure the asset liability mismatch of the balance sheet of the bank. In Table 7.1, we report the 2018 balance sheet of FDIC-insured commercial banks and savings institutions as provided by FDIC (2019). It concerns 5 406 financial institutions in the US. We notice that the total assets and liabilities are equal to \$17.9 tn. The most important items are loans and leases, investment securities and cash & due from depository institutions on the asset side, deposits and equity capital on the liability

TABLE 7.1: Assets and liabilities of FDIC-insured commercial banks and savings institutions (Amounts in \$ bn)

Total Assets	**17 943**
Loans secured by real estate	**4 888**
1-4 Family residential mortgages	2 119
Nonfarm nonresidential	1 446
Construction and development	350
Home equity lines	376
Multifamily residential real estate	430
Farmland	105
Real estate loans in foreign offices	62
Commercial & industrial loans	**2 165**
Loans to individuals	**1 743**
Credit cards	903
Other loans to individuals	839
Auto loans	455
Farm loans	82
Loans to depository institutions	84
Loans to foreign gov. & official inst.	11
Obligations of states in the U.S.	188
Other loans	862
Lease financing receivables	133
Gross total loans and leases	10 155
Less: Unearned income	2
Total loans and leases	10 152
Less: Reserve for losses	125
Net loans and leases	**10 028**
Securities	**3 723**
Available for sale (fair value)	2 590
Held to maturity (amortized cost)	1 129
U.S. Treasury securities	549
Mortgage-backed securities	2 187
State and municipal securities	330
Equity securities	3
Cash & due from depos. instit.	**1 694**
Fed. funds sold and reverse repos	622
Bank premises and fixed assets	130
Other real estate owned	7
Trading account assets	572
Intangible assets	399
Goodwill	334
Other Assets	769

Total liabilities and capital	**17 943**
Deposits	**13 866**
Foreign office deposits	1 253
Domestic office deposits	12 613
Interest-bearing deposits	9 477
Noninterest-bearing deposits	3 136
Estimated insured deposits	7 483
Time deposits	1 971
Brokered deposits	1 071
Federal funds purchased & repos	240
FHLB advances	571
Other borrowed money	557
Subordinated debt	69
Trading account liabilities	236
Other liabilities	381
Total liabilities	**15 921**
Total equity capital	**2 023**
Total bank equity capital	2 019
Perpetual preferred stock	9
Common stock	43
Surplus	1 277
Undivided profits	759
Other comprehensive income	-68
Net unrealized P&L on AFS	0

Source: Federal Deposit Insurance Corporation (2019), www.fdic.gov/bank/analytical/qbp.

side. Table 7.2 shows a simplified version of the balance sheet. The bank collects retail and corporate deposits and lends money to households and firms.

TABLE 7.2: A simplified balance sheet

Assets	Liabilities
Cash	Due to central banks
Loans and leases	Deposits
Mortgages	Deposit accounts
Consumer credit	Savings
Credit cards	Term deposits
Interbank loans	Interbank funding
Investment securities	Short-term debt
Sovereign bonds	Subordinated debt
Corporate bonds	Reserves
Other assets	Equity capital

Some deposits have a fixed maturity (e.g. a certificate of deposit), while others have an undefined maturity. This is for example the case of demand deposits or current accounts. These liabilities are then called non-maturity deposits (NMD), and include transaction deposits, NOW (negotiable order of withdrawal) accounts, money market deposit accounts and savings deposits. Term deposits (also known as time deposits or certificates of deposit) are deposits with a fixed maturity, implying that the customer cannot withdraw his funds before the term ends. Generally, the bank considers that the core deposits correspond to deposits of the retail customers and are a stable source of its funding. On the asset side, the bank proposes credit, loans and leases, and holds securities and other assets such as real estate, intangible assets[2] and goodwill[3]. In Chapter 3 on page 125, we have seen that loans concern both individuals, corporates and sovereigns. We generally distinguish loans secured by real estate, consumer loans, commercial and industrial loans. Leases correspond to contract agreements, where the bank purchases the asset on behalf of the customer, and the customer uses the asset in return and pays to the bank a periodic lease payment for the duration of the agreement[4]. Investment securities include repos, sovereign bonds, asset-backed securities, debt instruments and equity securities. We reiterate that the balance sheet does not concern off-balance sheet items. Indeed, the risk of credit lines (e.g. commitments, standby facilities or letters of credit) is measured by the credit risk[5], while derivatives (swaps, forwards, futures and options) are mainly managed within the market risk and the counterparty credit risk.

Another difference between assets and liabilities is that they are not '*priced*' at the same interest rate since the primary business of the bank is to capture the interest rate spread between its assets and its liabilities. The bank receives income from the loans and its investment portfolio, whereas the expenses of the bank concern the interest it pays on deposits and its debt, and the staff and operating costs. In Table 7.3, we report the 2018 income statement of FDIC-insured commercial banks and savings institutions. We can simplify the computation of this income statement and obtain the simplified version

[2] Intangible assets are non-physical assets that have a multi-period useful life such as servicing rights or customer lists. They are also intellectual assets (patents, copyrights, softwares, etc).

[3] Goodwill is the excess of the purchase price over the fair market value of the net assets acquired. The difference can be explained because of the brand name, good customer relations, etc.

[4] At the end of the contract, the customer may have the option to buy the asset.

[5] In this case, the difficult task is to estimate the exposure at default and the corresponding CCF parameter.

TABLE 7.3: Annual income and expense of FDIC-insured commercial banks and savings institutions (Amounts in $ mn)

Total interest income	**660 988**
Domestic office loans	492 201
Foreign office loans	21 965
Lease financing receivables	5 192
Balances due from depository institutions	24 954
Securities	92 908
Trading accounts	11 025
Federal funds sold	8 347
Other interest income	4 397
Total interest expense	**119 799**
Domestic office deposits	74 781
Foreign office deposits	8 877
Federal funds purchased	4 108
Trading liabilities and other borrowed money	28 629
Subordinated notes and debentures	2 780
Net interest income	**541 189**
Provision for loan and lease losses	**49 998**
Total noninterest income	**266 165**
Fiduciary activities	37 525
Service charges on deposit accounts	35 745
Trading account gains and fees	26 755
Interest rate exposures	7 148
Foreign exchange exposures	12 666
Equity security and index exposures	4 750
Commodity and other exposures	1 299
Credit exposures	367
Investment banking, advisory, brokerage and underwriting fees and commissions	12 522
Venture capital revenue	60
Net servicing fees	10 680
Net securitization income	230
Insurance commission fees and income	4 574
Net gains (losses) on sales of loans	12 593
Net gains (losses) on sales of other real estate owned	-99
Net gains (losses) on sales of other assets (except securities)	1 644
Other noninterest income	123 938
Total noninterest expense	**459 322**
Salaries and employee benefits	217 654
Premises and equipment expense	45 667
Other noninterest expense	190 944
Amortization expense and goodwill impairment losses	5 058
Securities gains (losses)	328
Income (loss) before income taxes and extraordinary items	**298 362**
Applicable income taxes	61 058
Extraordinary gains (losses), net	-267
Net charge-offs	47 479
Cash dividends	164 704
Retained earnings	72 045
Net operating income	**237 059**

Source: Federal Deposit Insurance Corporation (2019), www.fdic.gov/bank/analytical/qbp.

given in Table 7.4. Net interest income corresponds to the income coming from interest rates, whereas non-interest income is mainly generated by service fees and commissions. The income statement depends of course on the balance sheet items, but also on off-balance sheet items. Generally, loans, leases and investment securities are called the earning assets, whereas deposits are known as interest bearing liabilities.

TABLE 7.4: A simplified income statement

	Interest income
−	Interest expenses
=	Net interest income
+	Non-interest income
=	Gross income
−	Operating expenses
=	Net income
−	Provisions
=	Earnings before tax
−	Income tax
=	Profit after tax

7.1.1.2 Accounting standards

We understand that the goal of ALM is to control the risk of the balance sheet in order to manage and secure the future income of the bank. However, the ALM policy is constrained by accounting standards since the bank must comply with some important rules that distinguish banking and trading books. Accounting systems differ from one country to another country, but we generally distinguish four main systems: US GAAP[6], Japanese combined system[7], Chinese accounting standards and International Financial Reporting Standards (or IFRS). IFRS are standards issued by the IFRS Foundation and the International Accounting Standards Board (IASB) to provide a global accounting system for business affairs and capital markets. In March 2019, there were 144 jurisdictions that required the use of IFRS Standards for publicly listed companies and 12 jurisdictions that permitted its use. IFRS is then the world's most widely used framework. For example, it is implemented in European Union, Australia, Middle East, Russia, South Africa, etc. Since January 2018, IFRS 9 has replaced IAS 39 that was considered excessively complicated and inappropriate.

Financial instruments IAS 39 required financial assets to be classified in the four following categories:

- financial assets at fair value through profit and loss (FVTPL);
- available-for-sale financial assets (AFS);
- loans and receivables (L&R);
- held-to-maturity investments (HTM).

[6]GAAP stands for Generally Accepted Accounting Principles.

[7]Companies may choose one of the four accepted financial reporting frameworks: Japanese GAAP (which is the most widespread system), IFRS standards, Japan's modified international standards (JMIS) and US GAAP.

The FVTPL category had two subcategories. The first category (designated) included any financial asset that was designated on initial recognition as one to be measured at fair value with fair value changes in profit and loss. The second category (held-for-trading or HFT) included financial assets that were held for trading. Depending on the category, the bank measured the financial asset using the fair value approach[8] (AFS and FVTPL) or the amortized cost approach (L&R and HTM). In IFRS 9, the financial assets are divided into two categories:

- amortized cost (AC);
- fair value (FV).

For FV assets, we distinguish fair value through profit and loss (FVTPL) and fair value through other comprehensive income (FVOCI). Category changes between AC, FVTPL and FVOCI are recognized when the asset is derecognized or reclassified. In fact, the classification of an asset depends on two tests: the business model (BM) test and the solely payments of principal and interest (SPPI) test. In the BM test, the question is to know "*if the objective of the bank is to hold the financial asset to collect the contractual cash flows*" or not. In the SPPI test, the question is rather to understand if "*the contractual terms of the financial asset give rise on specified dates to cash flows that are solely payments of principal and interest on the principal amount outstanding*". It is obvious that the classification of an asset affects the ALM policy because it impacts differently the income statement.

On the liability side, there is little difference between IAS 39 and IFRS 9. All equity investments are measured at fair value, HFT financial liabilities are measured at FVTPL and all other financial liabilities are measured at amortized cost if the fair value option is applied.

Remark 69 *The main revision of IFRS 9 concerns impairment of financial assets since it establishes new models of expected credit loss for receivables and loans. This implies that banks can calculate loss provisioning as soon as the loan is entered the banking book.*

Hedging instruments Hedge accounting is an option and not an obligation. It considers that some financial assets are not held for generating P&L, but are used in order to offset a given risk. This implies that the hedging instrument is fully related to the hedged item. IAS 39 and IFRS 9 recognize three hedging strategies:

- a fair value hedge (FVH) is a hedge of the exposure to changes in fair value of a recognized asset or liability;
- a cash flow hedge (CFH) is a hedge of the exposure to variability in cash flows that is attributable to a particular risk;
- a net investment hedge (NIH) concerns currency risk hedging.

In the case of FVH, fair value of both the hedging instrument and the hedged item are recognized in profit and loss. In the case of CFH or NIH, the effective portion of the gain or loss on the hedging instrument is recognized in equity (other comprehensive income[9] or OCI), while the ineffective portion of the gain or loss on the hedging instrument is recognized in profit and loss.

[8] In the AFS case, gains and losses impact the equity capital and then the balance sheet, whereas gains and losses of FVTPL assets directly concerns the income statement.

[9] See Table 7.1 on page 370.

7.1.1.3 Role and importance of the ALCO

Remark 69 shows that IFRS 9 participates to the convergence of risk, finance and accounting that we recently observe. In fact, ALM is at the junction of these three concepts. This is why we could discuss how to organize the ALM function. Traditionally, it is located in the finance department because the ALM committee (ALCO) is in charge of both risk management and income management. In particular, it must define the funds transfer pricing (FTP) policy. Indeed, resources concerning interest and liquidity risks are transferred from business lines to the ALM portfolio. The ALCO and the ALM unit is in charge to manage the risks of this portfolio, and allocate the P&L across business lines:

> "*A major purpose of internal prices is to determine the P&L of the business lines. Transfer prices are internal prices of funds charged to business units or compensating cheap resources such as deposits. [...] Transfer pricing systems are notably designed for the banking book, for compensating resources collected from depositors and for charging funds used for lending. Internal prices also serve for exchanging funds between units with deficits of funds and units with excesses of funds. As they are used for calculating the P&L of a business line, they perform income allocation across business lines*" (Bessis, 2015, pages 109-110).

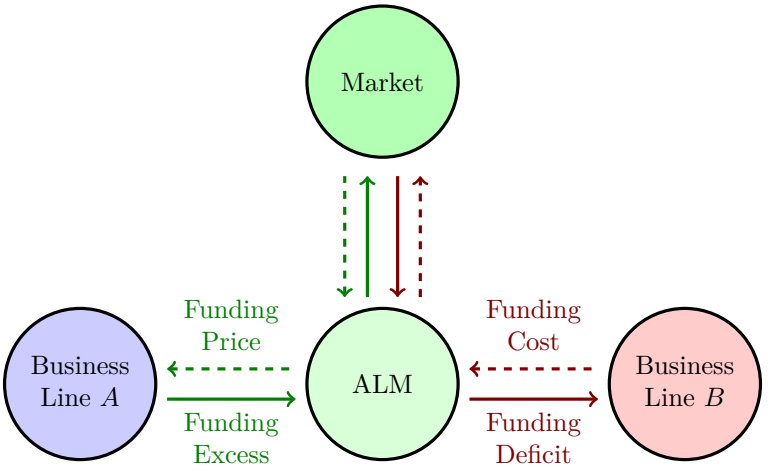


FIGURE 7.1: Internal and external funding transfer

This means that business lines with a funding excess will provide the liquidity to business lines with a funding deficit. For example, Figure 7.1 shows the relationships between the ALM unit and two business lines A and B. In this case, the business line A must be rewarded and receives the funding price, whereas the business B pays the funding cost. Internal funds transfer system avoids that business lines A and B directly go to the market. However, the ALM unit has access to the market for both lending the funding liquidity excess or borrowing the funding liquidity deficit of the bank. At first sight, we can assume that the internal funding price is equal to the external funding price and the internal funding cost is equal to the external funding cost. In this case, the ALM unit captures the bid/ask spread of the funding liquidity. In the real life, it is not possible and it is not necessarily desirable. Indeed, we reiterate that the goal of a bank is to perform liquidity transformation. This means that the liquidity excess of the business line A does not match necessarily the liquidity

deficit of the business line B. Second, the role of the ALM is also to be sure that business lines can pilot their commercial development. In this situation, it is important that internal funding prices and costs are less volatile than external funding prices and costs in order to better stabilize commercial margins. Since the funds transfer pricing policy is decided by the ALCO, we notice that the role of ALM cannot be reduced to a risk management issue. Even if the risk transfer is intentionally rational and fair, meaning that internal prices are related to market prices, the ALM remains a business issue because the assets and liabilities are generally not tradable and there are not always real market prices for these items. For example, what is the price of a \$100 deposit? It depends on the behavior of the customer, but also on the risk appetite of the bank. What is the margin of a \$100 loan? It is not the spread between the loan interest rate and the market interest rate, because there is no perfect matching between the two interest rates. In this case, the margin will depend on the risk management policy. This duality between income generation and risk management is the specificity of asset liability management. Therefore, the role of the ALCO is essential for a bank, because it impacts the risk management of its balance sheet, but also the income generated by its banking book.

7.1.2 Liquidity risk

In this section, we define the concept of liquidity gap, which is the main tool for measuring the ALM liquidity risk. In particular, we make the distinction between static and dynamic liquidity gap when we consider the new production and future projections. In order to calculate liquidity gaps, we also need to understand asset and liability amortization, and liquidity cash flow schedules. Finally, we present liquidity hedging tools, more precisely the standard instruments for managing the ALM liquidity risk.

7.1.2.1 Definition of the liquidity gap

Basel III uses two liquidity ratios (LCR and NSFR), which are related to the ALM liquidity risk. More generally, financial institutions (banks, insurance companies, pension funds and asset managers) manage funding risks by considering funding ratios or funding gaps. The general expression of a funding ratio is:

$$\mathcal{FR}(t) = \frac{A(t)}{L(t)} \tag{7.1}$$

where $A(t)$ is the value of assets and $L(t)$ is the value of liabilities at time t, while the funding gap is defined as the difference between asset value and liability value:

$$\mathcal{FG}(t) = A(t) - L(t) \tag{7.2}$$

If $\mathcal{FR}(t) > 1$ or $\mathcal{FG}(t) > 0$, the financial institution does not need funding because the selling of the assets covers the repayment of the liabilities. Equations (7.1) and (7.2) correspond to the bankruptcy or the liquidation point of view: if we stop the activity, are there enough assets to meet the liability requirements of the financial institution? Another point of view is to consider that the case $A(t) > L(t)$ requires financing the gap $A(t) - L(t)$, implying that the financial institution has to raise liability funding to match the assets. From that point of view, Equations (7.1) and (7.2) becomes[10]:

$$\mathcal{LR}(t) = \frac{L(t)}{A(t)} \tag{7.3}$$

[10] We use the letter $\mathcal{L}$ (liquidity) instead of $\mathcal{F}$ (funding) in order to make the difference between the two definitions.

and:

$$\mathcal{LG}(t) = L(t) - A(t) \tag{7.4}$$

In what follows, we consider the liquidity gap $\mathcal{LG}(t)$ instead of the funding gap $\mathcal{FG}(t)$, meaning that a positive (resp. negative) gap corresponds to a liquidity excess (resp. liquidity deficit).

Example 66 *We consider a simplified balance sheet with few items. The assets* $A(t)$ *are composed of loans that are linearly amortized in a monthly basis during the next year. Their values are equal to* 120*. The liabilities* $L(t)$ *are composed of three short-term in fine debt instruments, and the capital. The corresponding debt notional is respectively equal to* 65*,* 10 *and* 5 *whereas the associated remaining maturity is equal to two, seven and twelve months. The amount of capital is stable for the next twelve months and is equal to* 40*.*

In Table 7.5, we have reported the asset and liability values $A(t)$ and $L(t)$. Since the loans are linearly amortized in a monthly basis, $A(t)$ is equal to 110 after one month, 100 after two months, etc. The value of the first debt instrument remains 65 for the first and second months, and is then equal to zero because the maturity has expired. It follows that the value of the total debt is a piecewise constant function. It is equal to 80 until two months, 15 between three and seven months and 5 after. We can then calculate the liquidity gap. At the initial date, it is equal to zero by definition. At time $t = 1$, we deduce that $\mathcal{LG}(1) = +10$ because we have $A(1) = 110$ and $L(1) = 120$.

TABLE 7.5: Computation of the liquidity gap

Period	0	1	2	3	4	5	6	7	8	9	10	11	12
Loans	120	110	100	90	80	70	60	50	40	30	20	10	0
Assets	120	110	100	90	80	70	60	50	40	30	20	10	0
Debt #1	65	65	65										
Debt #2	10	10	10	10	10	10	10	10					
Debt #3	5	5	5	5	5	5	5	5	5	5	5	5	5
Debt (total)	80	80	80	15	15	15	15	15	5	5	5	5	5
Equity	40	40	40	40	40	40	40	40	40	40	40	40	40
Liabilities	120	120	120	55	55	55	55	55	45	45	45	45	45
$\mathcal{LG}(t)$	0	10	20	−35	−25	−15	−5	5	5	15	25	35	45

The time profile of the liquidity gap is given in Figure 7.2. We notice that it is positive at the beginning, implying that the bank has an excess of liquidity funding in the short-run. Then, we observe that the liquidity gap is negative and the bank needs liquidity funding. From the seventh month, the liquidity gap becomes again positive. At the end, the liquidity gap is always positive since assets and liabilities are fully amortized, implying that the balance sheet is only composed of the capital.

7.1.2.2 Asset and liability amortization

In order to calculate liquidity gaps, we need to understand the amortization of assets and liabilities, in particular the amortization of loans, mortgages, bonds and other debt instruments. The general rules applied to debt payment are the following:

- The annuity amount $A(t)$ at time t is composed of the interest payment $I(t)$ and the principal payment $P(t)$:

$$A(t) = I(t) + P(t)$$

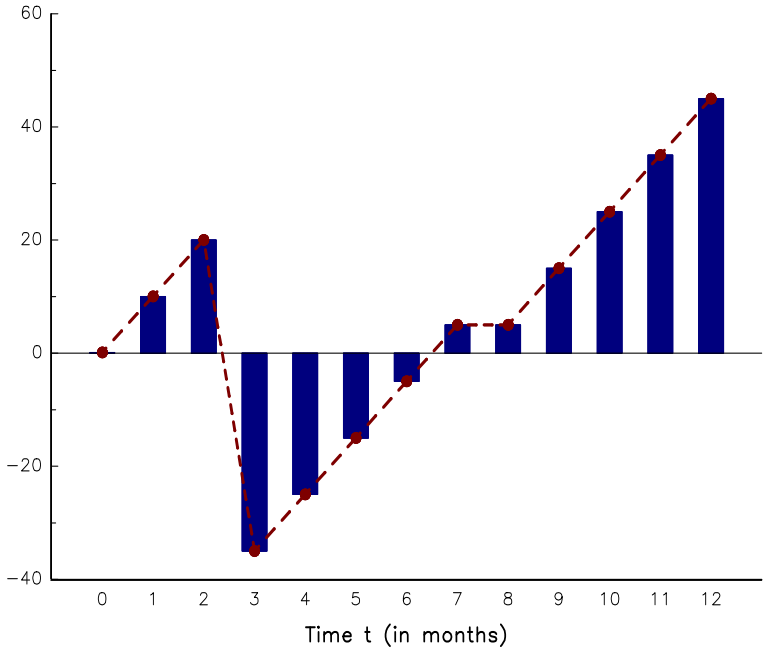


FIGURE 7.2: An example of liquidity gap

This implies that the principal payment at time t is equal to the annuity $A(t)$ minus the interest payment $I(t)$:

$$P(t) = A(t) - I(t)$$

It corresponds to the principal or the capital which is amortized at time t.

- The interest payment at time t is equal to the interest rate $i(t)$ times the outstanding principal balance (or the remaining principal) at the end of the previous period $N(t-1)$:

$$I(t) = i(t) N(t-1)$$

- The outstanding principal balance $N(t)$ is the remaining amount due. It is equal to the previous outstanding principal balance $N(t-1)$ minus the principal payment $P(t)$:

$$N(t) = N(t-1) - P(t) \tag{7.5}$$

At the initial date $t = 0$, the outstanding principal balance is equal to the notional of the debt instrument. At the maturity $t = n$, we must verify that the remaining amount due is equal to zero.

- The outstanding principal balance $N(t)$ is equal to the present value $C(t)$ of forward annuity amounts:

$$N(t) = C(t)$$

We can distinguish different types of debt instruments. For instance, we can assume that the capital is linearly amortized meaning that the principal payment $P(t)$ is constant over time (constant amortization debt). We can also assume that the annuity amount $A(t)$ is constant during the life of the debt instrument (constant payment debt). In this case, the principal payment $P(t)$ is an increasing function with respect to the time t. Another amortization

scheme corresponds to the case where the notional is fully repaid at the time of maturity (bullet repayment debt). This is for example the case of a zero-coupon bond.

Let us consider the case where the interest rate $i(t)$ is constant. For the constant amortization debt, we have:

$$P(t) = \frac{1}{n} N_0$$

where n is the number of periods and N_0 is the notional of the mortgage. The cumulative principal payment $Q(t)$ is equal to:

$$Q(t) = \sum_{s \leq t} P(s) = \frac{t}{n} N_0$$

We deduce that the outstanding principal balance $N(t)$ verifies:

$$N(t) = N_0 - Q(t) = \left(1 - \frac{t}{n}\right) N_0$$

We also have $I(t) = iC(t-1)$ where $C(t-1) = N(t-1)$ and:

$$A(t) = I(t) + P(t) = \left(\frac{1}{n} + i\left(1 - \frac{t-1}{n}\right)\right) N_0$$

In Exercise 7.4.1 on page 449, we derive the formulas of the constant payment debt. The constant annuity is equal to:

$$A(t) = A = \frac{i}{1 - (1+i)^{-n}} N_0$$

It is composed of the interest payment:

$$I(t) = \left(1 - \frac{1}{(1+i)^{n-t+1}}\right) A$$

and the principal payment:

$$P(t) = \frac{1}{(1+i)^{n-t+1}} A$$

Moreover, we show that the outstanding principal balance $N(t)$ verifies:

$$N(t) = \left(\frac{1 - (1+i)^{-(n-t)}}{i}\right) A$$

Finally, in the case of the bullet repayment debt, we have $I(t) = iN_0$, $P(t) = \mathbb{1}\{t = n\} \cdot N_0$, $A(t) = I(t) + P(t)$ and $N(t) = \mathbb{1}\{t \neq n\} \cdot N_0$.

Example 67 *We consider a 10-year mortgage, whose notional is equal to* \$100*. The annual interest rate i is equal to* 5%*, and we assume annual principal payments.*

Results are given in Tables 7.6, 7.7 and 7.8. For each payment structure, we have reported the value of the remaining capital $C(t-1)$ at the beginning of the period, the annuity paid at time t, the split between the interest payment $I(t)$ and the principal payment $P(t)$, the cumulative principal payment $Q(t)$. When calculating liquidity gaps, the most important quantity is the outstanding principal balance $N(t)$ given in the last column, because it corresponds to the amortization of the debt.

TABLE 7.6: Repayment schedule of the constant amortization mortgage

t	$C(t-1)$	$A(t)$	$I(t)$	$P(t)$	$Q(t)$	$N(t)$
1	100.00	15.00	5.00	10.00	10.00	90.00
2	90.00	14.50	4.50	10.00	20.00	80.00
3	80.00	14.00	4.00	10.00	30.00	70.00
4	70.00	13.50	3.50	10.00	40.00	60.00
5	60.00	13.00	3.00	10.00	50.00	50.00
6	50.00	12.50	2.50	10.00	60.00	40.00
7	40.00	12.00	2.00	10.00	70.00	30.00
8	30.00	11.50	1.50	10.00	80.00	20.00
9	20.00	11.00	1.00	10.00	90.00	10.00
10	10.00	10.50	0.50	10.00	100.00	0.00

TABLE 7.7: Repayment schedule of the constant payment mortgage

t	$C(t-1)$	$A(t)$	$I(t)$	$P(t)$	$Q(t)$	$N(t)$
1	100.00	12.95	5.00	7.95	7.95	92.05
2	92.05	12.95	4.60	8.35	16.30	83.70
3	83.70	12.95	4.19	8.77	25.06	74.94
4	74.94	12.95	3.75	9.20	34.27	65.73
5	65.73	12.95	3.29	9.66	43.93	56.07
6	56.07	12.95	2.80	10.15	54.08	45.92
7	45.92	12.95	2.30	10.65	64.73	35.27
8	35.27	12.95	1.76	11.19	75.92	24.08
9	24.08	12.95	1.20	11.75	87.67	12.33
10	12.33	12.95	0.62	12.33	100.00	0.00

TABLE 7.8: Repayment schedule of the bullet repayment mortgage

t	$C(t-1)$	$A(t)$	$I(t)$	$P(t)$	$Q(t)$	$N(t)$
1	100.00	5.00	5.00	0.00	0.00	100.00
2	100.00	5.00	5.00	0.00	0.00	100.00
3	100.00	5.00	5.00	0.00	0.00	100.00
4	100.00	5.00	5.00	0.00	0.00	100.00
5	100.00	5.00	5.00	0.00	0.00	100.00
6	100.00	5.00	5.00	0.00	0.00	100.00
7	100.00	5.00	5.00	0.00	0.00	100.00
8	100.00	5.00	5.00	0.00	0.00	100.00
9	100.00	5.00	5.00	0.00	0.00	100.00
10	100.00	105.00	5.00	100.00	100.00	0.00

Previously, we have assumed that the payment type is annual, but we can consider other periods for the amortization schedule. The most common frequencies are monthly, quarterly, semi-annually and annually[11]. Let i be the annual interest rate and p the frequency or the number of compounding periods per year. The consistency principle of the accumulation factor implies the following identity:

$$(1+i) = \left(1 + \frac{i^{(p)}}{p}\right)^p$$

where $i^{(p)}$ is the nominal interest rate expressed in a yearly basis. For example, if the nominal interest rate $i^{(\text{monthly})}$ is equal to 12%, the borrower pays a monthly interest rate of 1%, which corresponds to an annual interest rate of 12.6825%.

Remark 70 *The interest rate i is also called the annual equivalent rate (AER) or the effective annual rate (EAR).*

Example 68 *We consider a 30-year mortgage, whose notional is equal to* \$100*. The annual interest rate i is equal to* 5%*, and we assume monthly principal payments.*

This example is a variant of the previous example, since the maturity is higher and equal to 30 years, and the payment schedule is monthly. This implies that the number n of periods is equal to 360 months and the monthly interest rate is equal to 5%/12 or 41.7 bps. In Figure 7.3, we show the amortization schedule of the mortgage for the three cases: constant (or linear[12]) amortization, constant payment or annuity and bullet repayment. We notice that the constant annuity case is located between the constant amortization and the bullet repayment. We have also reported the constant annuity case when the interest rate is equal to 10%. We notice that we obtain the following ordering:

$$i_1 \geq i_2 \Rightarrow N(t \mid i_1) \geq N(t \mid i_2)$$

where $N(t)(i)$ is the outstanding principal balance given the interest rate i. In fact, constant annuity and constant amortization coincide when the interest rate goes to zero whereas constant annuity and bullet repayment coincide when the interest rate goes to infinity.

Example 69 *We consider the following simplified balance sheet:*

Assets				*Liabilities*			
Items	*Notional*	*Rate*	*Mat.*	*Items*	*Notional*	*Rate*	*Mat.*
Loan #1	100	5%	10	*Debt #1*	120	5%	10
Loan #2	50	8%	16	*Debt #2*	80	3%	5
Loan #3	40	3%	8	*Debt #3*	70	4%	10
Loan #4	110	2%	7	*Capital #4*	30		

The balance sheet is composed of four asset items and four liability items. Asset items correspond to different loans, whose remaining maturity is respectively equal to 10*,* 16*,* 8 *and* 7 *years. Liabilities contain three debt instruments and the capital, which is not amortized by definition. All the debt instruments are subject to monthly principal payments.*

In Figure 7.4, we have calculated the liquidity gap for different amortization schedule: constant payment, constant annuity and bullet repayment at maturity. We notice that constant payment and constant annuity give similar amortization schedule. This is not the

[11] Monthly is certainly the most used frequency for debt instruments.
[12] The two terms constant and linear can be used interchangeably.

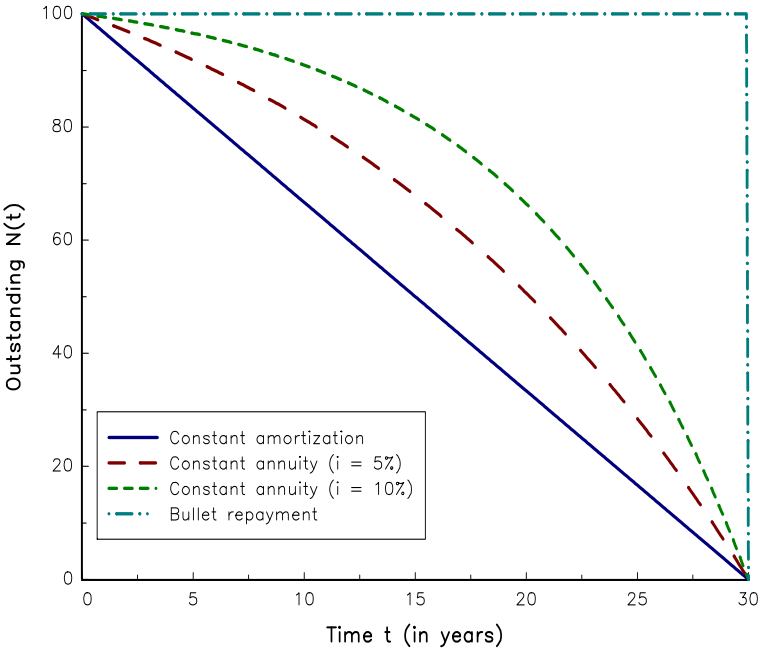


FIGURE 7.3: Amortization schedule of the 30-year mortgage

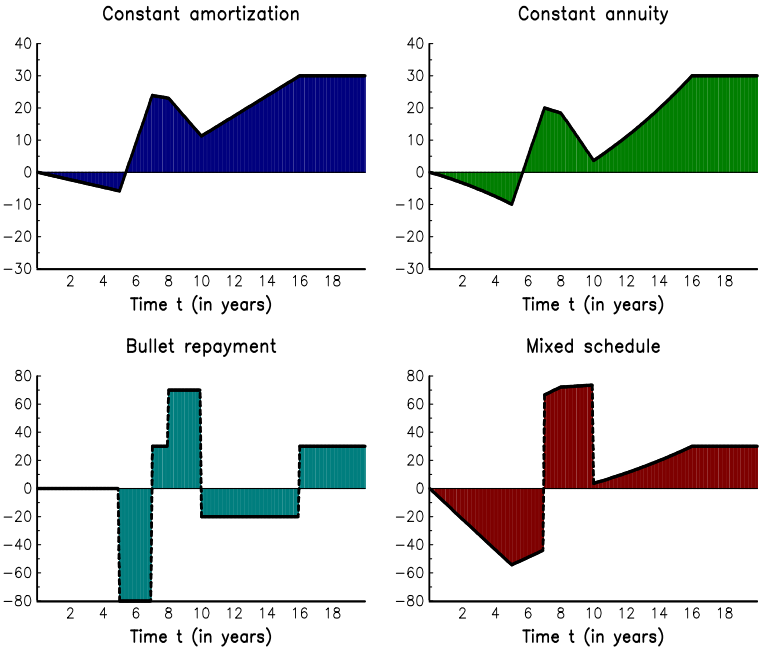


FIGURE 7.4: Impact of the amortization schedule on the liquidity gap

TABLE 7.9: Computation of the liquidity gap (mixed schedule)

t	Assets					Liabilities					$\mathcal{LG}_t$
	#1	#2	#3	#4	A_t	#1	#2	#3	#4	L_t	
1	99.4	49.9	39.6	110	298.8	119.2	78.7	70	30	297.9	−0.92
2	98.7	49.7	39.2	110	297.6	118.5	77.3	70	30	295.8	−1.83
3	98.1	49.6	38.8	110	296.4	117.7	76.0	70	30	293.7	−2.75
4	97.4	49.5	38.3	110	295.2	116.9	74.7	70	30	291.6	−3.66
5	96.8	49.3	37.9	110	294.0	116.1	73.3	70	30	289.4	−4.58
6	96.1	49.2	37.5	110	292.8	115.3	72.0	70	30	287.3	−5.49
7	95.4	49.1	37.1	110	291.6	114.5	70.7	70	30	285.2	−6.41
8	94.8	48.9	36.7	110	290.4	113.7	69.3	70	30	283.1	−7.32
9	94.1	48.8	36.3	110	289.2	112.9	68.0	70	30	280.9	−8.24
10	93.4	48.7	35.8	110	287.9	112.1	66.7	70	30	278.8	−9.15
11	92.8	48.5	35.4	110	286.7	111.3	65.3	70	30	276.7	−10.06
12	92.1	48.4	35.0	110	285.5	110.5	64.0	70	30	274.5	−10.97
0	100.0	50.0	40.0	110	300.0	120.0	80.0	70	30	300.0	0.00
1	92.1	48.4	35.0	110	285.5	110.5	64.0	70	30	274.5	−10.97
2	83.8	46.7	30.0	110	270.4	100.5	48.0	70	30	248.5	−21.90
3	75.0	44.8	25.0	110	254.8	90.1	32.0	70	30	222.1	−32.76
4	65.9	42.7	20.0	110	238.6	79.0	16.0	70	30	195.0	−43.55
5	56.2	40.5	15.0	110	221.7	67.4		70	30	167.4	−54.27
6	46.1	38.1	10.0	110	204.2	55.3		70	30	155.3	−48.91
7	35.4	35.5	5.0		75.9	42.5		70	30	142.5	66.56
8	24.2	32.7			56.9	29.0		70	30	129.0	72.12
9	12.4	29.7			42.1	14.9		70	30	114.9	72.81
10		26.4			26.4				30	30.0	3.62
11		22.8			22.8				30	30.0	7.19
12		18.9			18.9				30	30.0	11.06
13		14.8			14.8				30	30.0	15.24
14		10.2			10.2				30	30.0	19.77
15		5.3			5.3				30	30.0	24.68
16					0.0				30	30.0	30.00

case of bullet repayment. In the fourth panel, we consider a more realistic situation where we have both constant principal (loan #3 and debt #2), constant annuity (loan #1, loan #2 and debt #1) and bullet repayment (loan #4 and debt #2). Computation details for this last mixed schedule are given in Table 7.9. The top panel presents the liquidity gap $\mathcal{LG}(t)$ of the first twelve months while the bottom panel corresponds to the annual schedule. The top panel is very important since it corresponds to the first year, which is the standard horizon used by the ALCO for measuring liquidity requirements. We see that the bank will face a liquidity deficit during the first year.

The previous analysis does not take into account two important phenomena. The first one concerns customer behaviorial options such as prepayment decisions. We note $N^c(t)$ the conventional outstanding principal balance that takes into account the prepayment risk. We have:

$$N^c(t) = N(t) \cdot \mathbb{1}\{\boldsymbol{\tau} > t\}$$

where $N(t)$ is the theoretical outstanding principal balance and $\boldsymbol{\tau}$ is the prepayment time of the debt instrument. The prepayment time in ALM modeling is equivalent to the survival or default time that we have seen in credit risk modeling. Then $\boldsymbol{\tau}$ is a random variable, which is described by its survival function $\mathbf{S}(t)$. Let $p(t)$ be the probability that the debt

instrument has not been repaid at time t. We have:

$$p(t) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\tau} > t\}] = \mathbf{S}(t)$$

By construction, $N^c(t)$ is also random. Therefore, we can calculate its mathematical expectation, and we have $\bar{N}^c(t) = \mathbb{E}[N^c(t)] = p(t) \cdot N(t)$ For example, if we assume that $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$ where λ is the prepayment intensity, we obtain $\bar{N}^c(t) = e^{-\lambda t} \cdot N(t)$. By definition, we always have $N^c(t) \leq N(t)$ and $\bar{N}^c(t) \leq N(t)$.

In Figure 7.5, we consider the constant payment mortgage given in Example 68 on page 381. The first panel shows the theoretical or contractual outstanding principal balance. In the second and third panels, we consider that there is a prepayment at time $\boldsymbol{\tau} = 10$ and $\boldsymbol{\tau} = 20$. This conventional schedule coincides with the contractual schedule, but is equal to zero once the prepayment time occurs. Finally, the fourth panel presents the conventional amortization schedule $\bar{N}^c(t)$ when the prepayment time is exponentially distributed. When λ is equal to zero, we retrieve the previous contractual schedule $N(t)$. Otherwise, the mortgage amortization is quicker.

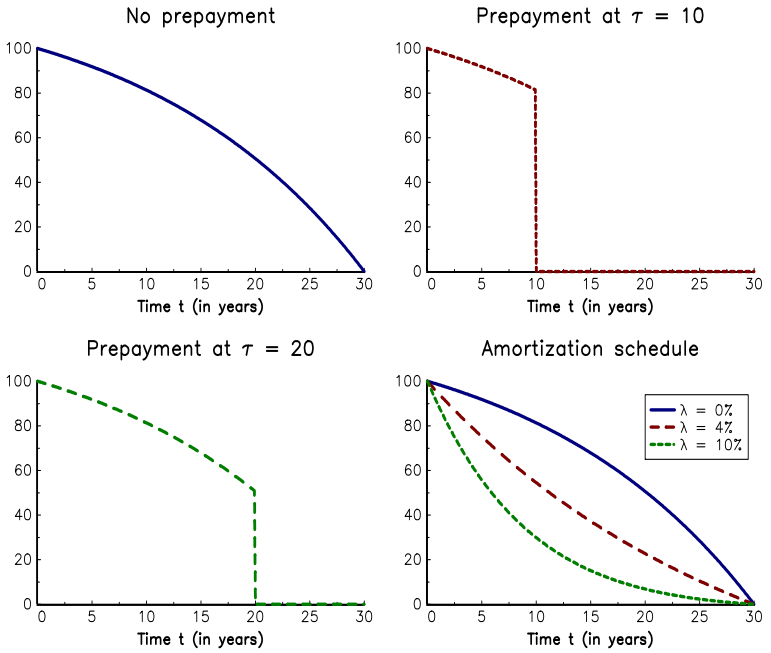


FIGURE 7.5: Conventional amortization schedule with prepayment risk

The second important phenomenon that impacts amortization schedule is the new production of assets and liabilities. If we consider a balance sheet item, its outstanding amount at time t is equal to the outstanding amount at time $t-1$ minus the amortization between t and $t-1$ plus the new production at time t:

$$N(t) = N(t-1) - \text{AM}(t) + \text{NP}(t) \tag{7.6}$$

This relationship is illustrated in Figure 7.6 and can be considered as an accounting identity (Demey *et al.*, 2003). In the case where there is no prepayment, the amortization $\text{AM}(t)$ is exactly equal to the principal payment $P(t)$ and we retrieve Equation (7.5) except the term $\text{NP}(t)$. However, there is a big difference between Equations (7.6) and (7.5). The first one describes the amortization of a debt instrument, for example a loan or a mortgage. The

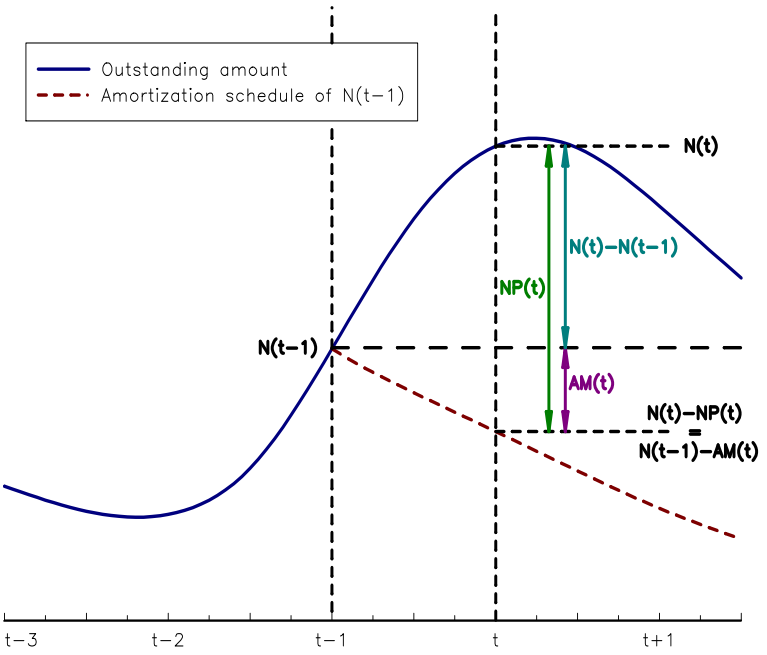


FIGURE 7.6: Impact of the new production on the outstanding amount

Source: Demey *et al.* (2003).

second one describes the amortization of a balance sheet item, that is the aggregation of several debt instruments. The new production $\text{NP}(t)$ corresponds to the financial transactions that appear in the balance sheet between t and $t-1$. They concern the new credit lines, customer loans, mortgages, deposits, etc. that have been traded by the bank during the last period $[t-1, t]$. The introduction of the new production leads to the concept of dynamic liquidity gap, in contrast to the static liquidity gap.

Remark 71 *As we will see in the next section, dynamic liquidity analysis is then more complex since the function* $\text{NP}(t)$ *is not always known and depends on many parameters. Said differently,* $\text{NP}(t)$ *is more a random variable. However, it is more convenient to treat* $\text{NP}(t)$ *as a deterministic function than a stochastic function in order to obtain closed-form formula and not to use Monte Carlo methods*[13].

7.1.2.3 Dynamic analysis

According to BCBS (2016d) and EBA (2018a), we must distinguish three types of analysis:

- Run-off balance sheet
  A balance sheet where existing non-trading book positions amortize and are not replaced by any new business.

[13]Equation (7.6) can also be written as follows:

$$\text{NP}(t) = N(t) - (N(t-1) - \text{AM}(t))$$

Written in this form, this equation indicates how to calculate the new production. In particular, this relationship can be used to define an estimator of $\text{NP}(t)$.

- Constant balance sheet
  A balance sheet in which the total size and composition are maintained by replacing maturing or repricing cash flows with new cash flows that have identical features.

- Dynamic balance sheet
  A balance sheet incorporating future business expectations, adjusted for the relevant scenario in a consistent manner.

The run-off balance sheet analysis has been exposed in the previous section. The constant or dynamic balance sheet analysis assumes that we include the new production when calculating the liquidity gap. For the constant analysis, this task is relatively easy since we consider a like-for-like replacement of assets and liabilities. The dynamic analysis is more difficult to implement because it highly depends "*on key variables and assumptions that are extremely difficult to project with accuracy over an extended period and can potentially hide certain key underlying risk exposures*" (BCBS, 2016d, page 8).

Stock-flow analysis According to Demey *et al.* (2003), the non-static analysis requires a mathematical framework in order to distinguish stock and flow streams. We follow these authors, and more particularly we present the tools introduced in Chapter 1 of their book. We note $\text{NP}(t)$ the new production at time t and $\text{NP}(t,u)$ the part of this production[14] that is always reported in the balance sheet at time $u \geq t$. The amortization function $\mathbf{S}(t,u)$ is defined by the following equation:

$$\text{NP}(t,u) = \text{NP}(t) \cdot \mathbf{S}(t,u)$$

The amortization function is in fact a survival function, implying that the following properties hold: $\mathbf{S}(t,t) = 1$, $\mathbf{S}(t,\infty) = 0$ and $\mathbf{S}(t,u)$ is a decreasing function with respect to u. The amortization function is homogeneous if we have $\mathbf{S}(t,u) = \mathbf{S}(u-t)$ for all $u \geq t$. Otherwise, amortization function is non-homogeneous and may depend on the information $\mathcal{I}_{t:u}$ between t and u. In this case, we can write $\mathbf{S}(t,u) = \mathbf{S}(t,u;\mathcal{I}_{t:u})$ where $\mathcal{I}_{t:u}$ may contain the trajectory of interest rates, the history of prepayment times, etc. We define the amortization rate as the hazard rate associated to the survival function $\mathbf{S}(t,u)$:

$$\lambda(t,u) = -\frac{\partial \ln \mathbf{S}(t,u)}{\partial u}$$

In management, we generally make the distinction between stock and flow streams, but we know that the stock at time t is the sum of past flows. In the case of ALM, the outstanding amount plays the role of stock while the new production corresponds to a flow. Therefore, the outstanding amount at time t is the sum of past productions that are always present in the balance sheet at time t:

$$N(t) = \int_0^\infty \text{NP}(t-s,t)\,\mathrm{d}s$$

If follows that:

$$\begin{aligned} N(t) &= \int_0^\infty \text{NP}(t-s)\,\mathbf{S}(t-s,t)\,\mathrm{d}s \\ &= \int_{-\infty}^t \text{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s \end{aligned} \tag{7.7}$$

[14] We have $\text{NP}(t) = \text{NP}(t,t)$ and $\text{NP}(t,\infty) = 0$.

TABLE 7.10: Relationship between the new production and the outstanding amount

s	NP (s)	$\mathbf{S}(s,7)$	NP $(s,7)$	$\mathbf{S}(s,10)$	NP $(s,10)$	$\mathbf{S}(s,12)$	NP $(s,12)$
1	110	0.301	33.13	0.165	18.18	0.111	12.19
2	125	0.368	45.98	0.202	25.24	0.135	16.92
3	95	0.449	42.69	0.247	23.43	0.165	15.70
4	75	0.549	41.16	0.301	22.59	0.202	15.14
5	137	0.670	91.83	0.368	50.40	0.247	33.78
6	125	0.819	102.34	0.449	56.17	0.301	37.65
7	115	1.000	115.00	0.549	63.11	0.368	42.31
8	152			0.670	101.89	0.449	68.30
9	147			0.819	120.35	0.549	80.68
10	159			1.000	159.00	0.670	106.58
11	152					0.819	124.45
12	167					1.000	167.00
$N(t)$			472.14		640.36		720.69

In the discrete-time analysis, the previous relationship becomes:

$$\begin{aligned} N(t) &= \sum_{s=0}^{\infty} \text{NP}(t-s,t) \\ &= \sum_{s=-\infty}^{t} \text{NP}(s) \cdot \mathbf{S}(s,t) \end{aligned}$$

The outstanding amount $N(t)$ at time t is then the sum of each past production NP (s) times its amortization function $\mathbf{S}(s,t)$. In Table 7.10, we provide an example of calculating the outstanding amount using the previous convolution method. In the second column, we report the production of each year s. We assume that the amortization function is homogeneous and is an exponential distribution with an intensity λ equal to 20%. The third and fourth columns give the values of the amortization function and the production that is present in the balance sheet at time $t = 7$. We obtain $N(7) = 472.14$. The four last columns correspond to the cases $t = 10$ and $t = 12$.

Demey *et al.* (2003) introduce the concept of stock amortization. We recall that the amortization function $\mathbf{S}(t,u)$ indicates the proportion of \$1 entering in the balance sheet at time t that remains present at time $u \geq t$. Similarly, the stock amortization function $\mathbf{S}^{\star}(t,u)$ measures the proportion of \$1 of outstanding amount at time t that remains present at time $u \geq t$. In order to obtain an analytical and tractable function $\mathbf{S}^{\star}(t,u)$, we must assume that the new production is equal to zero after time t. This corresponds to the run-off balance sheet analysis. Demey *et al.* (2003) show that the non-amortized outstanding amount is equal to:

$$N(t,u) = \int_{-\infty}^{t} \text{NP}(s)\, \mathbf{S}(s,u)\, \mathrm{d}s$$

where t is the current time and u is the future date. For instance, $N(5,10)$ indicates the outstanding amount that is present in the balance sheet at time $t = 5$ and will remain in the balance sheet five years after. It follows that:

$$N(t,u) = N(t) \cdot \mathbf{S}^{\star}(t,u)$$

and we deduce that:

$$\begin{aligned}\mathbf{S}^{\star}(t,u) &= \frac{N(t,u)}{N(t)} \\ &= \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,u)\,\mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s}\end{aligned}$$

Dynamics of the outstanding amount Using Equation (7.7), we obtain[15]:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\int_{-\infty}^{t} \mathrm{NP}(s)\,f(s,t)\,\mathrm{d}s + \mathrm{NP}(t) \tag{7.8}$$

where $f(t,u) = -\partial_u \mathbf{S}(t,u)$ is the density function of the amortization. This is the continuous-time version of the amortization schedule given by Equation (7.6):

$$N(t) - N(t-1) = -\mathrm{AM}(t) + \mathrm{NP}(t)$$

where:

$$\mathrm{AM}(t) = \int_{-\infty}^{t} \mathrm{NP}(s)\,f(s,t)\,\mathrm{d}s$$

As already said, we notice the central role of the new production when building a dynamic gap analysis. It is obvious that the new production depends on several parameters, for example the commercial policy of the bank, the competitive environment, etc.

Estimation of the dynamic liquidity gap We can then define the dynamic liquidity gap at time t for a future date $u \geq t$ as follows[16]:

$$\begin{aligned}\mathcal{LG}(t,u) = &\sum_{k \in \mathcal{L}iabilities} \left(N_k(t,u) + \int_t^u \mathrm{NP}_k(s)\,\mathbf{S}_k(s,u)\,\mathrm{d}s\right) - \\ &\sum_{k \in \mathcal{A}ssets} \left(N_k(t,u) + \int_t^u \mathrm{NP}_k(s)\,\mathbf{S}_k(s,u)\,\mathrm{d}s\right)\end{aligned}$$

where k represents a balance sheet item. This is the difference between the liability outstanding amount and the asset outstanding amount. For a given item k, the dynamic outstanding amount is composed of the outstanding amount $N_k(t,u)$ that will be non-amortized at time u plus the new production between t and u that will be in the balance sheet at time u. The difficulty is then to estimate the new production and the amortization function. As said previously, the new production generally depends on the business strategy of the bank.

[15] We have:

$$\begin{aligned}\mathrm{d}\left(\int_{-\infty}^{t} \mathrm{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s\right) &= \mathrm{NP}(t)\,\mathbf{S}(t,t)\,\mathrm{d}t + \int_{-\infty}^{t} \mathrm{NP}(s)\,\frac{\partial\,\mathbf{S}(s,t)}{\partial\,t}\,\mathrm{d}s \\ &= \mathrm{NP}(t)\,\mathrm{d}t - \int_{-\infty}^{t} \mathrm{NP}(s)\left(-\frac{\partial\,\mathbf{S}(s,t)}{\partial\,t}\right)\mathrm{d}s\end{aligned}$$

[16] In the case of the run-off balance sheet, we set $\mathrm{NP}_k(s) = 0$ and we obtain the following formula:

$$\mathcal{LG}(t,u) = \sum_{k \in \mathcal{L}iabilities} N_k(t,u) - \sum_{k \in \mathcal{A}ssets} N_k(t,u)$$

Concerning the amortization function, we can calibrate $\mathbf{S}_k(t,u)$ using a sample of new productions if we assume that the amortization function is homogenous: $\mathbf{S}_k(t,u) = \mathbf{S}_k(u-t)$. It follows that:

$$\hat{\mathbf{S}}_k(u-t) = \frac{\sum_{j\in k} \mathrm{NP}_j(t,u)}{\sum_{j\in k} \mathrm{NP}_j(t)}$$

Moreover, we can show that $\hat{\mathbf{S}}_k(u-t)$ is a convergent estimator and its asymptotic distribution is given by:

$$\hat{\mathbf{S}}_k(u-t) - \mathbf{S}_k(u-t) \to \mathcal{N}(0, H \cdot \mathbf{S}_k(u-t) \cdot (1-\mathbf{S}_k(u-t)))$$

where H is the Herfindahl index associated to the sample of new productions[17].

Remark 72 *This result can be deduced from the empirical estimation theory. Let* $\mathbf{S}(t)$ *be the survival function of the survival time* $\boldsymbol{\tau}$*. The empirical survival function of the weighted sample* $\{(w_j, \boldsymbol{\tau}_j), j = 1,\ldots,n\}$ *is equal to:*

$$\hat{\mathbf{S}}(t) = \frac{\sum_{j=1}^n w_j \cdot D_j}{\sum_{j=1}^n w_j}$$

where $D_j = \mathbb{1}(\boldsymbol{\tau}_j > t)$ *is a Bernoulli random variable with parameter* $p = \mathbf{S}(t)$*. If we assume that the sample observations are independent, we deduce that:*

$$\mathrm{var}\left(\hat{\mathbf{S}}(t)\right) = \frac{\sum_{j=1}^n w_j^2 \cdot \mathrm{var}(D_j)}{\left(\sum_{j=1}^n w_j\right)^2} = \sum_{j=1}^n \frac{w_j^2}{\left(\sum_{j'=1}^n w_{j'}\right)^2} \cdot \mathbf{S}(t) \cdot (1-\mathbf{S}(t))$$

Example 70 *We consider a sample of five loans that belong to the same balance sheet item. Below, we have reported the value taken by* $\mathrm{NP}_j(t,u)$*:*

$u-t$	0	1	2	3	4	5	6	7	8	9	10	11
#1	100	90	80	70	60	50	40	30	20	10	5	0
#2	70	65	55	40	20	10	5	0				
#3	100	95	85	80	60	40	20	10	0			
#4	50	47	44	40	37	33	27	17	10	7	0	
#5	20	18	16	14	10	8	5	3	0			

In Figure 7.7, we have estimated the amortization function $\hat{\mathbf{S}}(u-t)$. We have also computed the variance of the estimator and reported the 95% confidence interval[18].

Liquidity duration Another important tool to measure the mismatch between assets and liabilities is to calculate the liquidity duration, which is defined as the average time of the amortization of the new production $\mathrm{NP}(t)$. In a discrete-time analysis, the amortization value between two consecutive dates is equal to $\mathrm{NP}(t,u) - \mathrm{NP}(t,u+1)$. Therefore, the liquidity duration is the weighted average life (WAL) of the principal repayments:

$$\mathcal{D}(t) = \frac{\sum_{u=t}^{\infty} (\mathrm{NP}(t,u) - \mathrm{NP}(t,u+1)) \cdot (u-t)}{\sum_{u=t}^{\infty} (\mathrm{NP}(t,u) - \mathrm{NP}(t,u+1))}$$

[17] We have $H = \sum_{j\in k} w_j^2$ where:

$$w_j = \frac{\mathrm{NP}_j(t)}{\sum_{j'\in k} \mathrm{NP}_{j'}(t)}$$

[18] We have assumed that the sample is composed of 100 loans or 20 copies of the five previous loans. Otherwise, the confidence interval is too large because the sample size is small.

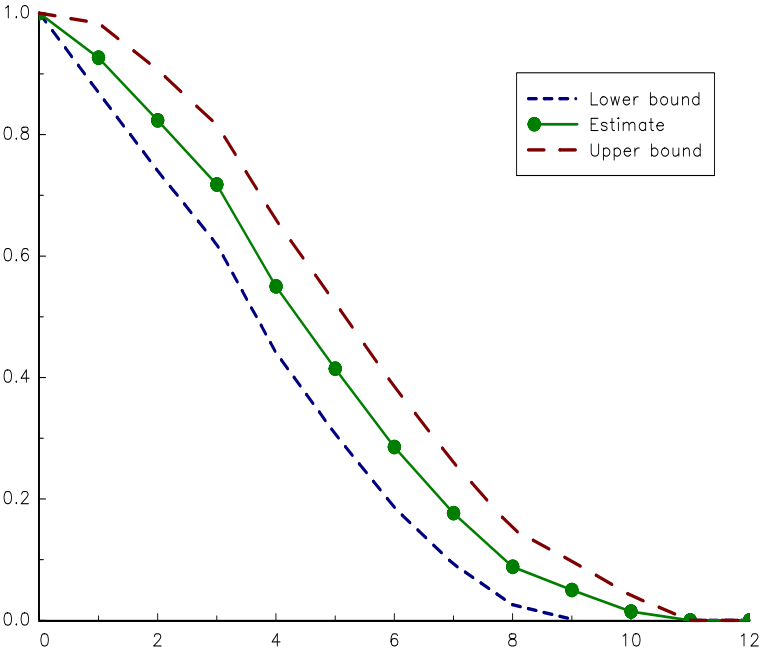


FIGURE 7.7: Estimation of the amortization function $\hat{\mathbf{S}}(u-t)$

Since we have:

$$\mathrm{NP}(t,u) - \mathrm{NP}(t,u+1) = -\mathrm{NP}(t) \cdot (\mathbf{S}(t,u+1) - \mathbf{S}(t,u))$$

and[19]:

$$\begin{aligned}\sum_{u=t}^{\infty} (\mathrm{NP}(t,u) - \mathrm{NP}(t,u+1)) &= -\mathrm{NP}(t) \cdot \sum_{u=t}^{\infty} (\mathbf{S}(t,u+1) - \mathbf{S}(t,u)) \\ &= \mathrm{NP}(t)\end{aligned}$$

we obtain the following formula:

$$\mathcal{D}(t) = -\sum_{u=t}^{\infty} (\mathbf{S}(t,u+1) - \mathbf{S}(t,u)) \cdot (u-t)$$

In the continuous-time analysis, the liquidity duration is equal to:

$$\begin{aligned}\mathcal{D}(t) &= -\int_t^{\infty} \frac{\partial \mathbf{S}(t,u)}{\partial u} (u-t)\, \mathrm{d}u \\ &= \int_t^{\infty} (u-t)\, f(t,u)\, \mathrm{d}u\end{aligned}$$

where $f(t,u)$ is the density function associated to the survival function $\mathbf{S}(t,u)$.

Remark 73 *If we consider the stock approach of the liquidity duration, we have:*

$$\mathcal{D}^{\star}(t) = \int_t^{\infty} (u-t)\, f^{\star}(t,u)\, \mathrm{d}u$$

[19] Because we have $\mathbf{S}(t,t) = 1$ and $\mathbf{S}(t,\infty) = 0$.

where $f^\star(t,u)$ *is the density function associated to the survival function* $\mathbf{S}^\star(t,u)$*:*

$$\begin{aligned} f^\star(t,u) &= -\frac{\partial\, \mathbf{S}^\star(t,u)}{\partial\, u} \\ &= \frac{\int_{-\infty}^{t} \mathrm{NP}(s)\, f(s,u)\, \mathrm{d}s}{\int_{-\infty}^{t} \mathrm{NP}(s)\, \mathbf{S}(s,t)\, \mathrm{d}s} \end{aligned}$$

Some examples We consider the three main amortization schemes: bullet repayment, constant (or linear) amortization and exponential amortization. In Exercise 7.4.3 on page 450, we have calculated the survival functions $\mathbf{S}(t,u)$ and $\mathbf{S}^\star(t,u)$, the liquidity duration $\mathcal{D}(t)$ and $\mathcal{D}^\star(t)$ and the outstanding dynamics $\mathrm{d}N(t)$ where m is the debt maturity and λ is the exponential parameter. Their expression is reported below in Table 7.11.

TABLE 7.11: Amortization function and liquidity duration of the three amortization schemes

Amortization	$\mathbf{S}(t,u)$	$\mathcal{D}(t)$
Bullet	$\mathbb{1}\{t \le u < t+m\}$	m
Constant	$\mathbb{1}\{t \le u < t+m\} \cdot \left(1 - \dfrac{u-t}{m}\right)$	$\dfrac{m}{2}$
Exponential	$e^{-\lambda(u-t)}$	$\dfrac{1}{\lambda}$
Amortization	$\mathbf{S}^\star(t,u)$	$\mathcal{D}^\star(t)$
Bullet	$\mathbb{1}\{t \le u < t+m\} \cdot \left(1 - \dfrac{u-t}{m}\right)$	$\dfrac{m}{2}$
Constant	$\mathbb{1}\{t \le u < t+m\} \cdot \left(1 - \dfrac{u-t}{m}\right)^2$	$\dfrac{m}{3}$
Exponential	$e^{-\lambda(u-t)}$	$\dfrac{1}{\lambda}$
Amortization	$\mathrm{d}N(t)$	
Bullet	$\mathrm{d}N(t) = (\mathrm{NP}(t) - \mathrm{NP}(t-m))\, \mathrm{d}t$	
Constant	$\mathrm{d}N(t) = \left(\mathrm{NP}(t) - \dfrac{1}{m}\displaystyle\int_{t-m}^{t} \mathrm{NP}(s)\, \mathrm{d}s\right) \mathrm{d}t$	
Exponential	$\mathrm{d}N(t) = (\mathrm{NP}(t) - \lambda N(t))\, \mathrm{d}t$	

We have represented these amortization functions $\mathbf{S}(t,u)$ and $\mathbf{S}^\star(t,u)$ in Figure 7.8. The maturity m is equal to 10 years and the exponential parameter λ is set to 30%. Besides the three previous amortization schemes, we also consider the constant payment mortgage (CPM), whose survival functions are equal to[20]:

$$\mathbf{S}(t,u) = \mathbb{1}\{t \le u < t+m\} \cdot \frac{1 - e^{-i(t+m-u)}}{1 - e^{-im}}$$

and:

$$\mathbf{S}^\star(t,u) = \frac{i(t+m-u) + e^{-i(t+m-u)} - 1}{im + e^{-im} - 1}$$

[20]These expressions are derived in Exercise 7.4.3 on page 450.

where i is the interest rate and m is the debt maturity. The CPM amortization scheme corresponds to the bottom/right panel[21] in Figure 7.8.

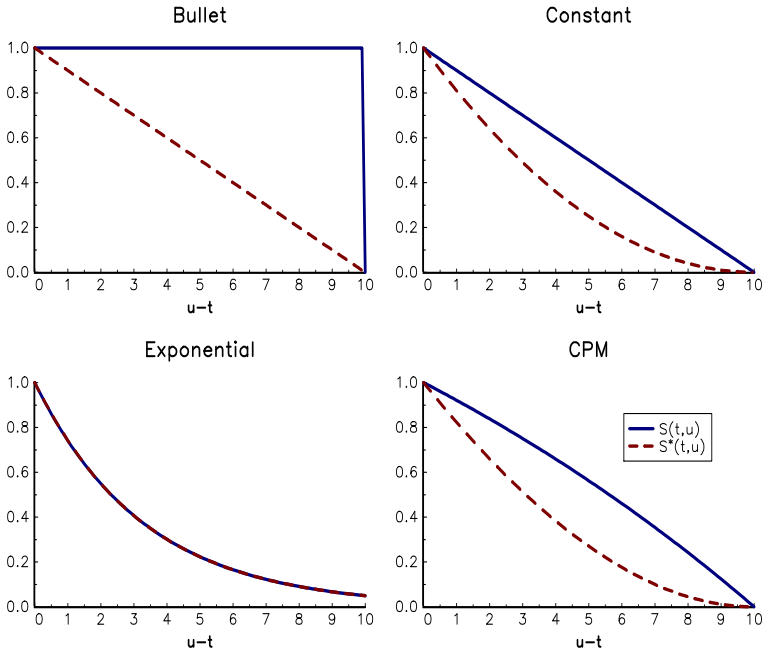


FIGURE 7.8: Amortization functions $\mathbf{S}(t,u)$ and $\mathbf{S}^{\star}(t,u)$

Remark 74 *We notice the convex profile of the constant and exponential amortization schemes, whereas the profile is concave for the CPM amortization scheme. Moreover, when the interest rate i goes to zero, the CPM profile corresponds to the constant profile.*

7.1.2.4 Liquidity hedging

When we face a risk that is not acceptable, we generally hedge it. In the case of the liquidity, the concept of hedging is unclear. Indeed, at first sight, it seems that there are no liquidity forwards, swaps or options in the market. On the other hand, liquidity hedging seems to be trivial. Indeed, the bank can lend to other market participants when having an excess of funding, or the bank can borrow when having a deficit of funding. For that, it may use the interbank market or the bond market. Nevertheless, there is generally an uncertainty about the liquidity gap, because the amortization schedule and the new production are not known for sure. This is why banks must generally adopt a conservative approach. For instance, they must not lend (or buy bonds) too much. In a similar way, they must not borrow too short. The liquidity gap analysis is particularly important in order to split the decision between daily, weekly, monthly and quarterly adjustments. Let us assume that the bank anticipates a liquidity deficit of \$10 mn for the next three months. It can borrow exactly \$10 mn for three months. One month later, the bank has finally an excess of liquidity. It is obvious that the previous lending is not optimal because the bank must pay a three-month interest rate while it could have paid a one-month interest rate.

The previous example shows that the management of the liquidity consists in managing interbank and bond operations. It is obvious that the funding program depends on the

[21] The interest rate i is set to 5%.

liquidity gap but also on the risk appetite of the bank. Some banks prefer to run a long-term liquidity program, others prefer to manage the liquidity on a shorter-term basis. The ALCO decisions may have therefore a big impact on the risk profile of the bank. The 2008 Global Financial Crisis has demonstrated that liquidity management is key during periods of stress. For instance, a bank, which has a structural liquidity excess, may stop to lend to the other participants in order to keep this liquidity for itself, while a bank, which has a structural liquidity need, may issue long-term debt in order to reduce day-to-day funding requirements. It is clear that ALCO decisions are beyond the scope of risk management and fall within strategic and business issues.

7.1.3 Interest rate risk in the banking book

The ALM of interest rate risk is extensively developed in the next section. However, we give here the broad lines, notably the regulation framework, which has been elaborated by the Basel Committee in April 2016 (BCBS, 2016d) and which is known as IRRBB (or interest rate risk in the banking book). IRRBB can be seen as the revision of the 2004 publication (BCBS, 2004b), but not solely. Indeed, this 2016 publication is relatively precise in terms of risk framework and defines a standardized framework, which was not the case in 2004. In particular, capital requirements are more closely supervised than previously, even if IRRBB continues to be part of the Basel capital framework's Pillar 2.

7.1.3.1 Introduction on IRRBB

Definition of IRRBB According to BCBS (2016d), "*IRRBB refers to the current or prospective risk to the bank' capital and earnings arising from adverse movements in interest rates that affect the bank's banking book positions. When interest rates change, the present value and timing of future cash flows change. This in turn changes the underlying value of a bank's assets, liabilities and off-balance sheet items and hence its economic value. Changes in interest rates also affect a bank's earnings by altering interest rate-sensitive income and expenses, affecting its net interest income*". We notice that the Basel Committee considers both economic value (EV) and earnings-based risk measures. EV measures reflect changes in the net present value of the balance sheet resulting from IRRBB, whereas earnings-based measures reflect changes in the expected future profitability of the bank. Since EV measures are generally used by supervisors[22] and earnings-based measures are more widely used by commercial banks[23], the Basel Committee thinks that the bank must manage these two risks because they capture two different time horizons. Economic value is calculated over the remaining life of debt instruments, implying a run-off balance sheet assumption. The earnings-based measure is calculated for a given time horizon, typically the next 12 month period. In this case, a constant or dynamic balance sheet assumption is more appropriate.

Categories of IRR For the Basel Committee, there are three main sources of interest rate risk: gap risk, basis risk and option risk. Gap risk refers to the mismatch risk arising from the term structure of banking book instruments. It includes repricing risk and yield curve risk. Repricing risk corresponds to timing differences in the maturity or the risk of changes in interest rates between assets and liabilities. For example, if the bank funds a long-term fixed-rate loan with a short-term floating-rate deposit, the future income may decrease if interest rates increases. Therefore, repricing risk has two components. The first one is the maturity difference between assets and liabilities. The second one is the change in

[22] Because they are more adapted for comparing banks.
[23] Because banks want to manage the volatility of earnings.

floating interest rates. Yield curve risk refers to non-parallel changes in the term structure of interest rates. A typical example concerns flattening, when short-term interest rates rise faster than long-term interest rates.

Basis risk occurs when changes in interest rates impact differently financial instruments with similar repricing tenors, because they are priced using different interest rate indices. Therefore, basis risk corresponds to the correlation risk of interest rate indices with the same maturity. For example, the one-month Libor rate is not perfectly correlated to the one-month Treasury rate. Thus, there is a basis risk when a one-month Treasury-based asset is funded with a one-month Libor-based liability, because the margin can change from one month to another month.

Option risk arises from option derivative positions or when the level or the timing of cash flows may change due to embedded options. A typical example is the prepayment risk. The Basel Committee distinguishes automatic option risk and behavioral option risk. Automatic options concern caps, floors, swaptions and other interest rate derivatives that are located in the banking book, while behavioral option risk includes fixed rate loans subject to prepayment risk, fixed rate loan commitments, term deposits subject to early redemption risk and non-maturity deposits (or NMDs).

Risk measures The economic value of a series of cash flows $\mathrm{CF} = \{\mathrm{CF}(t_k), t_k \geq t\}$ is the present value of these cash flows:

$$\begin{aligned}
\mathrm{EV} &= PV_t(\mathrm{CF}) \\
&= \mathbb{E}\left[\sum_{t_k \geq t} \mathrm{CF}(t_k) \cdot e^{-\int_t^{t_k} r(s)\,\mathrm{d}s}\right] \\
&= \sum_{t_k \geq t} \mathrm{CF}(t_k) \cdot B(t, t_k)
\end{aligned}$$

where $B_t(t_k)$ is the discount factor (e.g. the zero-coupon bond) for the maturity date t_k. To calculate the economic value of the banking book, we slot all notional repricing cash flows of assets and liabilities into a set of time buckets. Then, we calculate the net cash flows, which are equal to $\mathrm{CF}(t_k) = \mathrm{CF}_A(t_k) - \mathrm{CF}_L(t_k)$ where $\mathrm{CF}_A(t_k)$ and $\mathrm{CF}_L(t_k)$ are the cash flows of assets and liabilities for the time bucket t_k. Finally, the economic value is given by:

$$\begin{aligned}
\mathrm{EV} &= \sum_{t_k \geq t} \mathrm{CF}(t_k) \cdot B(t, t_k) \\
&= \sum_{t_k \geq t} \mathrm{CF}_A(t_k) \cdot B(t, t_k) - \sum_{t_k \geq t} \mathrm{CF}_L(t_k) \cdot B(t, t_k) \\
&= \mathrm{EV}_A - \mathrm{EV}_L
\end{aligned}$$

It is equal to the present value of assets minus the present value of liabilities. By construction, the computation of EV depends on the yield curve. We introduce the notation s in order to take into account a stress scenario of the yield curve. Then, we define the EV change as the difference between the EV for the base scenario and the EV for the given scenario s:

$$\begin{aligned}
\Delta\,\mathrm{EV}_s &= \mathrm{EV}_0 - \mathrm{EV}_s \\
&= \sum_{t_k \geq t} \mathrm{CF}_0(t_k) \cdot B_0(t, t_k) - \sum_{t_k \geq t} \mathrm{CF}_s(t_k) \cdot B_s(t, t_k)
\end{aligned}$$

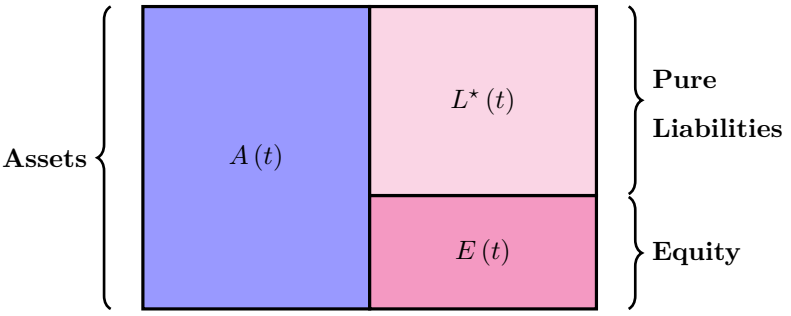


FIGURE 7.9: Relationship between $A(t)$, $L^\star(t)$ and $E(t)$

In this equation, the base scenario is denoted by 0 and corresponds to the current term structure of interest rates. The stress scenario s of the yield curve impacts the discount factors, but also the cash flows that depend on the future interest rates. $\Delta \text{EV}_s > 0$ indicates then a loss if the stress scenario s occurs. The Basel Committee defines the concept of economic value of equity (EVE or EV_E) as a specific form of EV where equity is excluded from the cash flows. We recall that the value of assets is equal to the value of liabilities at the current time t. If we distinguish pure liabilities $L^\star(t)$ from the bank equity capital $E(t)$, we obtain the balance sheet given in Figure 7.9. Since there is a perfect match between assets and liabilities, the value of the capital is equal to[24] $E(t) = A(t) - L^\star(t)$. It follows that:

$$\text{EVE} = \text{EV}_A - \text{EV}_{L^\star}$$

We can then define ΔEVE_s as the loss ΔEV_s where we have excluded the equity from the computation of the cash flows. Said differently, ΔEVE_s is the capital loss if the stress scenario s occurs.

Remark 75 *Changes in economic value can also be measured with the PV01 metric or the economic value-at-risk (EVaR). PV01 is calculated by assuming a single basis point change in interest rates. EVaR is the value-at-risk measure applied to the economic value of the banking book. Like the VaR, it requires specifying the holding period and the confidence level. The Basel Committee motivates the choice of EVE instead of PV01 and EVaR, because they would like to measure the impact of losses on the capital in a stress testing framework. In particular, PV01 ignores basis risks whereas EVaR is designed for normal market circumstances.*

Earnings-based measures are computed using the net interest income (NII), which is the difference between the interest payments on assets and the interest payments of liabilities. Said differently, NII is the difference between interest rate revenues received by the bank and interest rate costs paid by the bank. For a given scenario s, we define the change in net interest income as follows:

$$\Delta \text{NII}_s = \text{NII}_0 - \text{NII}_s$$

Like for the risk measure ΔEVE_s, $\Delta \text{NII}_s > 0$ indicates a loss if the stress scenario s occurs.

[24] We have:

$$\begin{aligned} A(t) &= L(t) \\ &= L^\star(t) + E(t) \end{aligned}$$

Finally, the economic value and earnings-based risk measures are equal to the maximum of losses by considering the different scenarios:

$$\mathcal{R}(\text{EVE}) = \max_s (\Delta \text{EVE}_s, 0)$$

and:

$$\mathcal{R}(\text{NII}) = \max_s (\Delta \text{NII}_s, 0)$$

Since IRRBB is part of the second pillar, there are no minimum capital requirements $\mathcal{K}$. Nevertheless, the Basel Committee imposes that $\mathcal{R}(\text{EVE})$ must be lower than 15% of the bank's tier 1 capital.

7.1.3.2 Interest rate risk principles

The Basel Committee defines nine IRR principles for banks and three IRR principles for supervisors. The first and second principles recall that banks must specifically manage IRRBB (and also CSRBB[25]) and have a governing body that oversights IRRBB. The third and fourth principles explain that the risk appetite of the bank for IRRBB must be defined with respect to both economic value and earnings-based risk measures arising from interest rate shocks and stress scenarios. The objective is to measure the change in the net present value of the banking book and the future profitability. To compute ΔEVE, banks must consider a run-off balance sheet assumption, whereas they must use a constant or dynamic balance sheet and a rolling 12-month period for computing ΔNII. For that, they have to consider multiple interest rate scenarios, for example historical and hypothetical scenarios. Besides these internal scenarios, six external scenarios are defined by the Basel Committee[26]: (1) parallel shock up; (2) parallel shock down; (3) steepener shock (short rates down and long rates up); (4) flattener shock (short rates up and long rates down); (5) short rates shock up; and (6) short rates shock down. The fifth principle deals with behaviorial and modeling assumptions, in particular embedded optionalities. The last three principles deals with risk management and model governance process, the disclosure of the information and the capital adequacy policy.

The role of supervisors is strengthened. They should collect on a regular basis sufficient information from the bank to assess its IRRBB exposure. This concerns modeling assumptions, interest rate and option exposures, yield curve parameters, statistical methodologies, etc. An important task is also the identification of outlier banks. The outlier/materiality test compares the bank's maximum ΔEVE (or $\mathcal{R}(\text{EVE})$) with 15% of its tier 1 capital. If this threshold is exceeded, supervisors must require mitigation actions, hedging programs and/or additional capital.

7.1.3.3 The standardized approach

Overview of the standardized framework There are five steps for measuring the bank's IRRBB:

1. The first step consists in allocating the interest rate sensitivities of the banking book to three categories:
   (a) amenable to standardization[27];
   (b) less amenable to standardization[28];
   (c) not amenable to standardization[29].

[25] Credit spread risk in the banking book.
[26] These scenarios are described in the next paragraph on page 397.
[27] The Basel Committee distinguish two main categories: fixed rate positions and floating rate positions.
[28] They concern explicit automatic interest rate options.
[29] This category is composed of NMDs, fixed rate loans subject to prepayment risk and term deposits subject to early redemption risk.

2. Then, the bank must slot cash flows (assets, liabilities and off-balance sheet items) into 19 predefined time buckets[30]: overnight (O/N), O/N–1M, 1M–3M, 3M–6M, 6M–9M, 9M–1Y, 1Y–18M, 1.5Y–2Y, 2Y–3Y, 3Y–4Y, 4Y–5Y, 5Y–6Y, 6Y–7Y, 7Y–8Y, 8Y–9Y, 9Y–10Y, 10Y–15Y, 15Y–20Y, 20Y+. This concerns positions amenable to standardization. For positions less amenable to standardization, they are excluded from this step. For positions with embedded automatic interest rate options, the optionality is ignored.

3. The bank determines $\Delta\,\mathrm{EVE}_{s,c}$ for each interest rate scenario s and each currency c.

4. In the fourth step, the bank calculates the total measure for automatic interest rate option risk $\mathrm{KAO}_{s,c}$.

5. Finally, the bank calculates the EVE risk measure for each interest rate shock s:

$$\mathcal{R}\left(\mathrm{EVE}_s\right) = \max\left(\sum_c \left(\Delta\,\mathrm{EVE}_{s,c} + \mathrm{KAO}_{s,c}\right)^+ ; 0\right)$$

The standardized EVE risk measure is the maximum loss across all the interest rate shock scenarios:

$$\mathcal{R}\left(\mathrm{EVE}\right) = \max_s \mathcal{R}\left(\mathrm{EVE}_s\right)$$

The supervisory interest rate shock scenarios The six stress scenarios are based on three shock sizes that the Basel Committee has calibrated using the period 2010 – 2015: the parallel shock size $\mathbb{S}_0$, the short shock size $\mathbb{S}_1$ and the long shock size $\mathbb{S}_2$. In the table below, we report their values for some currencies[31]:

Shock size		USD/CAD/SEK	EUR/HKD	GBP	JPY	EM
$\mathbb{S}_0$	(parallel)	200	200	250	100	400
$\mathbb{S}_1$	(short)	300	250	300	100	500
$\mathbb{S}_2$	(long)	150	100	150	100	300

where EM is composed of ARS, BRL, INR, MXN, RUB, TRY and ZAR. Given $\mathbb{S}_0$, $\mathbb{S}_1$ and $\mathbb{S}_2$, we calculate the following generic shocks for a given maturity t:

	Parallel shock $\Delta R^{(\mathrm{parallel})}(t)$	Short rates shock $\Delta R^{(\mathrm{short})}(t)$	Long rates shock $\Delta R^{(\mathrm{long})}(t)$
Up	$+\mathbb{S}_0$	$+\mathbb{S}_1 \cdot e^{-t/\tau}$	$+\mathbb{S}_2 \cdot \left(1 - e^{-t/\tau}\right)$
Down	$-\mathbb{S}_0$	$-\mathbb{S}_1 \cdot e^{-t/\tau}$	$-\mathbb{S}_2 \cdot \left(1 - e^{-t/\tau}\right)$

where τ is equal to four years. Finally, the five standardized interest rate shock scenarios are defined as follows:

1. Parallel shock up:

$$\Delta R^{(\mathrm{parallel})}(t) = +\mathbb{S}_0$$

[30] The buckets are indexed by k from 1 to 19. For each bucket, the midpoint is used for defining the corresponding maturity t_k. We have $t_1 = 0.0028$, $t_2 = 0.0417$, $t_3 = 0.1667$, $t_4 = 0.375$, $t_5 = 0.625$, …, $t_{17} = 12.5$, $t_{18} = 17.5$ and $t_{19} = 25$.

[31] The values for a more comprehensive list of currencies are given in BCBS (2016d) on page 44.

2. Parallel shock down:
$$\Delta R^{(\text{parallel})}(t) = -\mathbb{S}_0$$
3. Steepener shock (short rates down and long rates up):
$$\Delta R^{(\text{steepnener})}(t) = 0.90 \cdot \left|\Delta R^{(\text{long})}(t)\right| - 0.65 \cdot \left|\Delta R^{(\text{short})}(t)\right|$$
4. Flattener shock (short rates up and long rates down):
$$\Delta R^{(\text{flattener})}(t) = 0.80 \cdot \left|\Delta R^{(\text{short})}(t)\right| - 0.60 \cdot \left|\Delta R^{(\text{long})}(t)\right|$$
5. Short rates shock up:
$$\Delta R^{(\text{short})}(t) = +\mathbb{S}_1 \cdot e^{-t/\tau}$$
6. Short rates shock down:
$$\Delta R^{(\text{short})}(t) = -\mathbb{S}_1 \cdot e^{-t/\tau}$$

Example 71 *We assume that* $\mathbb{S}_0 = 100$ *bps,* $\mathbb{S}_1 = 150$ *bps and* $\mathbb{S}_2 = 200$ *bps. We would like to calculate the standardized shocks for the one-year maturity.*

The parallel shock up is equal to +100 bps, while the parallel shock down is equal to −100 bps. For the short rates shock, we obtain:
$$\Delta R^{(\text{short})}(t) = 150 \times e^{-1/4} = 116.82 \text{ bps}$$
for the up scenario and −116.82 bps for the down scenario. Since we have $\left|\Delta R^{(\text{short})}(t)\right| = 116.82$ and $\left|\Delta R^{(\text{long})}(t)\right| = 44.24$, the steepener shock is equal to:
$$\Delta R^{(\text{steepnener})}(t) = 0.90 \times 44.24 - 0.65 \times 116.82 = -36.12 \text{ bps}$$
For the flattener shock, we have:
$$\Delta R^{(\text{flattener})}(t) = 0.80 \times 116.82 - 0.60 \times 44.24 = 66.91 \text{ bps}$$
In Figure 7.10, we have represented the six interest rate shocks $\Delta R(t)$ for the set of parameters $(\mathbb{S}_0 = 100, \mathbb{S}_1 = 150, \mathbb{S}_2 = 200)$.

In Figure 7.11, we consider the yield curve generated by the Nelson-Siegel model[32] with the following parameters $\theta_1 = 8\%$, $\theta_2 = -7\%$, $\theta_3 = 6\%$ and $\theta_4 = 10$. Then, we apply the standardized interest rate shocks by considering EUR and EM currencies. We verify that the parallel shock moves uniformly the yield curve, the steepener shock increases the slope of the yield curve, the flattener shock reduces the spread between long and short interest rates, and the short rates shock has no impact on the long maturities after 10 years. We also notice that the deformation of the yield curve is more important for EM currencies than for the EUR currency.

Treatment of NMDs NMDs are segmented into three categories: retail transactional, retail non-transactional and wholesale Then, the bank must estimate the stable and non-stable part of each category[33]. Finally, the stable part of NMDs must be split between core and non-core deposits. However, the Basel Committee imposes a cap ω^+ on the proportion of core deposits (see Table 7.12). For instance, core deposits cannot exceed 70% of the retail non-transactional stable deposits. The time bucket for non-core deposits is set to overnight (or the first time bucket), meaning that the corresponding time bucket midpoint is equal to $t_1 = 0.0028$. For core deposits, the bank determines the appropriate cash flow slotting, but the average maturity cannot exceed the cap t^+, which is given in Table 7.12.

[32] We recall that it is defined in Footnote 8 on page 131.
[33] This estimation must be based on the historical data of the last 10 years.

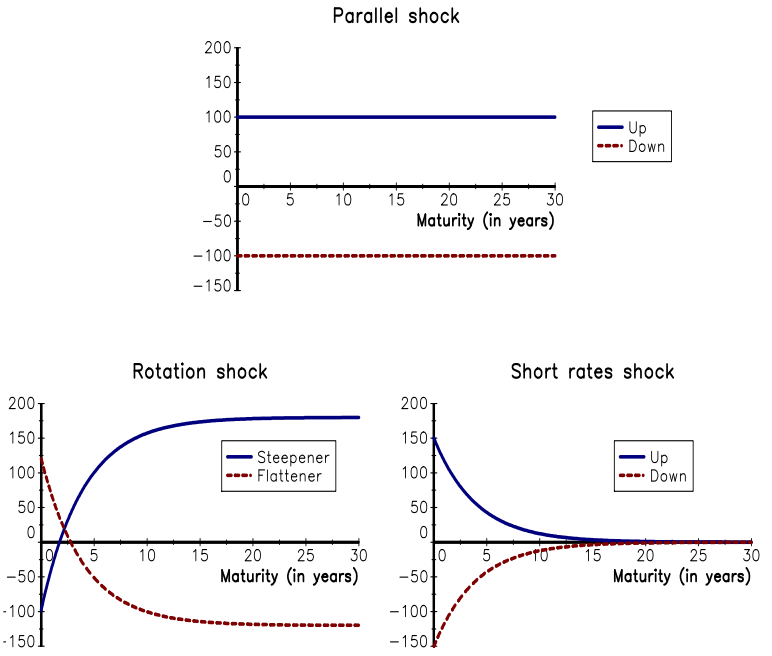


FIGURE 7.10: Interest rate shocks (in bps)

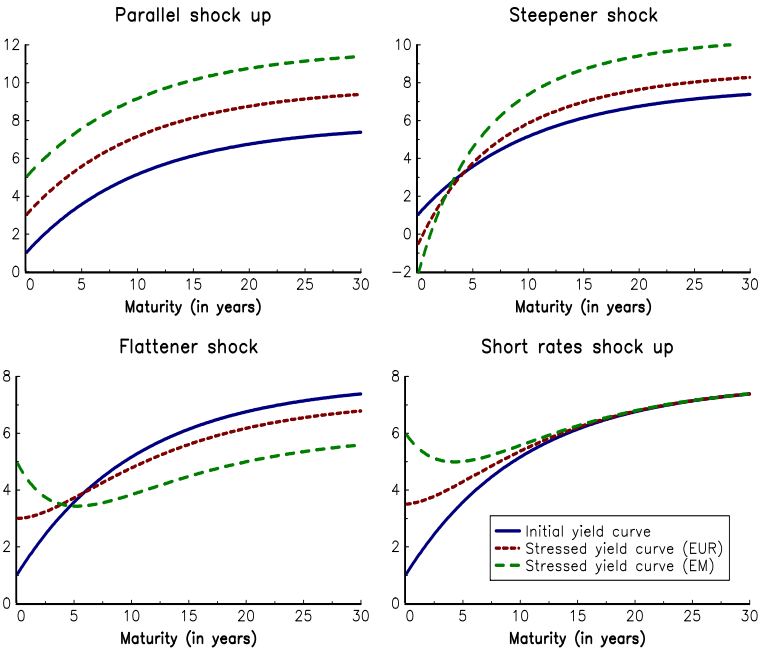


FIGURE 7.11: Stressed yield curve (in %)

TABLE 7.12: Cap on core deposits and maximum average maturity

Category	Cap ω^+	Cap t^+
Retail transactional	90%	5.0
Retail non-transactional	70%	4.5
Wholesale	50%	4.0

Behavioral options of retail customers This category mainly concerns fixed rate loans because of the prepayment risk, and fixed-term deposits because of the early redemption risk. The Basel Committee proposes to use a two-step procedure. First, the bank determine the baseline estimate of each category given the current yield curve. Then, the baseline estimate is multiplied according to the standardized interest rate scenarios. In the case of fixed rate loans subject to prepayment risk, the bank establishes the different homogenous prepayment categories. For each category, the bank estimates the baseline conditional prepayment rate CPR_0 and calculates the stressed conditional prepayment rate as follows:

$$\text{CPR}_s = \min\left(1, \gamma_s \cdot \text{CPR}_0\right)$$

where γ_s is the multiplier for the scenario s. The coefficient γ_s takes two values:

- $\gamma_s = 0.8$ for the scenarios 1, 3 and 5 (parallel up, steepener and short rates up);
- $\gamma_s = 1.2$ for the scenarios 2, 4 and 6 (parallel down, flattener and short rates down).

The cash flow for the time bucket t_k is the sum of two components:

$$\text{CF}_s\left(t_k\right) = \text{CF}_s^1\left(t_k\right) + \text{CF}_s^2\left(t_k\right)$$

where $\text{CF}_s^1\left(t_k\right)$ refers to the scheduled interest and principal repayment (without prepayment) and $\text{CF}_s^2\left(t_k\right)$ refers to the prepayment cash flow:

$$\text{CF}_s^2\left(t_k\right) = \text{CPR}_s \cdot N_s\left(t_{k-1}\right)$$

where $N_s\left(t_{k-1}\right)$ is the notional outstanding at time bucket t_{k-1} calculated with the stress scenario s.

The methodology for term deposits subject to early redemption risk is similar to the one of the fixed rate loans subject to prepayment risk. First, the bank estimates the baseline term deposit redemption ratio TDRR_0 for each homogeneous portfolio. Second, the stressed term deposit redemption ratio is equal to:

$$\text{TDRR}_s = \min\left(1, \gamma_s \cdot \text{TDRR}_0\right)$$

where γ_s is the multiplier for the scenario s. The coefficient γ_s takes two values:

- $\gamma_s = 1.2$ for the scenarios 1, 4 and 5 (parallel up, flattener and short rates up);
- $\gamma_s = 0.8$ for the scenarios 2, 3 and 6 (parallel down, steepener and short rates down).

Third, the term deposits which are expected to be redeemed early are slotted into the overnight time bucket, implying that the corresponding cash flows are given by:

$$\text{CF}_s\left(t_1\right) = \text{TDRR}_s \cdot N_0$$

where N_0 is the outstanding amount of term deposits.

Remark 76 *Fixed rate loans subject to prepayment risk and term deposits subject to early redemption risk follow the same methodology, but with two main differences. The first one concerns the impact of the stress scenario on the stress ratios* CPR_s *and* TDRR_s*. In the case of prepayment risk, the conditional prepayment rate generally increases when interest rates are falling and decreases when interest rates are rising. This is why we have* $\mathrm{CPR}_s > \mathrm{CPR}_0$ *for the scenarios where interest rates or the slope of the yield curve decrease (scenarios 1, 3 and 5). In the case of early redemption risk, the term deposit redemption ratio mainly depends on the short term interest rates. In particular, the ratio* TDRR_s *must increase when short rates increase, because this creates an incentive to negotiate a term deposit contract with a higher interest rate.*

Automatic interest rate options The computation of the automatic interest rate option risk KAO_s is given by:

$$\mathrm{KAO}_s = \sum_{i \in \mathcal{S}} \Delta\,\mathrm{FVAO}_{s,i} - \sum_{i \in \mathcal{B}} \Delta\,\mathrm{FVAO}_{s,i}$$

where:

- $i \in \mathcal{S}$ denotes an automatic interest rate option which is sold by the bank;
- $i \in \mathcal{B}$ denotes an automatic interest rate option which is bought by the bank;
- $\mathrm{FVAO}_{0,i}$ is the fair value of the automatic option i given the current yield curve and the current implied volatility surface;
- $\mathrm{FVAO}_{s,i}$ is the fair value of the automatic option i given the stressed yield curve and a relative increase in the implied volatility of 25%;
- $\Delta\,\mathrm{FVAO}_{s,i}$ is the change in the fair value of the option:

$$\Delta\,\mathrm{FVAO}_{s,i} = \mathrm{FVAO}_{s,i} - \mathrm{FVAO}_{0,i}$$

An example We consider a simplified USD-denominated balance sheet. The assets are composed of loans with the following cash flow slotting:

Instruments	Loans	Loans	Loans
Maturity	1Y	5Y	13Y
Cash flows	200	700	200

The loans are then slotted into three main buckets (short-term, medium-term and long-term loans). The average maturity is respectively equal to one-year, five-year and thirteen-year. The liabilities are composed of retail deposit accounts, term deposits, debt and tier-one equity capital. The cash flow slotting is given below:

Instruments	Non-core deposits	Term deposits	Core deposits	Debt		Equity capital
				ST	LT	
Maturity	O/N	7M	3Y	4Y	8Y	
Cash flows	100	50	450	100	100	200

The non-maturity deposits are split into non-core and core deposits. The maturity of non-core deposits is assumed to be overnight (O/N), whereas the estimated maturity of core deposits is around three years. We also have two debt instruments: one with a remaining

TABLE 7.13: Economic value of the assets

Bucket	t_k	$\mathrm{CF}_0(t_k)$	$R_0(t_k)$	$\mathrm{EV}_0(t_k)$
6	0.875	200	1.55%	197.31
11	4.50	700	3.37%	601.53
17	12.50	100	5.71%	48.98
EV_0				847.82

TABLE 7.14: Economic value of the pure liabilities

Bucket	t_k	$\mathrm{CF}_0(t_k)$	$R_0(t_k)$	$\mathrm{EV}_0(t_k)$
1	0.0028	100	1.00%	100.00
5	0.625	50	1.39%	49.57
9	2.50	450	2.44%	423.35
10	3.50	100	2.93%	90.26
14	7.50	100	4.46%	71.56
EV_0				734.73

maturity of four years and another one with a remaining maturity of eight years. The term deposits are slotted in a single bucket corresponding to a seven-month maturity.

We assume that the current yield curve is given by the Nelson-Siegel model with $\theta_1 = 8\%$, $\theta_2 = -7\%$, $\theta_3 = 6\%$ and $\theta_4 = 10$. In Table 7.13, we have reported the current economic value of the assets. It is respectively equal to 197.31, 601.53 and 48.98 for the three buckets and 847.82 for the total of assets. We have done the same exercise for the pure liabilities (Table 7.14). We obtain an economic value equal to 734.73. We deduce that the current economic value of equity is $\mathrm{EVE}_0 = 847.82 - 734.73 = 113.09$. Since the balance sheet is expressed in USD, we use the USD shocks for the interest rates scenarios: $\mathbb{S}_0 = 200$ bps, $\mathbb{S}_1 = 300$ bps and $\mathbb{S}_2 = 150$ bps. In Table 7.15, we have reported the stressed values of interest rates $R_s(t_k)$ and economic value $\mathrm{EV}_s(t_k)$ for every bucket of the balance sheet. By computing the stressed economic value of assets and pure liabilities, we deduce the stressed economic value of equity. For instance, in the case of the first stress scenario, we have $\mathrm{EVE}_1 = 781.79 - 697.39 = 84.41$. It follows that the economic value of equity will be reduced if the standardized parallel shock up occurs: $\Delta\,\mathrm{EVE}_1 = 113.10 - 84.41 = 28.69$. We observe that the economic value of equity decreases for scenarios 1, 3 and 5, and increases for scenarios 2, 4 and 6. Finally, we deduce that the risk measure $\mathcal{R}(\mathrm{EVE}) = \max_s(\Delta\,\mathrm{EVE}_s, 0) = 28.69$ represents 14.3% of the equity. This puts under the threshold 15% of the materiality test.

7.1.4 Other ALM risks

Even if liquidity and interest rate risks are the main ALM risks, there are other risks that impact the banking book of the balance sheet, in particular currency risk and credit spread risk.

7.1.4.1 Currency risk

We recall that the standardized approach for implementing IRRBB considers each currency separately. Indeed, the risk measures $\Delta\,\mathrm{EVE}_{s,c}$ and $\mathrm{KAO}_{s,c}$ are calculated for each interest rate scenario s and each currency c. Then, the aggregated value $\sum_c (\Delta\,\mathrm{EVE}_{s,c} + \mathrm{KAO}_{s,c})^+$ is calculated across the different currencies and the maximum is selected for the global risk measure of the bank.

TABLE 7.15: Stressed economic value of equity

Bucket	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$
			Assets			
$R_s(t_6)$	3.55%	−0.45%	0.24%	3.30%	3.96%	−0.87%
$R_s(t_{11})$	5.37%	1.37%	3.65%	3.54%	4.34%	2.40%
$R_s(t_{17})$	7.71%	3.71%	6.92%	4.96%	5.84%	5.58%
$\text{EV}_s(t_6)$	193.89	200.80	199.57	194.31	193.20	201.52
$\text{EV}_s(t_{11})$	549.76	658.18	594.03	596.91	575.74	628.48
$\text{EV}_s(t_{17})$	38.15	62.89	42.13	53.83	48.18	49.79
EV_s	781.79	921.87	835.74	845.05	817.11	879.79
			Pure liabilities			
$R_s(t_1)$	3.00%	−1.00%	−0.95%	3.40%	4.00%	−2.00%
$R_s(t_5)$	3.39%	−0.61%	−0.08%	3.32%	3.96%	−1.17%
$R_s(t_9)$	4.44%	0.44%	2.03%	3.31%	4.05%	0.84%
$R_s(t_{10})$	4.93%	0.93%	2.90%	3.40%	4.18%	1.68%
$R_s(t_{14})$	6.46%	2.46%	5.31%	4.07%	4.92%	4.00%
$\text{EV}_s(t_1)$	99.99	100.00	100.00	99.99	99.99	100.01
$\text{EV}_s(t_5)$	48.95	50.19	50.02	48.97	48.78	50.37
$\text{EV}_s(t_9)$	402.70	445.05	427.77	414.27	406.69	440.69
$\text{EV}_s(t_{10})$	84.16	96.80	90.34	88.77	86.39	94.30
$\text{EV}_s(t_{14})$	61.59	83.14	67.17	73.70	69.13	74.07
EV_s	697.39	775.18	735.31	725.71	710.98	759.43
			Equity			
EVE_s	84.41	146.68	100.43	119.34	106.13	120.37
$\Delta\,\text{EVE}_s$	28.69	−33.58	12.67	−6.24	6.97	−7.27

One of the issues concerns currency hedging. Generally, it is done by rolling reverse FX forward contracts, implying that the hedging cost is approximatively equal to $i^\star - i$, where i is the domestic interest rate and $i^\star$ is the foreign interest rate. This relationship comes from the covered interest rate parity (CIP). We deduce that the hedging cost can be large when $i^\star \gg i$. This has been particularly true for European and Japanese banks, because these regions have experienced some periods of low interest rates. The question of full hedging, partial hedging or no hedging has then been readdressed after the 2008 Global Financial Crisis. Most of banks continue to fully hedge the banking book including the equity capital, but it is not obvious that it is optimal. Another issue has concerned the access to dollar funding of non-US banks. Traditionally, "*their branches and subsidiaries in the United States were a major source of dollar funding, but the role of these affiliates has declined*" (Aldasoro and Ehlers, 2018, page 15). Today, we notice that a lot of non-US banks issue many USD-denominated debt instruments in order to access dollar funding[34]. Banks must now manage a complex multi-currency balance sheet, implying that currency management has become an important topic in ALM.

7.1.4.2 Credit spread risk

According to BCBS (2016d), credit spread risk in the banking book (CSRBB) is driven "*by changes in market perception about the credit quality of groups of different credit-risky instruments, either because of changes to expected default levels or because of changes to*

[34]See for instance annual reports of European and Japanese banks.

market liquidity". In Figure 7.12, we have reproduced the scheme provided by the Basel Committee in order to distinguish IRRBB and CSRBB. Therefore, CSRBB can be seen as the ALM spread risk of credit-risky instruments which is not explained by IRRBB and idiosyncratic credit risk. However, the definition provided by the Basel Committee is too broad, and does not avoid double counting with credit and jump-to-default risk[35]. At of the date of the publication of this book, the debate on CSRBB is far from finished, even if CSRBB must be monitored and assessed since 2018.

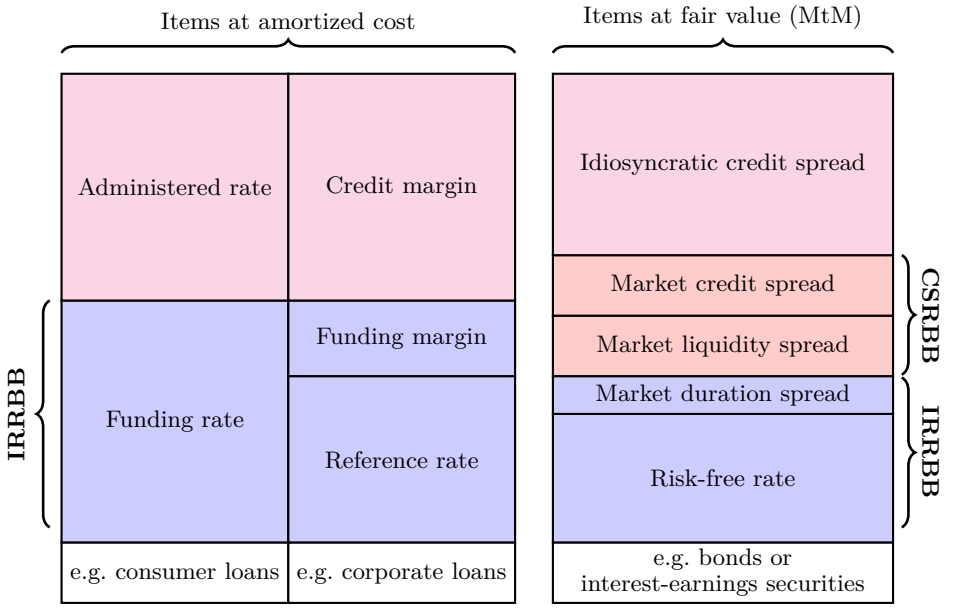


FIGURE 7.12: Components of interest rates

Source: BCBS (2016d, page 34).

7.2 Interest rate risk

In this section, we focus on the ALM tools that are related to the interest rate risk in the banking book. We first introduce the concept of duration gap and show how it is related to the economic value risk of the banking book. Then, we present the different ways to calculate earnings-at-risk (EaR) measures and focus more particularly on the net interest income sensitivity and the interest rate hedging strategies. The third part is dedicated to funds transfer pricing, whose objective is to centralize all interest rate risks, to manage them and to allocate profit between the different business units. Finally, we present an econometric model for simulating and evaluating interest rate scenarios.

[35] See for example the position of the European Banking Federation (EBF): www.ebf.eu/regulation-supervision/credit-spread-risk-in-the-banking-book-ebf-position.

7.2.1 Duration gap analysis

In this section, we focus on the duration gap analysis, which is an approximation of the repricing gap analysis we have previously presented. The idea is to obtain an estimation of $\Delta\,\text{EVE}$. Although this approach is only valid in the case of parallel interest rate shocks[36], it is an interesting method because we obtain closed-form formulas. In the case of non-parallel interest rate scenarios or if we want to obtain more accurate results, it is better to implement the repricing gap analysis, which consists in computing the stressed economic value of assets and liabilities in order to deduce the impact on the economic value of equity.

7.2.1.1 Relationship between Macaulay duration and modified duration

We consider a financial asset, whose price is given by the present value of cash flows:

$$V = \sum_{t_k \geq t} B\left(t, t_k\right) \cdot \text{CF}\left(t_k\right)$$

where $\text{CF}\left(t_k\right)$ is the cash flow at time t_k and $B\left(t, t_k\right)$ is the associated discount factor. The Macaulay duration $\mathcal{D}$ is the weighted average of the cash flow maturities:

$$\mathcal{D} = \sum_{t_k \geq t} w\left(t, t_k\right) \cdot \left(t_k - t\right)$$

where $w\left(t, t_k\right)$ is the weight associated to the cash flow at time t_k:

$$w\left(t, t_k\right) = \frac{B\left(t, t_k\right) \cdot \text{CF}\left(t_k\right)}{V}$$

In the case of a zero-coupon bond whose maturity date is T, the Macaulay duration is equal to the remaining maturity $T - t$.

Let us define the yield to maturity y as the solution of the following equation:

$$V = \sum_{t_k \geq t} \frac{\text{CF}\left(t_k\right)}{\left(1 + y\right)^{\left(t_k - t\right)}}$$

We have:

$$\begin{aligned} \frac{\partial V}{\partial y} &= \sum_{t_k \geq t} -\left(t_k - t\right) \cdot \left(1 + y\right)^{-\left(t_k - t\right) - 1} \cdot \text{CF}\left(t_k\right) \\ &= -\frac{\mathcal{D}}{\left(1 + y\right)} \cdot V \\ &= -\mathfrak{D} \cdot V \end{aligned}$$

where $\mathfrak{D}$ is the modified duration:

$$\mathfrak{D} = \frac{\mathcal{D}}{1 + y}$$

We deduce that the modified duration is the price sensitivity measure:

$$\mathfrak{D} = \frac{1}{V} \cdot \frac{\partial V}{\partial y} = -\frac{\partial \ln V}{\partial y}$$

[36] The duration gap analysis can be viewed as the first-order approximation of the repricing gap analysis.

If the yield to maturity is low, we have $\mathfrak{D} \approx \mathcal{D}$. Since the Macaulay duration is easier to interpret, the modified duration is more relevant to understand the impact of an interest rate stress scenario. Indeed, we have:

$$\Delta V \approx -\mathfrak{D} \cdot V \cdot \Delta y$$

Nevertheless, we can use the following alternative formula to evaluate the impact of an interest rate parallel shift:

$$\Delta V \approx -\mathcal{D} \cdot V \cdot \frac{\Delta y}{1+y}$$

Remark 77 *Using a continuous-time framework, the yield to maturity is defined as the root of the following equation:*

$$V = \sum_{t_k \geq t} e^{-y(t_k - t)} \cdot \mathrm{CF}\left(t_k\right)$$

We deduce that:

$$\begin{aligned} \frac{\partial V}{\partial y} &= \sum_{t_k \geq t} -\left(t_k - t\right) \cdot e^{-y(t_k - t)} \cdot \mathrm{CF}\left(t_k\right) \\ &= -\mathcal{D} \cdot V \end{aligned}$$

It follows that the modified duration $\mathfrak{D}$ is defined as the Macaulay duration $\mathcal{D}$ in continuous-time modeling.

Example 72 *We consider the following cash flows stream $\{t_k, \mathrm{CF}\left(t_k\right)\}$:*

t_k	1	4	7	11
$\mathrm{CF}\left(t_k\right)$	200	500	200	100

The current zero-coupon interest rates are: $R\left(1\right) = 2\%$, $R\left(4\right) = 3\%$, $R\left(7\right) = 4\%$, and $R\left(11\right) = 5\%$.

If we consider the discrete-time modeling framework, we obtain $V = 850.77$, $y = 3.61\%$, $\mathcal{D} = 4.427$ and $\mathfrak{D} = 4.273$. A parallel shock of $+1\%$ decreases the economic value since we obtain $V\left(R + \Delta R\right) = 816.69$. It follows that $\Delta V = -34.38$. Using the duration-based approximation, we have[37]:

$$\begin{aligned} \Delta V &\approx -\mathfrak{D} \cdot V \cdot \Delta R \\ &= -4.273 \times 850.77 \times 1\% \\ &= -36.35 \end{aligned}$$

In the case of the continuous-time modeling framework, the results become $V = 848.35$, $y = 3.61\%$ and $\mathcal{D} = \mathfrak{D} = 4.422$. If we consider a parallel shock of $+1\%$, the exact value of ΔV is equal to -35.37, whereas the approximated value is equal to -37.51. In Table 7.16, we also report the results for a parallel shock of -1%. Moreover, we indicate the case where we stress the yield to maturity and not the yield curve. Because $V\left(y + \Delta R\right) \neq V\left(R + \Delta R\right)$, we observe a small difference between the approximation and the true value of ΔV.

[37] This approximation is based on the assumption that the yield curve is flat. However, numerical experiments show that it is also valid when the term structure of interest rates is increasing or decreasing.

TABLE 7.16: Impact of a parallel shift of the yield curve

	Discrete-time		Continuous-time	
ΔR	$+1\%$	-1%	$+1\%$	-1%
$V(R+\Delta R)$	816.69	887.52	812.78	886.09
ΔV	-34.38	36.75	-35.57	37.74
$V(y+\Delta R)$	815.64	888.42	811.94	887.02
ΔV	-35.13	37.64	-36.41	38.67
Approximation	-36.35	36.35	-37.51	37.51

Remark 78 *From a theoretical point of view, duration analysis is valid under the assumption that the term structure of interest rates is flat and the change in interest rates is a parallel shift. This framework can be extended by considering the convexity:*

$$\mathfrak{C} = \frac{1}{V} \cdot \frac{\partial^2 V}{\partial y^2}$$

In this case, we obtain the following second-order approximation:

$$\Delta V \approx -\mathfrak{D} \cdot V \cdot \Delta y + \frac{1}{2}\mathfrak{C} \cdot V \cdot (\Delta y)^2$$

7.2.1.2 Relationship between the duration gap and the equity duration

Let V_j and $\mathcal{D}_j$ be the market value and the Macaulay duration associated to the j^{th} cash flow stream. Then, the market value of a portfolio that is composed of m cash flow streams is equal to the sum of individual market values:

$$V = \sum_{j=1}^{m} V_j$$

while the duration of the portfolio is the average of individual durations:

$$\mathcal{D} = \sum_{j=1}^{m} w_j \cdot \mathcal{D}_j$$

where:

$$w_j = \frac{V_j}{V}$$

This result is obtained by considering a common yield to maturity.

We recall that $E(t) = A(t) - L^\star(t)$ and $\text{EV}_E = \text{EV}_A - \text{EV}_{L^\star}$. Using the previous result, we deduce that the duration of equity is equal to:

$$\begin{aligned} \mathcal{D}_E &= \frac{\text{EV}_A}{\text{EV}_A - \text{EV}_{L^\star}} \cdot \mathcal{D}_A - \frac{\text{EV}_{L^\star}}{\text{EV}_A - \text{EV}_{L^\star}} \cdot \mathcal{D}_{L^\star} \\ &= \frac{\text{EV}_A}{\text{EV}_A - \text{EV}_{L^\star}} \cdot \mathcal{D}_{\mathcal{G}ap} \end{aligned} \tag{7.9}$$

where the duration gap (also called DGAP) is defined as the difference between the duration of assets and the duration of pure liabilities scaled by the ratio $\text{EV}_{L^\star} / \text{EV}_A$:

$$\mathcal{D}_{\mathcal{G}ap} = \mathcal{D}_A - \frac{\text{EV}_{L^\star}}{\text{EV}_A} \cdot \mathcal{D}_{L^\star} \tag{7.10}$$

Another expression of the equity duration is:

$$\mathcal{D}_E = \frac{\text{EV}_A}{\text{EV}_E} \cdot \mathcal{D}_{\mathcal{G}ap} = \mathcal{L}_{A/E} \cdot \frac{\mathcal{D}_{\mathcal{G}ap}}{\mathcal{L}_{E/A}} \tag{7.11}$$

We notice that the equity duration is equal to the duration gap multiplied by the leverage ratio, where $\mathcal{L}_{A/E}$ is the ratio between the economic value of assets and the economic value of equity.

By definition of the modified duration, we have[38]:

$$\begin{aligned} \Delta \text{EVE} &= \Delta \text{EV}_E \\ &\approx -\mathfrak{D}_E \cdot \text{EV}_E \cdot \Delta y \\ &= -\mathcal{D}_E \cdot \text{EV}_E \cdot \frac{\Delta y}{1+y} \end{aligned}$$

Using Equation (7.11), we deduce that:

$$\Delta \text{EVE} \approx -\mathcal{D}_{\mathcal{G}ap} \cdot \text{EV}_A \cdot \frac{\Delta y}{1+y} \tag{7.12}$$

Formulas (7.10) and (7.12) are well-known and are presented in many handbooks of risk management (Crouhy *et al.*, 2013; Bessis, 2015).

7.2.1.3 An illustration

We consider the following balance sheet:

Assets	V_j	$\mathcal{D}_j$	Liabilities	V_j	$\mathcal{D}_j$
Cash	5	0.0	Deposits	40	3.2
Loans	40	1.5	CDs	20	0.8
Mortgages	40	6.0	Debt	30	1.7
Securities	15	3.8	Equity capital	10	
Total	100		Total	100	

We have $\text{EV}_A = 100$, $\text{EV}_{L^\star} = 90$ and $\text{EV}_E = 10$. We deduce that the leverage is equal to:

$$\mathcal{L}_{A/E} = \frac{\text{EV}_A}{\text{EV}_E} = \frac{100}{10} = 10$$

The duration of assets is equal to:

$$\mathcal{D}_A = \frac{5}{100} \times 0 + \frac{40}{100} \times 1.5 + \frac{40}{100} \times 6.0 + \frac{15}{100} \times 3.8 = 3.57 \text{ years}$$

For the pure liabilities, we obtain:

$$\mathcal{D}_{L^\star} = \frac{40}{90} \times 3.2 + \frac{20}{90} \times 0.8 + \frac{30}{90} \times 1.7 = 2.17 \text{ years}$$

It follows that the duration gap is equal to:

$$\mathcal{D}_{\mathcal{G}ap} = 3.57 - \frac{90}{100} \times 2.17 = 1.62 \text{ years}$$

[38] We recall that EVE is an alternative expression for designating EV_E.

while the value $\mathcal{D}_E$ of the equity duration is 16.20 years. Since $\mathcal{D}_{\mathcal{G}ap}$ is equal to 1.62 years, the average duration of assets exceeds the average duration of liabilities. This is generally the normal situation because of the bank's liquidity transformation (borrowing short and lending long). In the table below, we have reported the impact of an interest rate shift on the economic value of equity when the yield to maturity is equal to 3%:

Δy	-2%	-1%	$+1\%$	$+2\%$
$\Delta\,\text{EVE}$	3.15	1.57	-1.57	-3.15
$\dfrac{\Delta\,\text{EVE}}{\text{EVE}}$	31.46%	15.73%	-15.73%	-31.46%

Since the duration gap is positive, the economic value of equity decreases when interest rates increase, because assets will fall more than liabilities. For instance, an interest rate rise of 1% induces a negative variation of 1.57 in EVE. This impact is large and represents a relative variation of -15.73%.

7.2.1.4 Immunization of the balance sheet

In order to reduce the sensitivity of the bank balance sheet to interest rate changes, we have to reduce the value of $|\Delta\,\text{EVE}|$. Using Equation (7.12), this is equivalent to control the value of the duration gap. In particular, a full immunization implies that:

$$\begin{aligned}\Delta\,\text{EVE} = 0 &\Leftrightarrow \mathcal{D}_{\mathcal{G}ap} = 0 \\ &\Leftrightarrow \mathcal{D}_A - \frac{\text{EV}_{L^\star}}{\text{EV}_A}\cdot\mathcal{D}_{L^\star} = 0 \end{aligned} \tag{7.13}$$

If we consider the normal situation where the duration gap is positive, we have three solutions:

1. we can reduce the duration of assets;
2. we can increase the relative weight of the liabilities with respect to the assets;
3. we can increase the duration of liabilities.

Generally, it takes time to implement the first two solutions. For instance, reducing the duration of assets implies redefining the business model by reducing the average maturity of loans. It can be done by decreasing the part of mortgages and increasing the part of short-term loans (e.g. consumer credit or credit cards). In fact, the third solution is the easiest way to immunize the bank balance sheet to interest rate changes. For example, the bank can issue a long-term debt instrument. Therefore, hedging the balance sheet involves managing the borrowing program of the bank.

Let us consider the previous example. We found $\mathcal{D}_A = 3.57$ and $\dfrac{\text{EV}_{L^\star}}{\text{EV}_A} = \dfrac{90}{100}$. It follows that the optimal value of the liability duration must be equal to 3.97 years:

$$\mathcal{D}_{\mathcal{G}ap} = 0 \Leftrightarrow \mathcal{D}_{L^\star} = \frac{100}{90}\times 3.57 = 3.97 \text{ years}$$

We assume that the bank issues a 10-year zero-coupon bond by reducing its current debt amount. The notional of the zero-coupon bond must then satisfy this equation:

$$\frac{40}{90}\times 3.2 + \frac{20}{90}\times 0.8 + \frac{30-N}{90}\times 1.7 + \frac{N}{90}\times 10 = 3.97$$

or:

$$N = \frac{3.97\times 90 - (40\times 3.2 + 20\times 0.8 + 30\times 1.7)}{10 - 1.7} = 19.52$$

TABLE 7.17: Bank balance sheet after immunization of the duration gap

Assets	V_j	$\mathcal{D}_j$	Liabilities	V_j	$\mathcal{D}_j$
Cash	5	0.0	Deposits	40	3.2
Loans	40	1.5	CDs	20	0.8
Mortgages	40	6.0	Debt	10.48	1.7
Securities	15	3.8	Zero-coupon bond	19.52	10.0
			Equity capital	10	0.0
Total	100		Total	100	

After immunization, the duration of equity is equal to zero and we obtain the balance sheet given in Table 7.17.

Remark 79 *The duration gap analysis covers the gap risk, which is the first-order source of interest rate risk. It is not adapted for measuring basis and option risks. For these two risks, we need to use the repricing analysis.*

7.2.2 Earnings-at-risk

Earnings-at-risk assesses potential future losses due to a change in interest rates over a specified period. Several measures of earnings can be used: accounting earnings, interest margins, commercial margins, etc. For interest rate scenarios, we can use predefined[39], historical or Monte Carlo scenarios. Once earnings distributions are obtained, we can analyze the results for each scenario, derive the most severe scenarios, compute a value-at-risk, etc. In this section, we first focus on the income gap analysis, which is the equivalent of the duration gap analysis when analyzing interest rate income risks. Then we present the tools for calculating the net interest income (NII). Finally, we consider hedging strategies in the context where both Δ EVE and NII risk measures are managed.

7.2.2.1 Income gap analysis

Definition of the gap Since Δ EVE measures the price risk of the balance sheet, Δ NII measures the earnings risk of the income statement. It refers to the risk of changes in the interest rates on assets and liabilities from the point of view of the net income. Indeed, if interest rates change, this induces a gap (or repricing) risk because the bank will have to reinvest assets and refinance liabilities at a different interest rate level in the future. The gap is defined as the difference between rate sensitive assets (RSA) and rate sensitive liabilities (RSL):

$$\text{GAP}(t,u) = \text{RSA}(t,u) - \text{RSL}(t,u) \tag{7.14}$$

where t is the current date and u is the time horizon of the gap[40]. While Δ EVE considers all the cash flows, Δ NII is generally calculated using a short-term time horizon, for example the next quarter or the next year. Therefore, rate sensitive assets/liabilities correspond to assets/liabilities that will mature or reprice before the time horizon of the gap. This is why the interest rate gap risk is also called the repricing risk or the reset risk.

In order to calculate the interest rate gap, the bank must decide which items are rate sensitive. This includes two main categories. The first one corresponds to items that mature

[39] Such as the six scenarios of the standardized IRRBB approach.
[40] This means that $h = u - t$ is the maturity of the gap.

before the time horizon $t+h$, whereas the second one corresponds to floating rate instruments. For example, consider the following balance sheet expressed in millions of dollars:

Assets	Amount	Liabilities	Amount
Loans		Deposits	
Less than 1 year	200	Non-maturity deposits	150
1 to 2 years	100	Money market deposits	250
Greater than 2 years	100	Term deposits	
Mortgages		Fixed rate	250
Fixed rate	100	Variable rate	100
Variable rate	350	Borrowings	
Securities		Less than 1 year	50
Fixed rate	50	Greater than 1 year	100
Physical assets	100	Capital	100
Total	1000	Total	1000

If the time horizon of the gap is set to one year, the rate sensitive assets are loans with maturities of less than one year (200) and variable rate mortgages (350), while the rate sensitive liabilities are money market deposits (250), variable rate term deposits (100) and borrowings with maturities of less than one year (50). Therefore, we can split the balance sheet between rate sensitive, fixed rate and non-earning assets:

Assets	Amount	Liabilities	Amount
Rate sensitive	550	Rate sensitive	400
Fixed rate	350	Fixed rate	600
Non-earning	100	Non-earning	100

We deduce that the one-year gap is equal to \$150 million:

$$\text{GAP}(t, t+1) = 550 - 400 = 150$$

Approximation of Δ NII We consider the following definition of the net interest income:

$$\begin{aligned}\text{NII}(t,u) &= \text{RSA}(t,u)\cdot R_{\text{RSA}}(t,u) + \text{NRSA}(t,u)\cdot R_{\text{NRSA}}(t,u) - \\ &\quad \text{RSL}(t,u)\cdot R_{\text{RSL}}(t,u) - \text{NRSL}(t,u)\cdot R_{\text{NRSL}}(t,u)\end{aligned}$$

where RNSA and RNSL denote assets and liabilities that are not rate sensitive and $R_{\mathcal{C}}(t,u)$ is the average interest rate for the category $\mathcal{C}$ and the maturity date u. We have:

$$\Delta\,\text{NII}(t,u) = \text{NII}(t+h,u+h) - \text{NII}(t,u)$$

By considering a static gap[41], we deduce that:

$$\begin{aligned}\Delta\,\text{NII}(t,u) &= \text{RSA}(t,u)\cdot(R_{\text{RSA}}(t+h,u+h) - R_{\text{RSA}}(t,u)) + \\ &\quad \text{NRSA}(t,u)\cdot(R_{\text{NRSA}}(t+h,u+h) - R_{\text{NRSA}}(t,u)) - \\ &\quad \text{RSL}(t,u)\cdot(R_{\text{RSL}}(t+h,u+h) - R_{\text{RSL}}(t,u)) - \\ &\quad \text{NRSL}(t,u)\cdot(R_{\text{NRSL}}(t+h,u+h) - R_{\text{NRSL}}(t,u))\end{aligned}$$

Since interest income and interest expense do not change for fixed rate assets and liabilities between t and $t+h$ — $R_{\text{NRSA}}(t+h,u+h) = R_{\text{NRSA}}(t,u)$ and $R_{\text{NRSL}}(t+h,u+h) - R_{\text{NRSL}}(t,u)$, we have:

$$\Delta\,\text{NII}(t,u) = \text{RSA}(t,u)\cdot\Delta R_{\text{RSA}}(t,u) - \text{RSL}(t,u)\cdot\Delta R_{\text{RSL}}(t,u)$$

[41] This means that $\text{RSA}(t+h,u+h) = \text{RSA}(t,u)$, $\text{NRSA}(t+h,u+h) = \text{NRSA}(t,u)$, $\text{RSL}(t+h,u+h) = \text{RSL}(t,u)$ and $\text{NRSL}(t+h,u+h) = \text{NRSL}(t,u)$ where $h = u - t$.

By assuming that the impact of interest rate changes is the same for rate sensitive assets and liabilities, we finally obtain:

$$\Delta \operatorname{NII}(t,u) \approx \operatorname{GAP}(t,u) \cdot \Delta R \tag{7.15}$$

where ΔR is the parallel shock of interest rates. Income gap analysis is then described by Equations (7.14) and (7.15).

For instance, if we consider the previous example, the one-year gap is equal to \$150 million and we have the following impact on the income:

ΔR	-2%	-1%	0%	$+1\%$	$+2\%$
Δ NII	$-\$3$ mn	$-\$1.5$ mn	0	$+\$1.5$ mn	$+\$3$ mn

If interest rates rise by 2%, the bank expects that its income increases by \$3 million. On the contrary, the loss can be equal to \$3 million if interest rates fall by 2%.

Remark 80 *The previous analysis is valid for a given maturity $h = u - t$. For example, $\Delta \operatorname{NII}(t, t+0.25)$ measures the impact for the next three months while $\Delta \operatorname{NII}(t, t+1)$ measures the impact for the next year. It is common to consider the change in income for a given time period $[u_1, u_2[$ where $u_1 = t + h_1$ and $u_2 = t + h_2$. We notice that:*

$$\begin{aligned}
\Delta \operatorname{NII}(t, u_1, u_2) &= \Delta \operatorname{NII}(t, u_2) - \Delta \operatorname{NII}(t, u_1) \\
&= (\operatorname{GAP}(t, u_2) - \operatorname{GAP}(t, u_1)) \cdot \Delta R \\
&= \operatorname{GAP}(t, u_1, u_2) \cdot \Delta R \\
&= (\operatorname{RSA}(t, u_1, u_2) - \operatorname{RSL}(t, u_1, u_2)) \cdot \Delta R
\end{aligned}$$

where $\operatorname{GAP}(t, u_1, u_2)$, $\operatorname{RSA}(t, u_1, u_2)$ and $\operatorname{RSL}(t, u_1, u_2)$ are respectively the static gap, rate sensitive assets and rate sensitive liabilities for the period $[u_1, u_2[$.

7.2.2.2 Net interest income

Definition We recall that the net interest income of the bank is the difference between interest rate revenues of its assets and interest rate expenses of its liabilities:

$$\operatorname{NII}(t,u) = \sum_{i \in \mathcal{A}ssets} N_i(t,u) \cdot R_i(t,u) - \sum_{j \in \mathcal{L}iabilities} N_j(t,u) \cdot R_j(t,u) \tag{7.16}$$

where $\operatorname{NII}(t,u)$ is the net interest income at time t for the maturity date u, $N_i(t,u)$ is the notional outstanding at time u for the instrument i and $R_i(t,u)$ is the associated interest rate. This formula is similar to the approximated equation presented above, but it is based on a full repricing model. However, this formula is static and assumes a run-off balance sheet. In order to be more realistic, we can assume a dynamic balance sheet. However, the computation of the net interest income is then more complex because it requires modeling the liquidity gap and also behavioral options.

An example We consider a simplified balance sheet with the following asset and liability positions:

- The asset position is made up of two bullet loans A and B, whose remaining maturity is respectively equal to 18 months and 2 years. The outstanding notional of each loan is equal to 500. Moreover, we assume that the interest rate is equal to 6% for the first loan and 5% for the second loan.

TABLE 7.18: Interest income schedule and liquidity gap

$u-t$	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
Loan A	7.50	7.50	7.50	7.50	7.50	7.50		
Loan B	6.25	6.25	6.25	6.25	6.25	6.25	6.25	6.25
IR revenues	13.25	13.25	13.25	13.25	13.25	13.25	6.25	6.25
Debt C	6.00	6.00	6.00	6.00				
Equity	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IR expenses	6.00	6.00	6.00	6.00	0.00	0.00	0.00	0.00
NII (t,u)	7.25	7.25	7.25	7.25	13.25	13.25	6.25	6.25
$\mathcal{LG}(t,u)$	0	0	0	0	-800	-800	-300	-300

- The liability position is made up of a bullet debt instrument C, whose remaining maturity is 1 year and outstanding notional is 800. We assume that the interest rate is equal to 3%.
- The equity capital is equal to 200.

To calculate the net interest income, we calculate the interest rate revenues and costs. By assuming a quarterly pricing, the quarterly income of the instruments are:

$$\begin{aligned} I_A &= \frac{1}{4} \times 6\% \times 500 = 7.50 \\ I_B &= \frac{1}{4} \times 5\% \times 500 = 6.25 \\ I_C &= \frac{1}{4} \times 3\% \times 800 = 6.00 \end{aligned}$$

We obtain the interest income schedule given in Table 7.18. However, calculating the net interest income as the simple difference between interest rate revenues and expenses ignores the fact that the balance sheet is unbalanced. In the last row in Table 7.18, we have reported the liquidity gap. At time $u = t + 1.25$, the value of the liabilities is equal to 200 because the borrowing has matured. It follows that the liquidity gap is equal to -800. At time $u = t + 1.75$, the loan A will mature. In this case, the liabilities is made up of the equity capital whereas the assets is made up of the loan B. We deduce that the liquidity gap is equal to $200 - 500 = -300$.

TABLE 7.19: Balance sheet under the constraint of a zero liquidity gap

	$u-t$	1.25	1.50	1.75	2.00
Approach #1	Debt D	500	500		
	Debt E	300	300	300	300
Approach #2	Loan F			500	500
	Debt G	800	800	800	800

At this stage, we can explore several approaches to model the net interest income, and impose a zero liquidity gap. In the first approach, the bank borrows 500 for the period $[t+1, t+1.50]$ and 300 for the period $[t+1, t+2]$. This corresponds to debt instruments D and E in Table 7.19. We note $\tilde{R}_L(t,u)$ the interest rate for these new liabilities. We notice that $\tilde{R}_L(t,u)$ is a random variable at time t, because it will be known at time $t+1$.

We have:

$$\begin{aligned}\text{NII}(t,u) &= 13.25 - \frac{1}{4} \times 800 \times \tilde{R}_L(t,u) \\ &= 13.25 - \frac{1}{4} \times 800 \times \left(\tilde{R}_L(t,u) - 3\%\right) - \frac{1}{4} \times 800 \times 3\% \\ &= 7.25 - 200 \times \left(\tilde{R}_L(t,u) - 3\%\right)\end{aligned}$$

for $u = t + 0.25$ and $u = t + 0.5$, and:

$$\text{NII}(t,u) = 6.25 - \frac{1}{4} \times 300 \times \tilde{R}_L(t,u) = 4.00 - 75 \times \left(\tilde{R}_L(t,u) - 3\%\right)$$

for $u = t + 1.75$ and $u = t + 2.0$.

The drawback of the previous approach is that the size of the balance sheet has been dramatically reduced for the two last dates. This situation is not realistic, because it assumes that the assets are not replaced by the new production. This is why it is better to consider that Loan A is rolled into Loan F, and the debt instrument C is replaced by the debt instrument G (see Table 7.19). In this case, we obtain:

$$\begin{aligned}\text{NII}(t,u) &= 6.25 + \frac{1}{4} \times 500 \times \tilde{R}_A(t,u) - \frac{1}{4} \times 800 \times \tilde{R}_L(t,u) \\ &= 6.25 + \frac{1}{4} \times 500 \times \left(\tilde{R}_A(t,u) - 6\%\right) + \frac{1}{4} \times 500 \times 6\% - \\ &\quad \frac{1}{4} \times 800 \times \left(\tilde{R}_L(t,u) - 3\%\right) - \frac{1}{4} \times 800 \times 3\% \\ &= 7.25 + \frac{1}{4} \times 500 \times \left(\tilde{R}_A(t,u) - 6\%\right) - \frac{1}{4} \times 800 \times \left(\tilde{R}_L(t,u) - 3\%\right)\end{aligned}$$

If we note $\Delta R_L = \tilde{R}_L(t,u) - 3\%$ and $\Delta R_A = \tilde{R}_A(t,u) - 6\%$, we obtain the following figures[42]:

ΔR_A	-2%	-1%	0%	$+1\%$	$+2\%$	-2%	-2%	-1.5%
ΔR_L	-2%	-1%	0%	$+1\%$	$+2\%$	0%	-1%	0.0%
$t + 1.00$	7.25	7.25	7.25	7.25	7.25	7.25	7.25	7.25
$t + 1.25$	11.25	9.25	7.25	5.25	3.25	7.25	9.25	7.25
$t + 1.50$	11.25	9.25	7.25	5.25	3.25	7.25	9.25	7.25
$t + 1.75$	8.75	8.00	7.25	6.50	5.75	4.75	6.75	4.13
$t + 2.00$	8.75	8.00	7.25	6.50	5.75	4.75	6.75	4.13

The case $\Delta R_L = \Delta R_A$ is equivalent to use the income gap analysis. However, this approach is simple and approximative. It does not take into account the maturity of the instruments and the dynamics of the yield curve. Let us consider a period of falling interest rates. We assume that the yield of assets is equal to the short interest rate plus 2% on average while the cost of liabilities is generally equal to the short interest rate plus 1%. On average, the bank captures a net interest margin (NIM) of 1%. This means that the market interest rate was equal to 5% for Loan A, 4% for Loan B and 2% for Debt C. We can then think that Loan A has been issued a long time ago whereas Debt C is more recent. If the interest rate environment stays at 2%, we have $\tilde{R}_A(t,u) = 4\%$ and $\tilde{R}_L(t,u) = 3\%$, which implies that $\Delta R_A = 4\% - 6\% = -2\%$ and $\Delta R_L = 3\% - 3\% = 0\%$. We obtain the results given in the seventh column. We can also explore other interest rate scenarios or other business

[42] We have $\text{NII}(t, t+1) = 7.25$, $\text{NII}(t, t+1.25) = \text{NII}(t, t+1.5) = 7.25 - 200 \times \Delta R_L$ and $\text{NII}(t, t+1.75) = \text{NII}(t, t+2) = 7.25 + 125 \times \Delta R_A - 200 \times \Delta R_L$.

scenarios. For instance, the bank may be safer than before, meaning that the spread paid to the market is lower (eight column) or the bank may have an aggressive loan issuing model, implying that the interest rate margin is reduced (ninth column).

Remark 81 *The previous analysis gives the impression that the net interest income is known for $u < t + 1.5$ and stochastic after. In fact, this is not true. Indeed, we notice that the interest rates of Loans A and B are equal to 6% and 5% whereas the current interest rates are around 2%. Therefore, we can anticipate that the bank will be subject to prepayment issues. Our analysis does not take into account the behavior of clients and the impact of embedded options in the net interest income*[43].

Mathematical formulation We reiterate that the net interest income is equal to:

$$\mathrm{NII}(t,u) = \sum_{i \in \mathcal{A}ssets} N_i(t,u) \cdot R_i(t,u) - \sum_{j \in \mathcal{L}iabilities} N_j(t,u) \cdot R_j(t,u)$$

If we consider a future date $t' > t$, we have:

$$\begin{aligned}\mathrm{NII}(t',u) &= \sum_{i \in \mathcal{A}ssets} N_i(t',u) \cdot R_i(t',u) - \sum_{j \in \mathcal{L}iabilities} N_j(t',u) \cdot R_j(t',u) - \\ &\left(\sum_{i \in \mathcal{A}ssets} N_i(t',u) - \sum_{j \in \mathcal{L}iabilities} N_j(t',u) \right) \cdot R(t',u)\end{aligned}$$

The future NII requires the projection of the new production and the forecasting of asset and liability rates (or customer rates). The third term represents the liquidity gap that must be financed or placed[44]. In what follows, we assume that the future liquidity gap is equal to zero in order to obtain tractable formulas.

Since we have the identity $\Delta \mathrm{NII}(t',u) = \mathrm{GAP}(t',u) \cdot \Delta R$, we deduce that:

$$\begin{aligned}\mathrm{GAP}(t',u) &= \frac{\Delta \mathrm{NII}(t',u)}{\Delta R} \\ &= \sum_{i \in \mathcal{A}ssets} N_i(t',u) \cdot \left(\frac{\Delta R_i(t',u)}{\Delta R} - 1 \right) - \\ &\quad \sum_{j \in \mathcal{L}iabilities} N_j(t',u) \cdot \left(\frac{\Delta R_j(t',u)}{\Delta R} - 1 \right)\end{aligned}$$

If we consider a continuous-time analysis where $u = t' + \mathrm{d}t$, we obtain:

$$\begin{aligned}\mathrm{GAP}(t',u) &= \sum_{i \in \mathcal{A}ssets} N_i(t',u) \cdot \left(\frac{\partial R_i(t',u)}{\partial R} - 1 \right) - \\ &\quad \sum_{j \in \mathcal{L}iabilities} N_j(t',u) \cdot \left(\frac{\partial R_j(t',u)}{\partial R} - 1 \right)\end{aligned}$$

where R represents the market interest rate[45]. Demey *et al.* (2003) consider two opposite situations corresponding to two categories of asset/liability rates:

[43] This issue is analyzed in the third section of this chapter on page 427.
[44] The borrowing/lending interest rate is denoted by $R(t',u)$.
[45] We recall that the gap analysis assumes a flat yield curve.

$\mathcal{C}_1$ The asset/liability rates are deterministic and independent from market interest rates:

$$\frac{\partial R_i(t', u)}{\partial R} = \frac{\partial R_j(t', u)}{\partial R} = 0$$

This category corresponds to contractual rates that are generally fixed.

$\mathcal{C}_2$ The asset/liability rates depend on market interest rates:

$$\begin{cases} R_i(t', u) = R + m_A \\ R_j(t', u) = R + m_L \end{cases}$$

where m_A and m_L are the commercial margins for assets and liabilities. It follows that:

$$\frac{\partial R_i(t', u)}{\partial R} = \frac{\partial R_j(t', u)}{\partial R} = 1$$

This category generally concerns floating rates that are based on a market reference rate plus a spread.

We deduce that the gap is the difference between liabilities and assets that belong to the first category $\mathcal{C}_1$:

$$\text{GAP}(t', u) = \sum_{\substack{j \in \mathcal{L}iabilities \\ j \in \mathcal{C}_1}} N_j(t', u) - \sum_{\substack{i \in \mathcal{A}ssets \\ i \in \mathcal{C}_1}} N_i(t', u)$$

Modeling customer rates Until now, we have used the variable R for defining the general level of interest rates and ΔR for defining a parallel shock on the yield curve. However, this definition is not sufficiently precise to understand the real nature of R. In fact, the study of client rates is essential to understand which interest rate is important for calculating earnings-at-risk measures. In what follows, we introduce the notation $R(t) = R(t, t + \mathrm{d}t)$ and $R(u) = R(u, u + \mathrm{d}u)$. The current date or the agreement date is denoted by t while $u > t$ is a future date.

We have already distinguished fixed rates and floating (or variable) rates. By definition, a fixed rate must be known and constant when the agreement is signed between the customer and the bank:

$$R(u) = R^\star = R(t)$$

On the contrary, the customer rate is variable if:

$$\Pr\left\{\tilde{R}(u) = R(t)\right\} < 1$$

In this case, the customer rate is a random variable at time t and depends on a reference rate, which is generally a market rate. Mathematically, we can write:

$$\begin{aligned} \tilde{R}(u) &= R(t) \cdot \mathbb{1}\{u < \tau\} + \tilde{R}(\tau) \cdot \mathbb{1}\{u \geq \tau\} \\ &= R^\star \cdot \mathbb{1}\{u < \tau\} + \tilde{R}(\tau) \cdot \mathbb{1}\{u \geq \tau\} \end{aligned} \tag{7.17}$$

where τ is the time at which the customer rate will change. τ is also called the next repricing date. For some products, τ is known while it may be stochastic in some situations[46]. If $\tilde{R}(\tau)$ is a function of a market rate, we can write:

$$\tilde{R}(\tau) = f(\tau, r(\tau))$$

[46] When the repricing date is known, it is also called the reset date.

We use the notation $r(\tau)$, because the market rate is generally a short-term interest rate. If we assume a linear relationship (noted $\mathcal{H}_{\mathcal{L}inear}$), we have:

$$\tilde{R}(\tau) = \rho \cdot r(\tau) + \tilde{m} \tag{7.18}$$

where ρ is the correlation between the customer rate and the market rate and $\tilde{m}$ is related to the commercial margin[47]. This is the simplest way for modeling $\tilde{R}(\tau)$, but there are some situations where the relationship is more complex. For example, Demey *et al.* (2003) study the case where the customer rate has a cap:

$$\tilde{R}(\tau) = r(\tau) \cdot \mathbb{1}\left\{r(\tau) < r^+\right\} + r^+ \cdot \mathbb{1}\left\{r(\tau) \geq r^+\right\} + \tilde{m}$$

where $r^+ + \tilde{m}$ is the cap.

Another challenge for modeling $\tilde{R}(u)$ is the case where the next repricing date τ is unknown. We generally assume that τ is exponentially distributed with parameter λ. If we consider the linear relationship (7.18), it follows that the expected customer rate is:

$$\begin{aligned} R(u) &= \mathbb{E}\left[\tilde{R}(u)\right] \\ &= R^\star \cdot e^{-\lambda(u-t)} + (\rho \cdot r(u) + \tilde{m}) \cdot \left(1 - e^{-\lambda(u-t)}\right) \end{aligned} \tag{7.19}$$

Sometimes, the relationship between the customer rate and the market rate is not instantaneous. For instance, Demey *et al.* (2003) consider the case where the customer rate is an average of the market rate over a window period h. Therefore, Equation (7.19) becomes[48]:

$$R(u) = R^\star \cdot e^{-\lambda(u-t)} + \lambda \int_{u-h}^{u} (\rho \cdot r(s) + \tilde{m}) \cdot e^{-\lambda(s-t)}\, \mathrm{d}s$$

Let us go back to the problem of determining the parallel shock ΔR. Using Equation (7.17),we have:

$$\begin{aligned} \Delta R &= \tilde{R}(u) - R(t) \\ &= \begin{cases} 0 & \text{if } u < \tau \\ \tilde{R}(\tau) - R^\star & \text{otherwise} \end{cases} \end{aligned}$$

Under the assumption $\mathcal{H}_{\mathcal{L}inear}$, we deduce that:

$$\Delta R = \tilde{R}(\tau) - R^\star = \rho \cdot \Delta r \tag{7.20}$$

where $\Delta r = r(\tau) - r(t)$ is the shock on the market rate. We notice that modeling the net interest income variation requires determining ρ and Δr. In the case where $\rho = 0$, we retrieve the previous result that $\Delta\,\mathrm{NII}$ is not sensitive to fixed rate items. Otherwise, Equation (7.20) shows that interest rate gaps must be conducted on a contract by contract basis or at least for each reference rate:

> "*Floating-rate interest gaps can be defined for all floating-rate references (1-month Libor, 1-year Libor, etc.). These floating-rate gaps are not fungible: they cannot be aggregated unless assuming a parallel shift of all rates*" (Bessis, 2015, page 47).

[47] The commercial margin is equal to:

$$\begin{aligned} m &= \tilde{R}(\tau) - r(\tau) \\ &= \tilde{m} - (1-\rho)\, r(\tau) \end{aligned}$$

When the correlation is equal to one, $\tilde{m}$ is equal to the commercial margin, otherwise it is greater.

[48] We assume that $u - h \geq t$.

Indeed, two contracts may have two different correlations with the same reference rate, and two contracts may have two different reference rates.

Equation (7.20) is valid only if we assume that the next repricing date is known. If τ is stochastic, Demey *et al.* (2003) obtain the following formula:

$$\begin{aligned}\Delta R(u) &= \mathbb{E}\left[\left(\tilde{R}(u) - R(t)\right) \cdot \mathbb{1}\{u \geq \tau\}\right] \\ &= \rho \cdot \Delta r \cdot \Pr\{\tau \leq u\}\end{aligned}$$

We conclude that the sensitivity of the customer rate to the market rate is equal to:

$$\rho(t,u) = \frac{\Delta R(u)}{\Delta r} = \rho \cdot \Pr\{\tau \leq u\}$$

It depends on two parameters: the correlation ρ between the two rates and the probability distribution of the repricing date τ. If τ follows an exponential distribution with parameter λ, we have $\rho(t,u) = \rho\left(1 - e^{-\lambda(u-t)}\right)$. We verify that $\rho(t,u) \leq \rho$. The upper limit case $\rho(t,u) = \rho$ is reached in the deterministic case (no random repricing), whereas the function $\rho(t,u)$ is equal to zero if ρ is equal to zero (no correlation). By definition of the exponential distribution, the average time between two repricing dates is equal to $1/\lambda$. In Figure 7.13, we have reported the function $\rho(t,u)$ for three values of the correlation : 0%, 50% and 100%. We show how λ impacts the sensitivity $\rho(t,u)$ and therefore Δ NII. This last parameter is particularly important when we consider embedded options and customer behavior[49]. For instance, $\lambda = 0.1$ implies that the contract is repriced every ten years on average (top/left panel). It is obvious that the sensitivity is lower for this contract than for a contract that is repriced every 2 years (top/right panel).

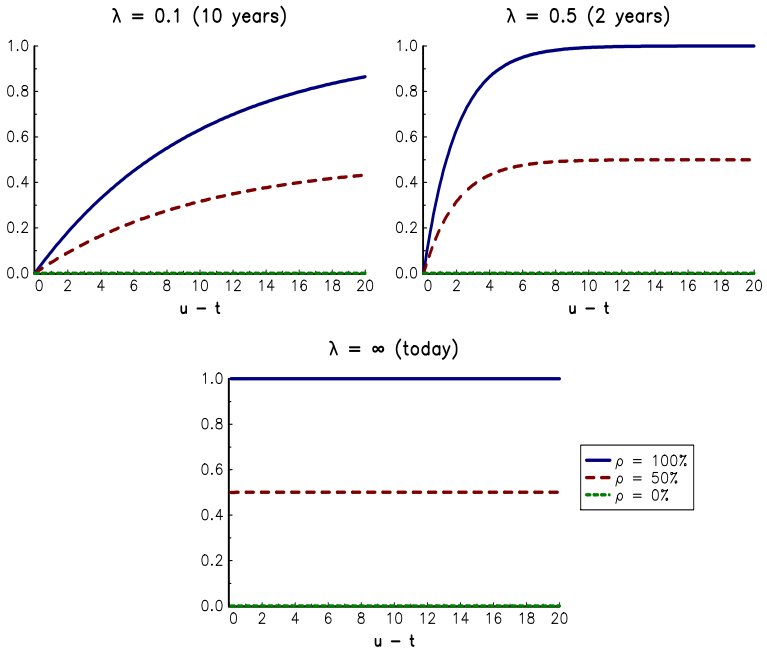


FIGURE 7.13: Sensitivity of the customer rate with respect to the market rate

[49] See Section 7.3 on page 427.

7.2.2.3 Hedging strategies

The question of hedging is not an easy task. There is no one optimal solution, but several answers. Moreover, this problem will be even more complicated when we will integrate the behavioral and embedded options.

To hedge or not to hedge Since the net interest income is sensitive to interest rate changes, it is important to define a hedging policy and to understand how it may impact the income statement of the bank. Let us define the hedged net interest income as the sum of the net interest income and the hedge P&L:

$$\mathrm{NII}_{\mathcal{H}}(t,u) = \mathrm{NII}(t,u) + \mathcal{H}(t,u)$$

In order to obtain a tractable formula of the hedge P&L $\mathcal{H}(t,u)$, we consider a forward rate agreement (FRA), which is an exchange contract between the future interest rate $r(u)$ at the pricing date u and the current forward rate $f(t,u)$ at the maturity date u. The hedge P&L is then:

$$\mathcal{H}(t,u) = N_{\mathcal{H}}(t,u) \cdot (f(t,u) - r(u))$$

where $N_{\mathcal{H}}(t,u)$ is the notional of the hedging strategy. We deduce that:

$$\begin{aligned}
\Delta\,\mathrm{NII}_{\mathcal{H}}(t,u) &= \Delta\,\mathrm{NII}(t,u) + \Delta\mathcal{H}(t,u) \\
&= \mathrm{GAP}(t,u) \cdot \Delta R(u) - N_{\mathcal{H}}(t,u) \cdot \Delta r(u) \\
&= (\mathrm{GAP}(t,u) \cdot \rho(t,u) - N_{\mathcal{H}}(t,u)) \cdot \Delta r(u)
\end{aligned}$$

because we have $\Delta R(u) = \rho(t,u) \cdot \Delta r(u)$. The hedged NII is equal to zero if the notional of the hedge is equal to the product of the interest rate gap and the sensitivity $\rho(t,u)$:

$$\Delta\,\mathrm{NII}_{\mathcal{H}}(t,u) = 0 \Leftrightarrow N_{\mathcal{H}}(t,u) = \mathrm{GAP}(t,u) \cdot \rho(t,u)$$

In this case, we obtain:

$$\mathrm{NII}_{\mathcal{H}}(t,u) - \mathrm{NII}(t,u) = \mathrm{GAP}(t,u) \cdot \rho(t,u) \cdot (f(t,u) - r(u))$$

We can draw several conclusions from the above mathematical framework:

- When the correlation between the customer rate and the market rate is equal to one, the notional of the hedge is exactly equal to the interest rate gap. Otherwise, it is lower.
- When the interest rate gap is closed, the bank does not need to hedge the net interest income.
- If the bank hedges the net interest income, the difference $\mathrm{NII}_{\mathcal{H}}(t,u) - \mathrm{NII}(t,u)$ is positive if the gap and the difference between $f(t,u)$ and $r(u)$ have the same sign. For example, if the gap is positive, a decrease of interest rates is not favorable. This implies that the hedged NII is greater than the non-hedged NII only if the forward rate $f(t,u)$ is greater than the future market rate $r(u)$. This situation is equivalent to anticipate that the forward rate is overestimated.

We conclude that hedging the interest rate gap is not systematic and depends on the expectations of the bank. It is extremely rare that the bank fully hedges the net interest income. The other extreme situation where the NII is fully exposed to interest rate changes is also not very common. Generally, the bank prefers to consider a partial hedging. Moreover,

we reiterate that the previous analysis is based on numerous assumptions[50]. Therefore, it is useless to compute a precise hedging strategy because of these approximations. This is why banks prefer to put in place macro hedging strategies with a limited number of instruments.

Hedging instruments In order to hedge the interest rate gap, the bank uses interest rate derivatives. They may be classified into two categories: those that hedge linear interest rate risks and those that hedge non-linear interest rate risks. The first category is made up of interest rate swaps (IRS) and forward rate agreements (FRA), while the second category concerns options such as caps, floors and swaptions. An IRS is a swap where two counterparties exchange a fixed rate against a floating rate or two floating rates. This is the hedging instrument which is certainly the most used in asset and liability management. The fixed rate is calibrated such that the initial value of the swap is equal to zero, meaning that the cost of entering into an IRS is low. This explains the popularity of IRS among ALM managers. However, these hedging instruments only concern linear changes in interest rates like the FRA instruments. In general, the ALM manager doesn't close fully all the interest rate gaps because this is not the purpose of a macro hedging strategy. In practice, two or three maturities are sufficient to highly reduce the risk.

Remark 82 *In order to hedge non-linear risks (slope of the yield curve, embedded options, etc.), the bank may use options. However, they are more expensive than IRS and are much less used by banks. One of the difficulties is the high degree of uncertainty around customer behavioral modeling.*

7.2.3 Simulation approach

We present here a general top-down econometric-based simulation framework in order to model the dynamics of the outstanding amount for the different items of the balance sheet. The underlying idea is that these items respond differently to key economic and market variables. The focus is then to model the earnings-at-risk profile of these items. The different profiles can also be aggregated in order to understand the income risk of each business line of the bank.

The framework is based on the cointegration theory and error correction models[51]. It is made up of two econometric models. We first begin by modeling the economic and market variables $x(t) = (x_1(t), \ldots, x_m(t))$ with a VECM:

$$\Phi_x(L)\,\Delta x(t) = \Pi_x x(t-1) + \varepsilon_x(t) \tag{7.21}$$

where $\Phi(L) = I_m - \Phi_1 L - \ldots - \Phi_p L^p$ is the lag polynomial and $\varepsilon_x(t) \sim \mathcal{N}(0, \Sigma_x)$. By definition, Equation (7.21) is valid if we have verified that each component of $x(t)$ is integrated of order one. The choice of the number p of lags is important. Generally, we consider a monthly econometric model, where the variables $x(t)$ are the economic growth $g(t)$, the inflation rate $\pi(t)$, the short-term market rate $r(t)$, the long-term interest rate $R(t)$, etc. In practice, $p = 3$ is used in order to have quarterly relationship between economic and market variables. The goal of this first econometric models is to simulate joint scenarios $\mathcal{S}_x$ of the economy and the market. Each scenario is represented by the current values of $x(t)$ and the future paths of $x(t+h)$:

$$\mathcal{S}_x = \{x(t+h) = (x_1(t+h), \ldots, x_m(t+h)), h = 0, 1, 2, \ldots\} \tag{7.22}$$

[50] They concern the sensitivity to markets rates, the behavior of customers, the new production, the interest rate shocks, etc.

[51] They are developed in Section 10.2.3 on page 655.

These scenarios do not necessarily correspond to extreme shocks, but they model the probability distribution of all future outcomes.

The second step consists in relating the growth of the outstanding amount $y_i(t)$ of item i to the variables $x(t)$. For instance, let us assume that:

$$y_i(t) = y_i(t-1) + 0.7 \times g(t) - 0.3 \times \pi(t)$$

This means that an economic growth of 1% implies that the outstanding amount of item i will increase by 70 bps, while the inflation has a negative impact on $y_i(t)$. The first idea is then to consider an ARX(q) model:

$$y_i(t) = \sum_{k=1}^{q} \phi_{i,k} y_i(t-k) + \sum_{j=1}^{m} \beta_{i,j} x_j(t) + \varepsilon_x(t)$$

However, this type of model has two drawbacks. It assumes that the current value of $y_i(t)$ is related to the current value of $x_j(t)$ and there are no substitution effects between the different items of the balance sheet. This is why it is better to consider again a VECM approach with exogenous variables:

$$\Phi_y(L)\Delta y(t) = \Pi_y y(t-1) + B_1 x(t) + B_2 \Delta x(t) + \varepsilon_y(t) \tag{7.23}$$

where $y(t) = (y_1(t), \ldots, y_n(t))$ and $\varepsilon_y(t) \sim \mathcal{N}(0, \Sigma_y)$. In this case, the current value of $y_i(t)$ is related to the current value of $x(t)$, the monthly variation $\Delta x(t)$ and the growth of the outstanding amount of the other items. Generally, the number q of lags is less than p. Indeed, the goal of the model (7.23) is to include short-term substitution effects between the different items whereas long-term substitution effects are more explained by the dynamics of economic and market variables.

Once the model (7.23) is estimated, we can simulate the future values of the outstanding amount for the different items with respect to the scenario $\mathcal{S}_x$ of the exogenous variables:

$$\mathcal{S}_y \mid \mathcal{S}_x = \{y(t+h) = (y_1(t+h), \ldots, y_n(t+h)), h = 0, 1, 2, \ldots\}$$

This framework allows going beyond the static gap analysis of interest rates, because the outstanding amounts are stochastic. For example, Figure 7.14 shows an earnings-at-risk analysis of the net interest income for the next six months. For each month, we report the median of NII and the 90% confidence interval.

Remark 83 *The previous framework can be used for assessing a given scenario, for example a parallel shock of interest rates. By construction, it will not give the same result than the income gap analysis, because this latter does not take into account the feedback effects of interest rates on the outstanding amount.*

7.2.4 Funds transfer pricing

According to Bessis (2015), the main objective of funds transfer pricing systems is to exchange funds and determine the profit allocation between business units. This means that all liquidity and interest rate risks are transferred to the ALM unit, which is in charge of managing them. Business units can then lend or borrow funding at a given internal price. This price is called the funds transfer price or the internal transfer rate, and is denoted by FTP. For example, the FTP charges interests to the business unit for client loans, whereas the FTP compensates the business unit for raising deposits. This implies that the balance sheet of the different business units is immunized to changes of market rates, and the internal transfer rates determine the net interest income of each business unit.

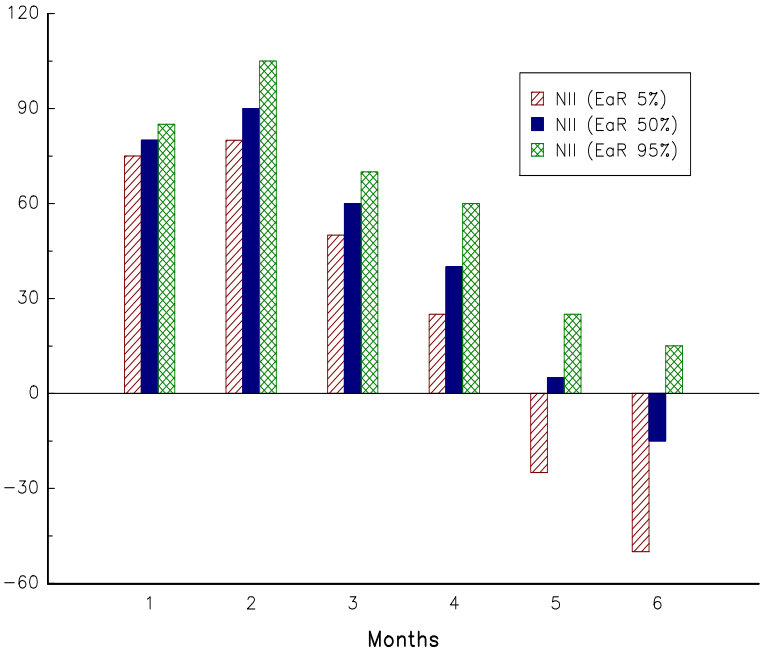


FIGURE 7.14: Earnings-at-risk analysis

7.2.4.1 Net interest and commercial margins

The net interest margin (NIM) is equal to the net interest income divided by the amount of assets:

$$\text{NIM}(t,u) = \frac{\sum_{i\in\mathcal{A}ssets} N_i(t,u)\cdot R_i(t,u) - \sum_{j\in\mathcal{L}iabilities} N_j(t,u)\cdot R_j(t,u)}{\sum_{i\in\mathcal{A}ssets} N_i(t,u)}$$

Let $\text{RA}(t,u)$ and $\text{RL}(t,u)$ be the interest earning assets and interest bearing liabilities (or asset and liability amounts that are sensitive to interest rates). Another expression of the NIM is:

$$\text{NIM}(t,u) = \frac{\text{RA}(t,u)\cdot R_{\text{RA}}(t,u) - \text{RL}(t,u)\cdot R_{\text{RL}}(t,u)}{\text{RA}(t,u)}$$

where R_{RA} and R_{RL} represent the weighted average interest rate of interest earning assets and interest bearing liabilities. The net interest margin differs from the net interest spread (NIS), which is the difference between interest earning rates and interest bearing rates:

$$\begin{aligned}\text{NIS}(t,u) &= \frac{\sum_{i\in\mathcal{A}ssets} N_i(t,u)\cdot R_i(t,u)}{\sum_{i\in\mathcal{A}ssets} N_i(t,u)} - \frac{\sum_{j\in\mathcal{L}iabilities} N_j(t,u)\cdot R_j(t,u)}{\sum_{j\in\mathcal{L}iabilities} N_j(t,u)}\\ &= R_{\text{RA}}(t,u) - R_{\text{RL}}(t,u)\end{aligned}$$

Example 73 *We consider the following interest earning and bearing items:*

Assets	$N_i(t,u)$	$R_i(t,u)$	*Liabilities*	$N_j(t,u)$	$R_j(t,u)$
Loans	*100*	*5%*	*Deposits*	*100*	*0.5%*
Mortgages	*100*	*4%*	*Debts*	*60*	*2.5%*

The interest income is equal to $100\times 5\% + 100\times 4\% = 9$ and the interest expense is $100\times 0.5\% + 60\times 2.5\% = 2$. We deduce that the net interest income is equal to $9-2=7$.

Moreover, we obtain[52] $\text{RA}(t,u) = 200$, $R_{\text{RA}}(t,u) = 4.5\%$, $\text{RL}(t,u) = 160$ and $R_{\text{RL}}(t,u) = 1.25\%$. We deduce that:

$$\text{NIM}(t,u) = \frac{200 \times 4.5\% - 160 \times 1.25\%}{200} = \frac{7}{200} = 3.5\%$$

and:

$$\text{NIS}(t,u) = 4.5\% - 1.25\% = 3.25\%$$

The net interest margin and spread are expressed in percent. NIM is the profitability ratio of the assets whereas NIS is the interest rate spread captured by the bank.

Remark 84 *In Figure 7.15, we have reported the average net interest margin in % for all US banks from 1984 to 2019. The average NIM was equal to* 3.36% *at the end of the first quarter of 2019. During the last 15 years, the average value is equal to* 3.78%*, the maximum* 4.91% *has been reached during Q1 1994 whereas the minimum* 2.95% *was observed in Q1 2015.*

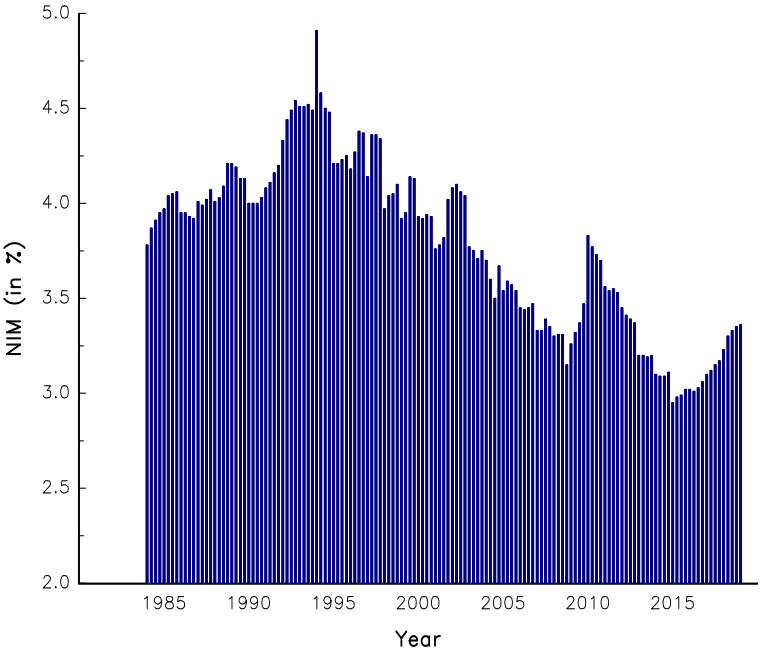


FIGURE 7.15: Evolution of the net interest margin in the US

Source: Federal Financial Institutions Examination Council (US), Net Interest Margin for all US Banks [USNIM], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/USNIM, July 9, 2019.

[52] We have:

$$R_{\text{RA}}(t,u) = \frac{100 \times 5\% + 100 \times 4\%}{100 + 100} = 4.5\%$$

and:

$$R_{\text{RL}}(t,u) = \frac{100 \times 0.5\% + 60 \times 2.5\%}{100 + 60} = 1.25\%$$

Let us now see how to calculate the commercial margin rate. A first idea is to approximate it by the net interest margin or the net interest spread. However, these quantities are calculated at the global level of the bank, not at the level of a business unit and even less at the level of a product. Let us consider an asset i. From a theoretical point of view, the commercial margin rate is the spread between the client rate of this asset $R_i(t,u)$ and the corresponding market rate $r(t,u)$:

$$m_i(t,u) = R_i(t,u) - r(t,u)$$

Here, we assume that $R_i(t,u)$ and $r(t,u)$ have the same maturity u. If we consider a liability j, we obtain a similar formula:

$$m_j(t,u) = r(t,u) - R_j(t,u)$$

In this framework, we assume that the business unit borrows at the market rate $r(t,u)$ in order to finance the asset i or lends to the market at the same rate $r(t,u)$. A positive commercial margin rate implies that $R_i(t,u) > r(t,u)$ and $r(t,u) > R_j(t,u)$. In the case where we can perfectly match the asset i with the liability j, the commercial margin rate is the net interest spread:

$$\begin{aligned} m(t,u) &= m_i(t,u) + m_j(t,u) \\ &= R_i(t,u) - R_j(t,u) \end{aligned}$$

As already said, a funds transfer pricing system is equivalent to interpose the ALM unit between the business unit and the market. In the case of assets, we decompose the commercial margin rate of the bank as follows:

$$\begin{aligned} m_i(t,u) &= R_i(t,u) - r(t,u) \\ &= \underbrace{(R_i(t,u) - \text{FTP}_i(t,u))}_{m_i^{(c)}(t,u)} + \underbrace{(\text{FTP}_i(t,u) - r(t,u))}_{m_i^{(t)}(t,u)} \end{aligned}$$

where $m_i^{(c)}(t,u)$ and $m_i^{(t)}(t,u)$ are the commercial margin rate of the business unit and the transformation margin rate of the ALM unit. For liabilities, we also have:

$$\begin{aligned} m_j(t,u) &= m_j^{(c)}(t,u) + m_j^{(t)}(t,u) \\ &= (\text{FTP}_j(t,u) - R_j(t,u)) + (r(t,u) - \text{FTP}_j(t,u)) \end{aligned}$$

The goal of FTP is then to lock the commercial margin rate $m_i^{(c)}(t,u)$ (or $m_j(t,u)$) over the lifetime of the product contract.

Let us consider Example 73. The FTP for the loans and the mortgages is equal to 3%, while the FTP for deposits is equal to 1.5% and the FTP for debts is equal to 2.5%. If we assume that the market rate is equal to 2.5%, we obtain the following results:

Assets	$m_i^{(c)}(t,u)$	$m_i^{(t)}(t,u)$	Liabilities	$m_j^{(c)}(t,u)$	$m_j^{(t)}(t,u)$
Loans	2%	0.5%	Deposits	1.0%	1.0%
Mortgages	1%	0.5%	Debts	0.0%	0.0%

It follows that the commercial margin of the bank is equal to:

$$\begin{aligned} M^{(c)} &= 100 \times 2\% + 100 \times 1\% + 100 \times 1\% + 60 \times 0\% \\ &= 4 \end{aligned}$$

For the transformation margin, we have:

$$\begin{aligned} M^{(t)} &= 100 \times 0.5\% + 100 \times 0.5\% + 100 \times 1.0\% + 60 \times 0\% \\ &= 2.0 \end{aligned}$$

We don't have $M^{(c)} + M^{(t)} = \text{NII}$ because assets and liabilities are not compensated:

$$\begin{aligned} \text{NII} - \left(M^{(c)} + M^{(t)}\right) &= (\text{RA}(t,u) - \text{RL}(t,u)) \cdot r(t,u) \\ &= 40 \times 2.5\% \\ &= 1 \end{aligned}$$

In fact, in a funds transfer pricing system, the balance sheet issue is the problem of the ALM unit. It is also interesting to notice that we can now calculate the commercial margin of each product: $M^{(c)}_{Loans} = 2$, $M^{(c)}_{Mortgages} = 1$ and $M^{(c)}_{Deposits} = 1$. We can then aggregate them by business units. For example, if the business unit is responsible for loans and deposits, its commercial margin is equal to 3.

7.2.4.2 Computing the internal transfer rates

Since the business unit knows the internal prices of funding, the commercial margin rates are locked and the commercial margin has a smooth profile. The business unit can then focus on its main objective, which is selling products and not losing time in managing interest rate and liquidity risks. However, in order to do correctly its job, the internal prices must be fair. The determination of FTPs is then crucial because it has a direct impact on the net income of the business unit. A system of arbitrary or wrong prices can lead to a false analysis of the income allocation, where some business units appear to be highly profitable when the exact opposite is true. The consequence is then a wrong allocation of resources and capital.

The reference rate If we consider the transformation margin rate, we have $m_i^{(t)}(t,u) = \text{FTP}_i(t,u) - r(t,u)$. The internal prices are fair if the corresponding mark-to-market is equal to zero on average, because the goal of FTP is to smooth the net interest income of each business unit and to allocate efficiently the net interest income between the different business units. For a contract with a bullet maturity, this implies that:

$$\text{FTP}_i(t,u) = \mathbb{E}[r(t,u)]$$

The transformation margin can then be interpreted as an interest rate swap[53] receiving a fixed leg $\text{FTP}_i(t,u)$ and paying a floating leg $r(t,u)$. It follows that the funds transfer price is equal to the market swap rate at the initial date t with the same maturity than the asset item i (Demey *et al.*, 2003).

In practice, it is impossible to have funds transfer prices that depend on the initial date and the maturity of each contract. Let us first assume that the bank uses the short market rate $r(u)$ for determining the funds transfer prices and considers globally the new production $\text{NP}(t)$ instead of the different individual contracts. The mark-to-market of the transformation margin satisfies then the following equation:

$$\mathbb{E}_t\left[\int_t^\infty B(t,u)\,\text{NP}(t)\,\mathbf{S}(t,u)\,(\text{FTP}(t,u) - r(u))\,\mathrm{d}u\right] = 0$$

[53] In the case of liabilities, the transformation margin is an interest rate swap paying the fixed leg $\text{FTP}_i(t,u)$ and receiving the floating leg $r(t,u)$.

As noticed by Demey *et al.* (2003), we need another constraint to determine explicitly the internal transfer rate, because the previous equation is not sufficient. For instance, if we assume that the internal transfer rate is constant over the lifetime of the new production — $\text{FTP}(t,u) = \text{FTP}(t)$, we obtain:

$$\text{FTP}(t,u) = \frac{\mathbb{E}_t\left[\int_t^\infty B(t,u)\,\mathbf{S}(t,u)\,r(u)\,\mathrm{d}u\right]}{\mathbb{E}_t\left[\int_t^\infty B(t,u)\,\mathbf{S}(t,u)\,\mathrm{d}u\right]}$$

The drawback of this approach is that the commercial margin is not locked, and the business unit is exposed to the interest rate risk. On the contrary, we can assume that the commercial margin rate of the business unit is constant:

$$R(u) - \text{FTP}(t,u) = m$$

Demey *et al.* (2003) show that[54]:

$$\text{FTP}(t,u) = R(u) + \frac{\mathbb{E}_t\left[\int_t^\infty B(t,u)\,\mathbf{S}(t,u)\,(r(u) - R(u))\,\mathrm{d}u\right]}{\mathbb{E}_t\left[\int_t^\infty B(t,u)\,\mathbf{S}(t,u)\,\mathrm{d}u\right]}$$

The term structure of funds transfer prices According to Bessis (2015), there are two main approaches for designing a funds transfer pricing system: cash netting and central cash pool systems. In the first case, the business unit transfers to the ALM unit only the net cash balance, meaning that the internal transfer rates apply only to a fraction of asset and liability items. This system presents a major drawback, because business units are exposed to interest rate and liquidity risks. On the contrary, all funding and investment items are transferred into the ALM book in the second approach. In this case, all items have their own internal transfer rate. In order to reduce the complexity of the FTP system, assets and liabilities are generally classified into homogeneous pools in terms of maturity, credit, etc. In this approach, each pool has its own FTP. For example, the reference rate of long maturity pools is a long-term market rate while the reference rate of short maturity pools is a short-term market rate. In Figure 7.16, we have represented the term structure of the FTPs. Previously, we have seen that the reference rate is the market swap rate, meaning that the reference curve is the IRS curve. In practice, the FTP curve will differ from the IRS curve for several reasons. For instance, the reference curve can be adjusted by adding a credit spread in order to reflect the credit-worthiness of the bank, a bid-ask spread in order to distinguish assets and liabilities, a behavior-based spread because of prepayment and embedded options, and a liquidity spread. Therefore, we can decompose the funds transfer price as follows:

$$\text{FTP}(t,u) = \text{FTP}^{\text{IR}}(t,u) + \text{FTP}^{\text{Liquidity}}(t,u) + \text{FTP}^{\text{Other}}(t,u)$$

where $\text{FTP}^{\text{IR}}(t,u)$ is the interest rate component, $\text{FTP}^{\text{IR}}(t,u)$ is the liquidity component and $\text{FTP}^{\text{Other}}(t,u)$ corresponds to the other components. The FTP curve can then be different than the IRS curve for the reasons presented above. But it can also be different because of business or ALM decisions. For instance, if the bank would like to increase its mortgage market share, it can reduce the client rate $R_i(t,u)$ meaning that the commercial

[54] Using this formulation, we can show the following results:

- for a loan with a fixed rate, the funds transfer price is exactly the swap rate with the same maturity than the loan and the same amortization scheme than the new production;
- if the client rate $R(u)$ is equal to the short-term market rate $r(u)$, the funds transfer price $\text{FTP}(t,u)$ is also equal to $r(u)$.

margin $m_i^{(c)}(t,u)$ decreases, or it can maintain the commercial margin by reducing the internal transfer rate $\text{FTP}_i(t,u)$. Another example concerns the investment maturity of retail deposits. Each time this maturity is revisited, it has a big impact on the retail business unit because a shorter maturity will reduce the internal transfer price and a longer maturity will increase the internal transfer price. Therefore, the FTP of deposits highly impacts the profitability of the retail business unit.

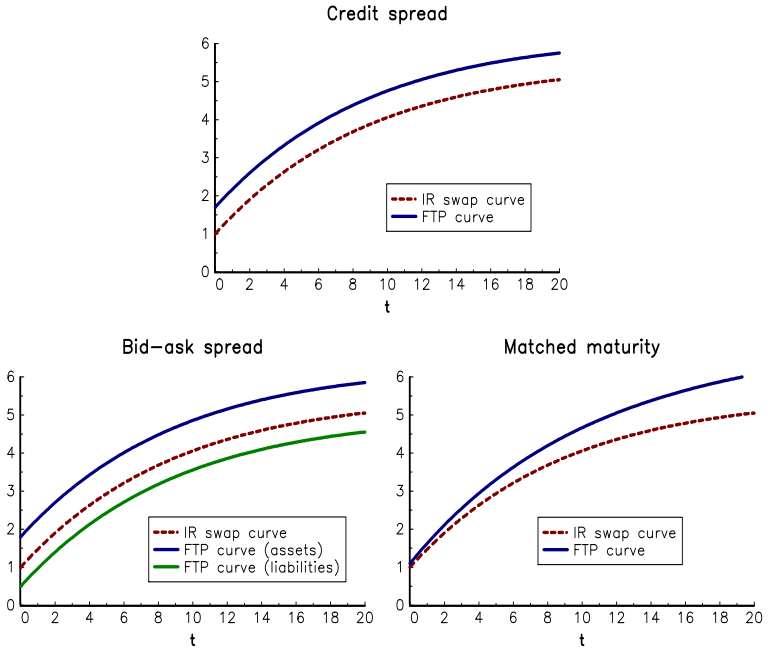


FIGURE 7.16: The term structure of FTP rates

7.3 Behavioral options

In this section, we focus on three behavioral options that make it difficult to calculate liquidity and interest rate risks. They have been clearly identified by the BCBS (2016d) and concern non-maturity deposits, prepayment risk and redemption (or early termination) issues. For NMDs, the challenge is to model the deposit volume and the associated implicit duration. For the two other risks, the goal is to calculate prepayment rates and redemption ratios on a yearly basis.

7.3.1 Non-maturity deposits

Let us assume that the deposit balance of the client A is equal to \$500. In this case, we can assume that the duration of this deposit is equal to zero day, because the client could withdraw her deposit volume today. Let us now consider 1 000 clients, whose deposit balance is equal to \$500. On average, we observe that the probability to withdraw \$500 at once is equal to 50%. The total amount that may be withdrawn today is then between \$0 and \$500 000. However, it is absurd to think that the duration of deposits is equal to zero,

because the probability that \$500 000 are withdrawn is less than $10^{-300}\%$! Since we have $\Pr\{S > 275000\} < 0.1\%$, we can decide that 55% of the deposit balance has a duration of zero day, 24.75% has a duration of one day, 11.14% has a duration of two days, etc. It follows that the duration of deposits depends on the average behavior of customers and the number of account holders, but many other parameters may have an impact on non-maturity deposits. From a contractual point of view, deposits have a very short-term duration. From a statistical point of view, we notice that a part of these deposits are in fact very stable because of the law of large numbers.

NMDs are certainly the balance sheet item that is the most difficult to model. There are multiple reasons. The first reason is the non-specification of a maturity in the contract. The second reason is that NMDs are the most liquid instruments and their transaction costs are equal to zero, implying that subscriptions and redemptions are very frequent. This explains that the volume of deposits is the most volatile among the different banking products at the individual level. Another reason is the large number of embedded options that creates significant gamma and vega option risks (Blöchlinger, 2015). Finally, the volume of NMDs is very sensitive to the monetary policy (Bank of Japan, 2014), because NMDs are part of the M1 money supply, but also of the M2 money supply. Indeed, NMDs is made up of demand deposits (including overnight deposits and checkable accounts) and savings accounts. M1 captures demand deposits (and also currency in circulation) while M2 − M1 is an approximation of savings accounts. In what follows, we do not make a distinction between NMDs, but it is obvious that the bank must distinguish demand deposits and savings accounts in practice. Generally, academics model behavioral options related to NMDs by analyzing substitution effects between NMDs and term deposits. In the real life, demand-side substitution is more complex since it also concerns the cross-effects between demand deposits and savings accounts.

7.3.1.1 Static and dynamic modeling

In the case of non-maturity deposits, it is impossible to make the distinction between the entry dates. This means that the stock amortization function $\mathbf{S}^\star(t,u)$ must be equal to the amortization function $\mathbf{S}(t,u)$ of the new production. This implies that the hazard rate $\lambda(t,u)$ of the amortization function $\mathbf{S}(t,u)$ does not depend on the entry date t:

$$\lambda(t,u) = \lambda(u)$$

Indeed, we have by definition:

$$\mathbf{S}(t,u) = \exp\left(-\int_t^u \lambda(s)\,\mathrm{d}s\right)$$

and we verify that[55]:

$$\mathbf{S}^\star(t,u) = \frac{\int_{-\infty}^t \mathrm{NP}(s)\,S(s,u)\,\mathrm{d}s}{\int_{-\infty}^t \mathrm{NP}(s)\,S(s,t)\,\mathrm{d}s} = \mathbf{S}(t,u)$$

According to Demey *et al.* (2003), the concept of new production has no meaning. Then, we must focus on the modeling of the current volume of NMDs, which is given by Equation

[55] This result is based on the following computation:

$$\frac{\int_{-\infty}^t \mathrm{NP}(s)\,e^{-\int_s^u \lambda(v)\,\mathrm{d}v}\,\mathrm{d}s}{\int_{-\infty}^t \mathrm{NP}(s)\,e^{-\int_s^t \lambda(v)\,\mathrm{d}v}\,\mathrm{d}s} = \frac{\int_{-\infty}^t \mathrm{NP}(s)\,e^{-\left(\int_s^t \lambda(v)\,\mathrm{d}v+\int_t^u \lambda(v)\,\mathrm{d}v\right)}\,\mathrm{d}s}{\int_{-\infty}^t \mathrm{NP}(s)\,e^{-\int_s^t \lambda(v)\,\mathrm{d}v}\,\mathrm{d}s} = e^{-\int_t^u \lambda(v)\,\mathrm{d}v}$$

(7.7) on page 386:

$$N(t) = \int_{-\infty}^{t} \text{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s$$

It follows that:

$$\begin{aligned}\frac{\mathrm{d}N(t)}{\mathrm{d}t} &= \text{NP}(t)\,\mathbf{S}(t,t) - \int_{-\infty}^{t} \text{NP}(s)\,f(s,t)\,\mathrm{d}s \\ &= \text{NP}(t) - \lambda(t)\int_{-\infty}^{t} \text{NP}(s)\,\mathbf{S}(s,t)\,\mathrm{d}s\end{aligned}$$

or:

$$\mathrm{d}N(t) = (\text{NP}(t) - \lambda(t)\,N(t))\,\mathrm{d}t \tag{7.24}$$

Therefore, the variation of $N(t)$ is the difference between deposit inflows $\text{NP}(t)$ and deposit outflows $\lambda(t)\,N(t)$. In the case where the new production and the hazard rate are constant – $\text{NP}(t) = \text{NP}$ and $\lambda(t) = \lambda$, we obtain[56] $N(t) = N_\infty + (N_0 - N_\infty)\,e^{-\lambda(t-t_0)}$ where $N_0 = N(t_0)$ is the current value and $N_\infty = \lambda^{-1}\,\text{NP}$ is the long-term value of $N(t)$. In this case, Equation (7.24) becomes:

$$\mathrm{d}N(t) = \lambda(N_\infty - N(t))\,\mathrm{d}t \tag{7.25}$$

We recognize the deterministic part of the Ornstein-Uhlenbeck process:

$$\mathrm{d}N(t) = \lambda(N_\infty - N(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \tag{7.26}$$

where $W(t)$ is a Brownian motion. In this case, the solution is given by[57]:

$$N(t) = N_0 e^{-\lambda(t-t_0)} + N_\infty\left(1 - e^{-\lambda(t-t_0)}\right) + \sigma\int_{t_0}^{t} e^{-\lambda(t-s)}\,\mathrm{d}W(s) \tag{7.27}$$

The estimation of the parameters $(\lambda, N_\infty, \sigma)$ can be done using the generalized method of moments (GMM) or the method of maximum likelihood (ML). In this case, we can show that:

$$N(t) \mid N(s) = N_s \sim \mathcal{N}\left(\mu_{(s,t)}, \sigma^2_{(s,t)}\right)$$

where:

$$\mu_{(s,t)} = N_s e^{-\lambda(t-s)} + N_\infty\left(1 - e^{-\lambda(t-s)}\right)$$

and:

$$\sigma^2_{(s,t)} = \sigma^2\left(\frac{1 - e^{-2\lambda(t-s)}}{2\lambda}\right)$$

Example 74 *We consider a deposit account with the following characteristics:* $N_\infty = \$1\,000$, $\lambda = 10$ *and* $\sigma = \$1\,000$.

The frequency λ means that the average duration of the deposit balance is equal to $1/\lambda$. In our case, we find $1/10 = 0.1$ years or 1.2 months. The new production is $\text{NP} = \lambda N_\infty = \$10\,000$. This new production can be interpreted as the annual income of the client

[56] The solution of Equation (7.24) is given by:

$$N(t) - \frac{\text{NP}}{\lambda} = \left(N_0 - \frac{\text{NP}}{\lambda}\right)e^{-\lambda(t-t_0)}$$

[57] See Appendix A.3.8.2 on page 1075.

that is funded the deposit account. In Figure 7.17, the top panel represents the expected value $\mu_{(0,t)}$ of the deposit balance by considering different current values N_0, the top left panel corresponds to the density function[58] $f_{(s,t)}(x)$ of $N(t)$ given that $N(s) = N_s$ and the bottom panel shows three simulations of the stochastic process $N(t)$.

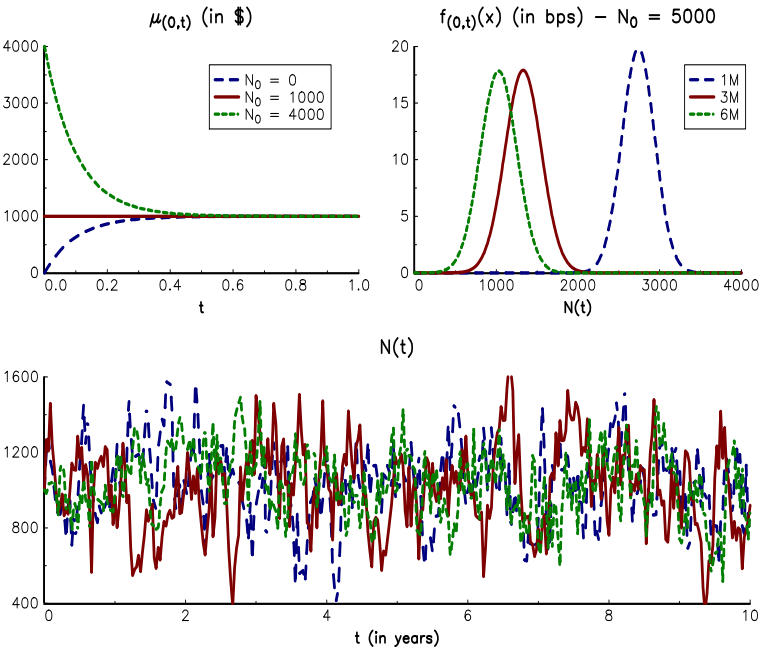


FIGURE 7.17: Statistics of the deposit amount $N(t)$

Another extension of Model (7.25) is to make the distinction between stable and non-stable deposits. Let g be the growth rate of deposits. The total amount of deposits $D(t)$ is given by:

$$D(t) = e^{g(t-s)} \sum_{i=1}^{n_t} N_i(t)$$

where n_t is the number of deposit accounts and $N_i(t)$ is the deposit balance of the i^{th} deposit account. It follows that:

$$\begin{aligned} D(t) &= e^{g(t-s)} \sum_{i=1}^{n_t} N_{\infty,i} + e^{g(t-s)} \sum_{i=1}^{n_t} (N_{s,i} - N_{\infty,i}) e^{-\lambda_i (t-s)} + \\ & e^{g(t-s)} \sum_{i=1}^{n_t} \sigma_i \sqrt{\frac{1 - e^{-2\lambda_i (t-s)}}{2\lambda_i}} \varepsilon_i(t) \end{aligned}$$

where $\varepsilon_i(t) \sim \mathcal{N}(0,1)$. By considering a representative agent, we can replace the previous equation by the following expression:

$$D(t) = D_\infty e^{g(t-s)} + (D_s - D_\infty) e^{(g-\lambda)(t-s)} + \varepsilon(t) \tag{7.28}$$

[58] We have:

$$f_{(s,t)}(x) = \frac{1}{\sigma(s,t)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu_{(s,t)}}{\sigma_{(s,t)}}\right)^2\right)$$

where $D_\infty = \sum_{i=1}^{n_t} N_{\infty,i}$, $D_s = \sum_{i=1}^{n_t} N_{s,i}$, λ^{-1} is the weighted average duration of deposits and $\varepsilon(t)$ is the stochastic part. Demey *et al.* (2003) notice that we can decompose $D(t)$ into two terms:

$$D(t) = D_{\text{long}}(s,t) + D_{\text{short}}(s,t)$$

where $D_{\text{long}}(s,t) = D_\infty e^{g(t-s)}$ and $D_{\text{short}}(s,t) = (D_s - D_\infty) e^{(g-\lambda)(t-s)} + \varepsilon(t)$. This breakdown seems appealing at first sight, but it presents a major drawback. Indeed, the short component $D_{\text{short}}(s,t)$ may be negative. In practice, it is better to consider the following equation:

$$D(t) = \underbrace{\varphi D_\infty e^{g(t-s)}}_{D_{\text{stable}}(s,t)} + \underbrace{(D_s - D_\infty) e^{(g-\lambda)(t-s)} + \varepsilon(t) + (1-\varphi) D_\infty e^{g(t-s)}}_{D_{\text{non-stable}}(s,t)}$$

where $D_{\text{stable}}(s,t)$ corresponds to the amount of stable deposits and $D_{\text{non-stable}}(s,t) = D(t) - D_{\text{stable}}(s,t)$ is the non-stable deposit amount. At time $t = s$, we verify that[59]:

$$D(t) = D_{\text{stable}} + D_{\text{non-stable}}(t)$$

The estimation of stable deposits is a two-step process. First, we estimate D_∞ by using the ML method. Second, we estimate the fraction $\varphi < 1$ of the long-run amount of deposits that can be considered as stable. Generally, we calibrate the parameter φ such that φN_∞ is the quantile of $D(t)$ at a given confidence level (e.g. 90% or 95%).

In Figure 7.18, we assume that the deposit amount $D(t)$ follows an Ornstein-Uhlenbeck process with parameters $D_\infty = \$1$ bn, $\lambda = 5$ and $\sigma = \$200$ mn. In the top/right panel, we have reported the $D_{\text{long}}/D_{\text{short}}$ breakdown. We verify that the short component may be negative, meaning the long component cannot be considered as a stable part. This is not the case with the $D_{\text{stable}}/D_{\text{non-stable}}$ breakdown given in the bottom panels. The big issue is of course the estimation of the parameter φ. One idea might be to calibrate φ such that $\Pr\{D(t) \leq \varphi D_\infty\} = 1 - \alpha$ given the confidence level α. If we consider the Ornstein-Uhlenbeck dynamics, we obtain the following formula:

$$\varphi = 1 - \frac{\sigma \Phi^{-1}(1-\alpha)}{D_\infty \sqrt{2\lambda}}$$

In our example, this ratio is respectively equal to 85.3%, 89.6% and 91.9% when α takes the value 99%, 95% and 90%.

Remark 85 *We recall that the Basel Committee makes the distinction between stable and core deposits. It is assumed that the interest rate elasticity of NMDs is less than one. Core deposits are the proportion of stable deposits, whose pass through sensitivity is particularly low, meaning they are "unlikely to reprice even under significant changes in interest rate environment" (BCBS, 2016d, page 26).*

7.3.1.2 Behavioral modeling

If we assume that the growth rate g is equal to zero, the linearization of Equation (7.28) corresponds to the Euler approximation of the Ornstein-Uhlenbeck process:

$$D(t) \approx D(s) + \lambda(D_\infty - D(s)) + \varepsilon(t) \tag{7.29}$$

[59] The previous results are based on the dynamic analysis between time s and t. If we prefer to adopt a static analysis, the amount of non-stable deposits must be defined as follows:

$$D_{\text{non-stable}}(t) = D(t) - \varphi D_\infty$$

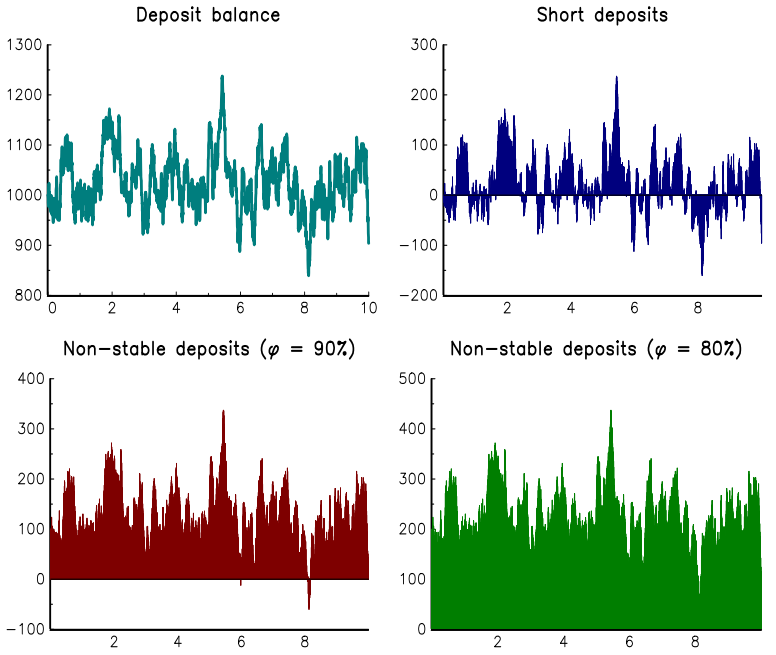


FIGURE 7.18: Stable and non-stable deposits

Here, $D(t)$ is the value of the non-maturity account balance or deposit volume. A similar expression is obtained by considering the individual deposit amount $N(t)$ instead of $D(t)$. In what follows, we use the same notation $D(t)$ for defining aggregated and individual deposit balances. Let us come back to the general case: $\mathrm{d}D(t) = (\mathrm{NP}(t) - \lambda(t) D(t))\,\mathrm{d}t$. By assuming that the new production is a function of the current balance, we have $\mathrm{NP}(t) = g(t, X(t)) D(t)$ where $g(t, X(t))$ depends on a set of explanatory variables $X(t)$. If follows that $\mathrm{d}\ln D(t) = (g(t, X(t)) - \lambda(t))\,\mathrm{d}t$ and:

$$\ln D(t) \approx \ln D(s) + g(s, X(s)) - \lambda(s) \tag{7.30}$$

Modeling the behavior of the client and introducing embedded options can be done by combining Equations (7.29) and (7.30):

$$\ln D(t) = \ln D(s) + \lambda(\ln D_\infty - \ln D(s)) + g(t, X(t)) + \varepsilon(t)$$

In this case, the main issue is to specify $g(t, X(t))$ and the explanatory variables that impact the dynamics of the deposit volume. Most of the time, $g(t, X(t))$ depends on two variables: the deposit rate $i(t)$ and the market rate $r(t)$. In what follows, we present several models that have been proposed for modeling either $D(t)$ or $i(t)$ or both. The two pioneer models are the deposit balance model of Selvaggio (1996) and the deposit rate model of Hutchison and Pennacchi (1996).

The Hutchison-Pennacchi-Selvaggio framework In Selvaggio (1996), the deposit rate $i(t)$ is exogenous and the bank account holder modifies his current deposit balance $D(t)$ to target a level $D^\star(t)$, which is defined as follows:

$$\ln D^\star(t) = \beta_0 + \beta_1 \ln i(t) + \beta_2 \ln Y(t)$$

where $Y(t)$ is the income of the account holder. The rational of this model is the following. In practice, the bank account holder targets a minimum positive balance in order to meet

his current liquidity and consumption needs, which are a function of his income $Y(t)$. For example, we can assume that the client with a monthly income of \$10 000 targets a larger amount than the client with a monthly income of \$1 000. Moreover, we can assume that the target balance depends on the deposit rate $i(t)$. The elasticity coefficient must be positive, meaning that the client has a high incentive to transfer his money into a term deposit account if the deposit rate is low. At time t, the account holder can face two situations. If $D_{t-1} < D_t^\star$, he will certainly increase his deposit volume in order to increase his cash liquidity. If $D_{t-1} > D_t^\star$, he will certainly transfer a part of his deposit balance into his term account. Therefore, the behavior of the bank account holder can be represented by a mean-reverting AR(1) process:

$$\ln D(t) - \ln D(t-1) = (1-\phi)(\ln D^\star(t) - \ln D(t-1)) + \varepsilon(t) \tag{7.31}$$

where $\varepsilon(t) \sim \mathcal{N}\left(0, \sigma^2\right)$ is a white noise process and $\phi \leq 1$ is the mean-reverting parameter. It follows that:

$$\begin{aligned} \ln D(t) &= \phi \ln D(t-1) + (1-\phi) \ln D^\star(t) + \varepsilon(t) \\ &= \phi \ln D(t-1) + \beta_0' + \beta_1' \ln i(t) + \beta_2' \ln Y(t) + \varepsilon(t) \end{aligned} \tag{7.32}$$

where $\beta_k' = (1-\phi)\beta_k$. Let $d(t) = \ln D(t)$ be the logarithm of the deposit volume. The model of Selvaggio (1996) is then a ARX(1) process:

$$d(t) = \phi d(t-1) + (1-\phi) d^\star(t) + \varepsilon(t) \tag{7.33}$$

where $d^\star(t) = \ln D^\star(t)$ is the exogenous variable.

In practice, the bank does not know the value $\theta = (\phi, \beta_0, \beta_1, \beta_2, \sigma)$ of the parameters. Moreover, these parameters are customer-specific and are different from one customer to another. The bank can then estimate the vector θ for a given customer if it had a sufficient history. For instance, we consider that a two-year dataset of monthly observations or a ten-year dataset of quarterly observations is generally sufficient to estimate five parameters. However, the variables $i(t)$ and $Y(t)$ rarely change, meaning that it is impossible to estimate θ for a given customer. Instead of using a time-series analysis, banks prefer then to consider a cross-section/panel analysis. Because Model (7.33) is linear, we can aggregate the behavior of the different customers. The average behavior of a customer is given by Equation (7.32) where the parameters ϕ, β_0, β_1, β_2 and σ are equal to the mean of the customer parameters. This approach has the advantage to be more robust in terms of statistical inference. Indeed, the regression is performed using a large number of observations (number of customers $\times$ number of time periods).

In the previous model, the deposit interest rate is given and observed at each time period. Hutchison and Pennacchi (1996) propose a model for fixing the optimal value of $i(t)$. They assume that the bank maximizes its profit:

$$i^\star(t) = \arg\max \Pi(t)$$

where the profit $\Pi(t)$ is equal to the revenue minus the cost:

$$\Pi(t) = r(t) \cdot D(t) - (i(t) + c(t)) \cdot D(t)$$

In this expression, $r(t)$ is the market interest rate and $c(t)$ is the cost of issuing deposits. By assuming that $D(t)$ is an increasing function of $i(t)$, the first-order condition is:

$$(r(t) - (i(t) + c(t))) \cdot \frac{\partial D(t)}{\partial i(t)} - D(t) = 0$$

We deduce that:

$$\begin{aligned} i^\star(t) &= r(t) - c(t) - \left(\frac{\partial D(t)}{\partial i(t)}\right)^{-1} D(t) \\ &= r(t) - \left(c(t) + \left(\frac{\partial d(t)}{\partial i(t)}\right)^{-1}\right) \\ &= r(t) - s(t) \end{aligned} \tag{7.34}$$

The deposit interest rate is then equal to the market interest rate $r(t)$ minus a spread[60] $s(t)$. Equations (7.32) and (7.34) are the backbone of various non-maturity deposit models.

The IRS framework Using arbitrage theory, Jarrow and van Deventer (1998) show that the deposit rate must be lower than the market rate[61] – $i(t) \leq r(t)$, and the current market value of deposits is the net present value of the cash flow stream $D(t)$:

$$V(0) = \mathbb{E}\left[\sum_{t=0}^{\infty} B(0, t+1)\,(r(t) - i(t))\,D(t)\right] \tag{7.35}$$

where $B(0,t)$ is the discount factor. Jarrow and van Deventer (1998) interpret $V(0)$ as an exotic interest rate swap, where the bank receives the market rate and pays the deposit rate. Since the present value of the deposit liability of the bank is equal to $L(0) = D(0) - V(0)$, the hedging strategy consists in "*investing* $D(0)$ *dollars in the shortest term bond* $B(0,1)$ *and shorting the exotic interest rate swap represented by* $V(0)$" (Jarrow and van Deventer, 1998, page 257). The complete computation of the hedging portfolio requires specifying $i(t)$ and $D(t)$. For example, Jarrow and van Deventer (1998) consider the following specification:

$$\ln D(t) = \ln D(t-1) + \beta_0 + \beta_1 r(t) + \beta_2 (r(t) - r(t-1)) + \beta_3 t \tag{7.36}$$

and:

$$i(t) = i(t) + \beta_0' + \beta_1' r(t) + \beta_2' (r(t) - r(t-1)) \tag{7.37}$$

The deposit balance and the deposit rate are linear in the market rate $r(t)$ and the variation of the market rate $\Delta r(t)$. The authors also add a trend in Equation (7.36) in order to take into account macroeconomic variables that are not included in the model.

The previous model is fully tractable in continuous-time. Beyond these analytical formulas, the main interest of the Jarrow-van Deventer model is to show that the modeling of non-maturity deposits is related to the modeling of interest rate swaps. Another important contribution of this model is the introduction of the replicating portfolio. Indeed, it is common to break down deposits into stable and non-stable deposits, and stable deposits into core and non-core deposits. The idea is then to replicate the core deposits with a hedging portfolio with four maturities (3, 5, 7 and 10 years). In this case, the funds transfer pricing of non-maturity deposits is made up of four internal transfer rates corresponding to the maturity pillars of the replicating portfolio.

[60] We notice that the spread $s(t)$ is the sum of the cost $c(t)$ and the Lerner index $\eta(t)$, where $\eta(t) = 1/e(t)$ and $e(t)$ is the interest rate elasticity of the demand.

[61] This inequality is obtained by assuming no arbitrage opportunities for individuals and market segmentation. In particular, Jarrow and van Deventer (1998) consider that the competition among banks is imperfect because of entry and mobility barriers to the banking industry.

Asymmetric adjustment models O'Brien (2001) introduces an asymmetric adjustment of the deposit rate:

$$\Delta i(t) = \alpha(t) \cdot (\hat{\imath}(t) - i(t-1)) + \eta(t)$$

where $\hat{\imath}(t)$ is the conditional equilibrium deposit rate and:

$$\alpha(t) = \alpha^+ \cdot \mathbb{1}\{\hat{\imath}(t) > i(t-1)\} + \alpha^- \cdot \mathbb{1}\{\hat{\imath}(t) < i(t-1)\}$$

If $\hat{\imath}_t > i(t-1)$, we obtain $\Delta i(t) = \alpha^+ \cdot (\hat{\imath}(t) - i(t-1)) + \eta(t)$, otherwise we have $\Delta i(t) = \alpha^- \cdot (\hat{\imath}(t) - i(t-1)) + \eta(t)$. The distinction between α^+ and α^- can be justified by the asymmetric behavior of banks and the rigidity of deposit rates. In particular, O'Brien (2001) suggests that $\alpha^- > \alpha^+$, implying that banks adjust more easily the deposit rate when the market rate decreases than when it increases. In this model, the deposit balance is a function of the spread $r(t) - i(t)$:

$$\ln D(t) = \beta_0 + \beta_1 \ln D(t-1) + \beta_2 (r(t) - i(t)) + \beta_3 \ln Y(t) + \varepsilon(t)$$

Moreover, O'Brien (2001) assumes that the conditional equilibrium deposit rate is a linear function of the market rate:

$$\hat{\imath}(t) = \gamma_0 + \gamma_1 \cdot r(t)$$

In the previous model, the asymmetric adjustment explicitly concerns the deposit interest rate $i(t)$ and implicitly impacts the deposit balance $D(t)$ because of the spread $r(t) - i(t)$. Frachot (2001) considers an extension of the Selvaggio model by adding a correction term that depends on the market interest rate $r(t)$ and a threshold:

$$\ln D(t) - \ln D(t-1) = (1-\phi)(\ln D^\star(t) - \ln D_{t-1}) + \delta_c(r(t), r^\star) \tag{7.38}$$

where $\delta_c(r(t), r^\star) = \delta \cdot \mathbb{1}\{r(t) \leq r^\star\}$ and $r^\star$ is the interest rate floor. When market interest rates are too low and below $r^\star$, the bank account holder does not make the distinction between deposit and term balances, and we have:

$$\delta_c(r(t), r^\star) = \begin{cases} \delta & \text{if } r(t) \leq r^\star \\ 0 & \text{otherwise} \end{cases}$$

Contrary to the Selvaggio model, the average behavior is not given by Equation (7.38) because of the non-linearity pattern. Let f be the probability density function of the threshold $r^\star$ among the different customers of the bank. On average, we have:

$$\begin{aligned} \mathbb{E}[\delta_c(r(t), r^\star)] &= \int_0^\infty \delta \cdot \mathbb{1}\{r(t) \leq x\} \cdot f(x)\, \mathrm{d}x \\ &= \delta \cdot (1 - \mathbf{F}(r(t))) \end{aligned}$$

The average behavior is then given by the following equation:

$$d(t) - d(t-1) = (1-\phi)(d^\star(t) - d(t-1)) + \delta(1 - \mathbf{F}(r(t)))$$

where $d(t) = \ln D(t)$ and $d^\star(t) = \ln D^\star(t)$. For example, if we assume that the distribution of $r^\star$ is uniform on the range $[0; r^\star_{\max}]$, we obtain $f(x) = 1/r^\star_{\max}$ and $\mathbf{F}(x) = \min(x/r^\star_{\max}, 1)$. We deduce that:

$$\begin{aligned} d(t) - d(t-1) &= (1-\phi)(d^\star(t) - d(t-1)) + \delta\left(1 - \min\left(\frac{r(t)}{r^\star_{\max}}, 1\right)\right) \\ &= (1-\phi)(d^\star(t) - d(t-1)) + \delta\frac{\max(r^\star_{\max} - r(t), 0)}{r^\star_{\max}} \end{aligned}$$

In the case where $r^\star \sim \mathcal{N}\left(\mu_\star, \sigma_\star^2\right)$, we obtain:

$$d(t) - d(t-1) = (1-\phi)\left(d^\star(t) - d(t-1)\right) + \delta\Phi\left(\frac{\mu_\star - r(t)}{\sigma_\star}\right)$$

Another asymmetric model was proposed by OTS (2001):

$$d(t) = d(t-1) + \Delta \ln\left(\beta_0 + \beta_1 \arctan\left(\beta_2 + \beta_3 \frac{i(t)}{r(t)}\right) + \beta_4 i(t)\right) + \varepsilon(t)$$

where Δ corresponds to the frequency. The '*Net Portfolio Value Model*' published by the Office of Thrift Supervision[62] is a comprehensive report that contains dozens of models in order to implement risk management and ALM policies. For instance, Chapter 6 describes the methodologies for modeling liabilities and Section 6.D is dedicated to demand deposits. These models were very popular in the US in the 1990s. In 2011, the Office of the Comptroller of the Currency (OCC) provided the following parameters for the monthly model[63] of transaction accounts: $\beta_0 = 0.773$, $\beta_1 = -0.065$, $\beta_2 = -5.959$, $\beta_3 = 0.997$ and $\beta_4 = 1$ bp. In the case of money market accounts, the parameters were $\beta_0 = 0.643$, $\beta_1 = -0.069$, $\beta_2 = -6.284$, $\beta_3 = 2.011$ and $\beta_4 = 1$ bp.

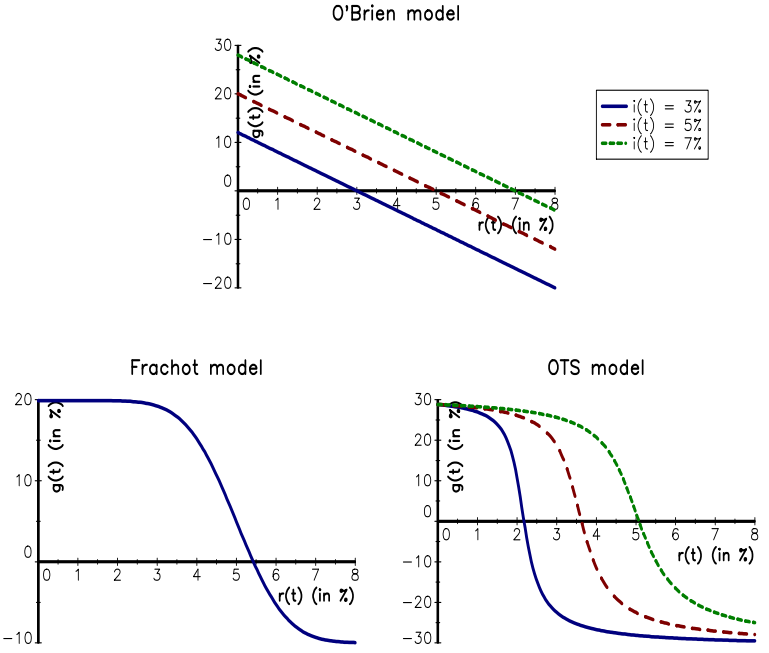


FIGURE 7.19: Impact of the market rate on the growth rate of deposits

In Figure 7.19, we compare the growth rate $g(t)$ of deposits for the different asymmetric models. For the O'Brien model, the growth rate is equal to $g(t) = \beta_2(r(t) - i(t))$. In the case of the Frachot model, the market rate has only a positive impact because $\delta_c(r(t), r^\star) \geq 0$. This is why we consider an extended version where the correction term is equal to

[62] The mission of OTS is to "*supervise savings associations and their holding companies in order to maintain their safety and soundness and compliance with consumer laws and to encourage a competitive industry that meets America's financial services needs*".

[63] We have $\Delta = 1/12$.

$\delta_c(r(t), r^\star) - \delta^-$. The growth rate is then $g(t) = \delta(1 - \mathbf{F}(r(t))) - \delta^-$. Finally, the growth rate of the OTS model is equal to $g(t) = \ln\left(\beta_0 + \beta_1 \arctan\left(\beta_2 + \beta_3 \frac{i(t)}{r(t)}\right) + \beta_4 i(t)\right)$. Using several value of the deposit rate $i(t)$, we measure the impact of the market rate $r(t)$ on the growth rate $g(t)$ using the following parameters: $\beta_2 = -4$ (O'Brien model), $\delta = 30\%$, $\mu_\star = 5\%$, $\sigma_\star = 1\%$ and $\delta^- = 10\%$ (Frachot model), and $\beta_0 = 1.02$, $\beta_1 = 0.2$, $\beta_2 = -7$, $\beta_3 = 5$ and $\beta_4 = 0$ (OTS model). The O'Brien model is linear while the Frachot model is non-linear. However, the Frachot model does not depend on the level of the deposit rate $i(t)$. The OTS model combines non-linear effects and the dependence on the deposit rate.

Remark 86 *These different models have been extended in order to take into account other explanatory variables such that the CDS of the bank, the inflation rate, the deposit rate competition, lagged effects, etc. We can then use standard econometric and time-series tools for estimating the unknown parameters.*

7.3.2 Prepayment risk

A prepayment is the settlement of a debt or the partial repayment of its outstanding amount before its maturity date. It is an important risk for the ALM of a bank, because it highly impacts the net interest income and the efficiency of the hedging portfolio. For example, suppose that the bank has financed a 10-year mortgage paying 5% through a 10-year bond paying 4%. The margin on this mortgage is equal to 1%. Five years later, the borrower prepays the mortgage because of a fall in interest rates. In this case, the bank receives the cash of the mortgage refund whereas it continues to pay a coupon of 4%. Certainly, the cash will yield a lower return than previously, implying that the margin is reduced and may become negative.

Prepayment risk shares some common features with default risk. Indeed, the prepayment time can be seen as a stopping time exactly like the default time for credit risk. Prepayment and default are then the two actions that may terminate the loan contract. This is why they have been studied together in some research. However, they also present some strong differences. In the case of the default risk, the income of the bank is reduced because both interest and capital payments are shut down. In the case of the prepayment risk, the bank recovers the capital completely, but no longer receives the interest due. Moreover, while default risk increases when the economic environment is bad or interest rates are high, prepayment risk is more pronounced in a period of falling interest rates.

In the 1980s, prepayment has been extensively studied in the case of RMBS. The big issue was to develop a pricing model for GNMA[64] mortgage-backed pass-through securities (Dunn and McConnell, 1981; Brennan and Schwartz, 1985; Schwartz and Torous, 1989). In these approaches, the prepayment option is assimilated to an American call option and the objective of the borrower is to exercise the option when it has the largest value[65] (Schwartz and Torous, 1992). However, Deng *et al.* (2000) show that "*there exists significant heterogeneity among mortgage borrowers and ignoring this heterogeneity results in serious errors in estimating the prepayment behavior of homeowners*". Therefore, it is extremely difficult to model the prepayment behavior, because it is not always a rational decision and many factors affect prepayment decisions (Keys *et al.*, 2016; Chernov *et al.*, 2017). This microeconomic approach is challenged by a macroeconomic approach, whose goal is to model the prepayment rate at the portfolio level and not the prepayment time at the loan level.

[64] The Government National Mortgage Association (GNMA or Ginnie Mae) has already been presented on page 139.

[65] This implies that the call option is in the money.

In what follows, we focus on mortgage loans, because it is the main component of prepayment risk. However, the analysis can be extended to other loans, for example consumer credit, student loans and leasing contracts. The case of student loans is very interesting since students are looking forward to repay their loan as soon as possible once they have found a job and make enough money.

7.3.2.1 Factors of prepayment

Following Hayre *et al.* (2000), prepayments are caused by two main factors: refinancing and housing turnover. Let i_0 be the original interest rate of the mortgage or the loan. We note $i(t)$ the interest rate of the same mortgage if the household would finance it at time t. It is clear that the prepayment time $\boldsymbol{\tau}$ depends on the interest rate differential, and we can assume that the prepayment probability is an increasing function of the difference $\Delta i(t) = i_0 - i(t)$:

$$\mathbb{P}(t) = \Pr\{\boldsymbol{\tau} \leq t\} = \vartheta(i_0 - i(t))$$

where $\partial_x \vartheta(x) > 0$. For instance, if the original mortgage interest rate is equal to 10% and the current mortgage interest rate is equal to 0%, nobody benefits from keeping the original mortgage, and it is preferable to fully refinance the mortgage. This situation is particularly true in a period of falling interest rates. The real life example provided by Keys *et al.* (2016) demonstrates the strong implication that a prepayment may have on household budgeting:

> "*A household with a 30-year fixed-rate mortgage of* \$200 000 *at an interest rate of* 6.0% *that refinances when rates fall to* 4.5% *(approximately the average rate decrease between 2008 and 2010 in the US) saves more than* \$60 000 *in interest payments over the life of the loan, even after accounting for refinance transaction costs. Further, when mortgage rates reached all-time lows in late 2012, with rates of roughly* 3.35% *prevailing for three straight months, this household with a contract rate of* 6.5% *would save roughly* \$130 000 *over the life of the loan by refinancing*" (Keys *et al.*, 2016, pages 482-483).

As already said, the prepayment value is the premium of an American call option, meaning that we can derive the optimal option exercise. In this case, the prepayment strategy can be viewed as an arbitrage strategy between the market interest rate and the cost of refinancing. In practice, we observe that the prepayment probability $\mathbb{P}(t)$ depends on other factors: loan type, loan age, loan balance, monthly coupon (Elie *et al.*, 2002). For example, it is widely accepted that the prepayment probability is an increasing function of the monthly coupon.

The second factor for explaining prepayments is housing turnover. In this case, the prepayment decision is not motivated by refinancing, but it is explained by the home sale due to life events. For instance, marriage, divorce, death, children leaving home or changing jobs explain a large part of prepayment rates. Another reason is the housing market dynamics, in particular home prices that have an impact on housing turnover. These different factors explain that we also observe prepayments even when interest rates increase. For example, the upgrading housing decision (i.e. enhancing the capacity or improving the quality of housing) is generally explained by the birth of a new child, an inheritance or a salary increase.

Remark 87 *In addition to these two main factors, we also observe that some borrowers choose to reduce their debt even if it is not an optimal decision. When they have some financial saving, which may be explained by an inheritance for example, they proceed to partial prepayments.*

7.3.2.2 Structural models

As with the credit risk, there are two families of prepayment models. The objective of structural models is to explain the prepayment time $\boldsymbol{\tau}$ of a borrower while reduced-form models are interested in the prepayment rate of a loan portfolio.

Value of the American option The objective is to find the optimal value $\boldsymbol{\tau}$ such that the borrower minimizes the paid cash flows or maximizes the prepayment option. Let us consider a mortgage, whose maturity is equal to T. In continuous-time, the risk-neutral value of cash flows is equal to[66]:

$$V(t) = \inf_{\boldsymbol{\tau} \leq T} \mathbb{E}^{\mathbb{Q}} \left[\int_t^{\boldsymbol{\tau}} m(u) \, e^{-\int_t^u r(s)\,\mathrm{d}s} \,\mathrm{d}u + e^{-\int_t^{\boldsymbol{\tau}} r(s)\,\mathrm{d}s} M(\boldsymbol{\tau}) \mid \mathcal{F}_t \right] \tag{7.39}$$

where $m(t)$ and $M(t)$ are the coupon and the mark-to-market value of the mortgage at time t. The first term that makes up $V(t)$ is the discounted value of the interest paid until the prepayment time $\boldsymbol{\tau}$ whereas the second term is the discounted value of the mortgage value at the prepayment time $\boldsymbol{\tau}$. Equation (7.39) is a generalization of the net present value of a mortgage in continuous-time[67]. The computation of the optimal stopping time can be done in a Hamilton-Jacobi-Bellman (HJB) framework. We introduce the state variable X_t, which follows a diffusion process:

$$\mathrm{d}X(t) = \mu(t, X(t)) \,\mathrm{d}t + \sigma(t, X(t)) \,\mathrm{d}W(t)$$

We note $V(t, X)$ the value of $V(t)$ when $X(t)$ is equal to X. In the absence of prepayment, we deduce that the value of $V(t, X)$ satisfies the following Cauchy problem[68]:

$$\begin{cases} -\partial_t V(t, X) + r(t) V(t, X) = \mathcal{A}_t V(t, X) + m(t) \\ V(T, X) = M(T) \end{cases}$$

where $\mathcal{A}_t$ is the infinitesimal generator of the diffusion process:

$$\mathcal{A}_t V(t, X) = \frac{1}{2}\sigma^2(t, X) \frac{\partial^2 V(t, X)}{\partial X^2} + \mu(t, X) \frac{\partial^2 V(t, X)}{\partial x^2}$$

The prepayment event changes the previous problem since we must verify that the value $V(t, X)$ is lower than the mortgage value $M(t)$ minus the refinancing cost $C(t)$. The option problem is then equivalent to solve the HJB equation or the variational inequality:

$$\min\left(\mathcal{L}_t V(t, X), V(t, X) + C(t) - M(t)\right) = 0$$

where:

$$\mathcal{L}_t V(t, X) = \mathcal{A}_t V(t, X) + m(t) + \partial_t V(t, X) - r(t) V(t, X)$$

This model can be extended to the case where there are several state variables or there is no maturity (perpetual mortgage).

---

[66] $r(t)$ is the discount rate.

[67] The net present value is equal to:

$$V(t) = \mathbb{E}^{\mathbb{Q}} \left[\int_t^{T} m(u) \, e^{-\int_t^u r(s)\,\mathrm{d}s} \,\mathrm{d}u + e^{-\int_t^{T} r(s)\,\mathrm{d}s} N(T) \mid \mathcal{F}_t \right]$$

where $N(T)$ is the outstanding amount at the maturity.

[68] We use the Feynmac-Kac representation given on page 1070.

The Agarwal-Driscoll-Laibson model There are several possible specifications depending on the choice of the state variables, the dynamics of interest rates, etc. For example, using a similar framework than previously, Agarwal *et al.* (2013) propose the following optimal refinancing rule:

$$i_0 - i(t) \geq \delta^\star = \frac{1}{\psi}\left(\phi + W\left(-e^{-\phi}\right)\right) \tag{7.40}$$

where $W(x)$ is the Lambert W function[69], $\psi = \sigma^{-1}\sqrt{2(r+\lambda)}$ and $\phi = 1+\psi(r+\lambda)(C/M)$. The parameters are the real discount rate r, the rate λ of exogenous mortgage prepayment, the volatility σ of the mortgage rate $i(t)$, the refinancing cost C and the remaining mortgage value M. Equation (7.40) has been obtained by solving the HJB equation and assuming that $\mathrm{d}X(t) = \sigma\,\mathrm{d}W(t)$ and $X(t) = i(t) - i_0$.

Using the numerical values $r = 5\%$, $\lambda = 10\%$, $\sigma = 2\%$, and $C/M = 1\%$, $\delta^\star$ is equal to 110 bps. This means that the borrower has to prepay his mortgage if the mortgage rate falls by at least 110 bps. In Table 7.20, we consider the impact of one parameter by considering the other parameters unchanged. First, we assume that the cost function is $C = 2000 + 1\% \times M$, meaning that there is a fixed cost of \$2 000. It follows that $\delta^\star$ is a decreasing function of the mortgage value M, because fixed costs penalize low mortgage values. We also verify that $\delta^\star$ is an increasing function of r, σ and λ. In particular, the parameter σ has a big influence, because it indicates if the mortgage rate is volatile or not. In the case of a high volatility, it may be optimal that the borrower is waiting that $i(t)$ highly decreases. This is why the HJB equation finds a high value of $\delta^\star$.

TABLE 7.20: Optimal refinancing rule $\delta^\star$

M (in KUSD)	$\delta^\star$	r	$\delta^\star$	σ	$\delta^\star$	λ	$\delta^\star$
10	612	1%	101	1%	79	2%	89
100	198	2%	103	2%	110	5%	98
250	150	5%	110	3%	133	10%	110
500	131	8%	116	5%	171	15%	120
1 000	121	10%	120	10%	239	20%	128

7.3.2.3 Reduced-form models

Rate, coupon or maturity incentive? The previous approach can only be applied to the refinancing decision, but it cannot deal with all types of prepayment. Moreover, there is no guarantee that the right decision variable is the difference $i_0 - i(t)$ between the current mortgage rate and the initial mortgage rate. For instance, $i_0 - i(t) = 1\%$ implies a high impact for a 20-year remaining maturity, but has a small effect when the maturity is less than one year. A better decision variable is the coupon or annuity paid by the borrower. In the case of a constant payment mortgage, we recall that the annuity is equal to:

$$A(i,n) = \frac{i}{1-(1+i)^{-n}} N_0$$

where N_0 is the notional of the mortgage, i is the mortgage rate and n is the number of periods. If the mortgage rate drops from i_0 to $i(t)$, the absolute difference of the annuity is equal to $\mathfrak{D}_A(i_0, i(t)) = A(i_0,n) - A(i(t),n)$, whereas the relative difference of the annuity

[69] The Lambert W function is related to Shannon's entropy and satisfies $W(x)\,e^{W(x)} = x$.

is given by:

$$\begin{aligned}\mathfrak{D}_R(i_0, i(t)) &= \frac{\mathfrak{D}_A(i_0, i(t))}{A(i_0, n)} \\ &= 1 - \left(\frac{1-(1+i_0)^{-n}}{1-(1+i(t))^{-n}}\right)\frac{i(t)}{i_0}\end{aligned}$$

where n is the remaining number of periods. In a similar way, the relative cumulative difference $\mathfrak{C}(i_0, i(t))$ is equal to:

$$\begin{aligned}\mathfrak{C}(i_0, i(t)) &= \frac{\sum_{t=1}^{n}\mathfrak{D}_A(i_0, i(t))}{N_0} \\ &= n\left(\frac{i_0}{1-(1+i_0)^{-n}} - \frac{i(t)}{1-(1+i(t))^{-n}}\right)\end{aligned}$$

Finally, another interesting measure is the minimum number of periods $\mathfrak{N}(i_0, i(t))$ such that the new annuity is greater than or equal to the initial annuity:

$$\mathfrak{N}(i_0, i(t)) = \{x \in \mathbb{N} : A(i(t), x) \geq A(i(t), n), A(i(t), x+1) < A(i(t), n)\}$$

where $\mathfrak{N}(i_0, i(t))$ measures the maturity reduction of the loan by assuming that the borrower continues to pay the same annuity.

TABLE 7.21: Impact of a new mortgage rate (100 KUSD, 5%, 10-year)

i	A	$\mathfrak{D}_A$ (in \$)		$\mathfrak{D}_R$	$\mathfrak{C}$	$\mathfrak{N}$
(in %)	(in \$)	Monthly	Annually	(in %)	(in %)	(in years)
5.0	1 061					
4.5	1 036	24	291	2.3	2.9	9.67
4.0	1 012	48	578	4.5	5.8	9.42
3.5	989	72	862	6.8	8.6	9.17
3.0	966	95	1 141	9.0	11.4	8.92
2.5	943	118	1 415	11.1	14.2	8.75
2.0	920	141	1 686	13.2	16.9	8.50
1.5	898	163	1 953	15.3	19.5	8.33
1.0	876	185	2 215	17.4	22.2	8.17
0.5	855	206	2 474	19.4	24.7	8.00

Let us illustrate the impact of a new rate $i(t)$ on an existing mortgage. We assume that the current outstanding amount is equal to \$100 000 and the amortization scheme is monthly. In Table 7.21, we show how the monthly annuity changes if the original rate is 5% and the remaining maturity is ten years. If the borrower refinances the mortgage at 2%, the monthly annuity is reduced by \$141, which represents 13.2% of the current monthly coupon. His total gain is then equal to 16.9% of the outstanding amount. If the borrower prefers to reduce the maturity and takes the annuity constant, he will gain 18 months. In Tables 7.22 and 7.23, we compute the same statistics when the remaining maturity is twenty years or the original rate is 10%. Banks have already experienced this kind of situation these last 30 years. For example, we report the average rate of 30-year and 15-year fixed rate mortgages in the US in Figure 7.20. We also calculate the differential rate between the 30-year mortgage rate lagged 15 years and the 15-year mortgage rate. We notice that this refinancing opportunity has reached 10% and more in the 1990s, and was above 3% most of the times these last 25 years. Of course, this situation is exceptional and explained by 30 years of falling interest rates.

TABLE 7.22: Impact of a new mortgage rate (100 KUSD, 5%, 20-year)

i (in %)	A (in \$)	$\mathfrak{D}_A$ (in \$) Monthly	Annually	$\mathfrak{D}_R$ (in %)	$\mathfrak{C}$ (in %)	$\mathfrak{N}$ (in years)
5.0	660					
4.5	633	27	328	4.1	6.6	18.67
4.0	606	54	648	8.2	13.0	17.58
3.5	580	80	960	12.1	19.2	16.67
3.0	555	105	1 264	16.0	25.3	15.83
2.5	530	130	1 561	19.7	31.2	15.17
2.0	506	154	1 849	23.3	37.0	14.50
1.5	483	177	2 129	26.9	42.6	14.00
1.0	460	200	2 401	30.3	48.0	13.50
0.5	438	222	2 664	33.6	53.3	13.00

TABLE 7.23: Impact of a new mortgage rate (100 KUSD, 10%, 10-year)

i (in %)	A (in \$)	$\mathfrak{D}_A$ (in \$) Monthly	Annually	$\mathfrak{D}_R$ (in %)	$\mathfrak{C}$ (in %)	$\mathfrak{N}$ (in years)
10.0	1 322					
9.0	1 267	55	657	4.1	6.6	9.33
8.0	1 213	108	1 299	8.2	13.0	8.75
7.0	1 161	160	1 925	12.1	19.3	8.33
6.0	1 110	211	2 536	16.0	25.4	7.92
5.0	1 061	261	3 130	19.7	31.3	7.58
4.0	1 012	309	3 709	23.3	37.1	7.25
3.0	966	356	4 271	26.9	42.7	6.92
2.0	920	401	4 816	30.4	48.2	6.67
1.0	876	445	5 346	33.7	53.5	6.50

Survival function with prepayment risk Previously, we have defined the amortization function $\mathbf{S}(t,u)$ as the fraction of the new production at time t that still remains in the balance sheet at time $u \geq t$: $\text{NP}(t,u) = \text{NP}(t)\,\mathbf{S}(t,u)$. We have seen that $\mathbf{S}(t,u)$ corresponds to a survival function. Therefore, we can use the property that the product of n_s survival functions is a survival function, meaning that we can decompose $\mathbf{S}(t,u)$ as follows:

$$\mathbf{S}(t,u) = \prod_{j=1}^{n_s} \mathbf{S}_j(t,u)$$

This implies that the hazard rate is an additive function:

$$\lambda(t,u) = \sum_{j=1}^{n_s} \lambda_j(t,u)$$

because we have:

$$e^{-\int_t^u \lambda(t,s)\,\mathrm{d}s} = \prod_{j=1}^{n_s} e^{-\int_t^u \lambda_j(t,s)\,\mathrm{d}s} = e^{-\int_t^u \left(\sum_{j=1}^{n_s} \lambda_j(t,s)\right)\mathrm{d}s}$$

If we apply this result to prepayment, we have:

$$\mathbf{S}(t,u) = \mathbf{S}_c(t,u) \cdot \mathbf{S}_p(t,u)$$

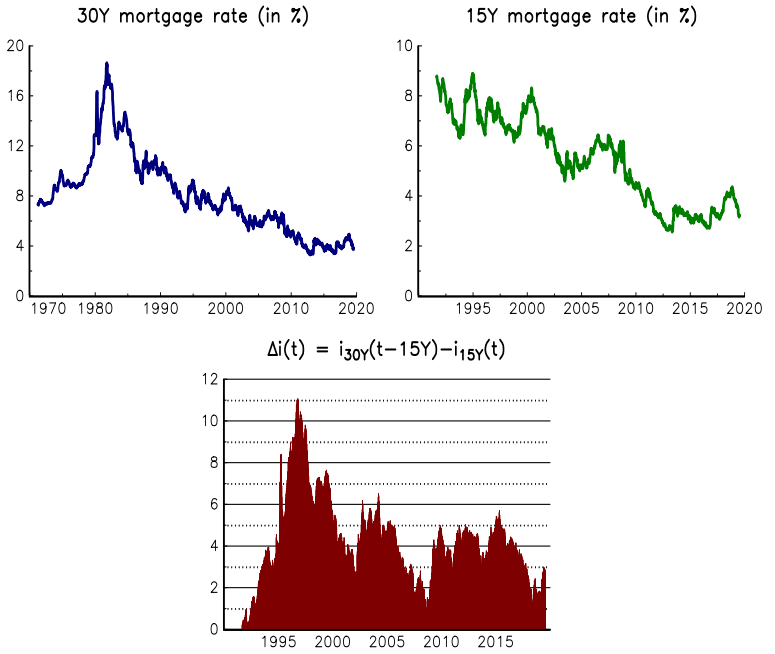


FIGURE 7.20: Evolution of 30-year and 15-year mortgage rates in the US

Source: Freddie Mac, 30Y/15Y Fixed Rate Mortgage Average in the United States [MORTGAGE30US/15US], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MORTGAGE30US, July 24, 2019.

where $\mathbf{S}_c(t,u)$ is the traditional amortization function (or the contract-based survival function) and $\mathbf{S}_p(t,u)$ is the prepayment-based survival function.

Example 75 *We consider a constant amortization mortgage (CAM) and assume that the prepayment-based hazard rate is constant and equal to* λ_p.

In Exercise 7.4.3 on page 450, we show that the survival function is equal to:

$$\mathbf{S}_c(t,u) = \mathbb{1}\{t \leq u \leq t+m\} \cdot \frac{1-e^{-i(t+m-u)}}{1-e^{-im}}$$

It follows that:

$$\begin{aligned}
\lambda_c(t,u) &= -\frac{\partial \ln \mathbf{S}_c(t,u)}{\partial u} \\
&= \frac{\partial \ln\left(1-e^{-im}\right)}{\partial u} - \frac{\partial \ln\left(1-e^{-i(t+m-u)}\right)}{\partial u} \\
&= \frac{ie^{-i(t+m-u)}}{1-e^{-i(t+m-u)}} \\
&= \frac{i}{e^{i(t+m-u)}-1}
\end{aligned}$$

Finally, we deduce that:

$$\lambda(t,u) = \mathbb{1}\{t \leq u \leq t+m\} \cdot \left(\frac{i}{e^{i(t+m-u)}-1} + \lambda_p\right)$$

In Figure 7.21, we report the survival function $\mathbf{S}(t,u)$ and the hazard rate $\lambda(t,u)$ of a 30-year mortgage at 5%. We also compare the amortization function $\mathbf{S}(t,u)$ obtained in continuous-time with the function calculated when we assume that the coupon is paid monthly. We notice that the continuous-time model is a good approximation of the discrete-time model.

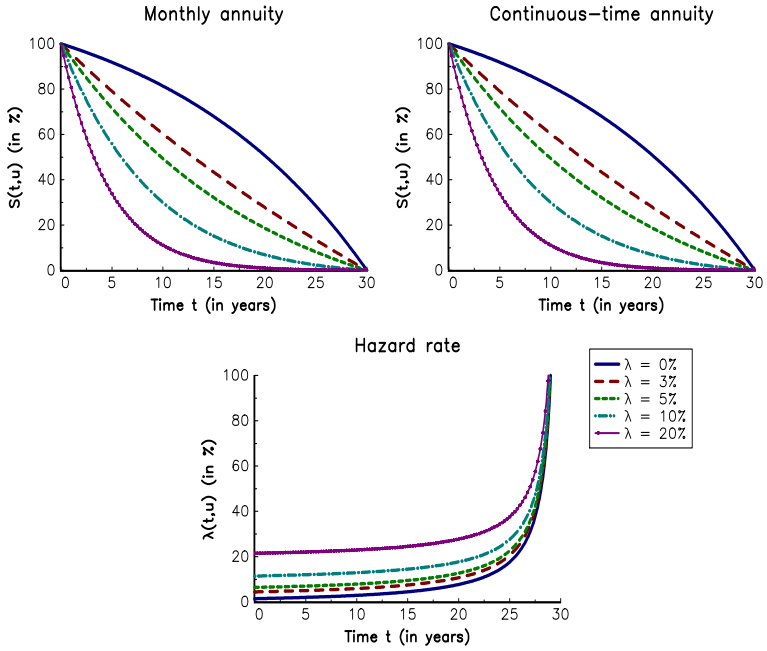


FIGURE 7.21: Survival function in the case of prepayment

Specification of the hazard function It is unrealistic to assume that the hazard function $\lambda_p(t,u)$ is constant because we do not make the distinction between economic and structural prepayments. In fact, it is better to decompose $\mathbf{S}_p(t,u)$ into the product of two survival functions:

$$\mathbf{S}_p(t,u) = \mathbf{S}_{\text{refinancing}}(t,u) \cdot \mathbf{S}_{\text{turnover}}(t,u)$$

where $\mathbf{S}_{\text{refinancing}}(t,u)$ corresponds to economic prepayments due to refinancing decisions and $\mathbf{S}_{\text{turnover}}(t,u)$ corresponds to structural prepayments because of housing turnover. In this case, we can assume that $\lambda_{\text{turnover}}(t,u)$ is constant and corresponds to the housing turnover rate. The specification of $\lambda_{\text{refinancing}}(t,u)$ is more complicated since it depends on several factors. For instance, Elie *et al.* (2002) show that $\lambda_{\text{refinancing}}(t,u)$ depends on the loan characteristics (type, age and balance), the cost of refinancing and the market rates. Moreover, they observe a seasonality in prepayment rates, which differs with respect to the loan type (monthly, quarterly or semi-annually).

As for deposit balances, the '*Net Portfolio Value Model*' published by the Office of Thrift Supervision (2001) gives very precise formulas for measuring prepayment. They assume that the prepayment rate is made up of three factors:

$$\lambda_p(t,u) = \lambda_{\text{age}}(u-t) \cdot \lambda_{\text{seasonality}}(u) \cdot \lambda_{\text{rate}}(u)$$

where λ_{age} measures the impact of the loan age, $\lambda_{\text{seasonality}}$ corresponds to the seasonality factor and λ_{rate} represents the influence of market rates. The first two components are

specified as follows:

$$\lambda_{\text{age}}(\text{age}) = \begin{cases} 0.4 \cdot \text{age} & \text{if} \quad \text{age} \leq 2.5 \\ 1 & \text{if} \quad \text{age} \geq 2.5 \end{cases}$$

and:

$$\lambda_{\text{seasonality}}(u) = 1 + 0.20 \times \sin\left(1.571 \times \left(\frac{12 + \text{month}(u) - 3}{3}\right) - 1\right)$$

where $\text{age} = u - t$ is the loan age and $\text{month}(u)$ is the month of the date u. We notice that λ_{age} is equal to zero for a new mortgage – $u - t = 0$, increases linearly with mortgage age and remains constant after 30 months or 2.5 years. The refinancing factor of the OTS model has the following expression:

$$\lambda_{\text{rate}}(u) = \beta_0 + \beta_1 \arctan\left(\beta_2 \cdot \left(\beta_3 - \frac{i_0}{i(u - 0.25)}\right)\right)$$

where $i(u - 0.25)$ is the mortgage refinancing rate (lagged three months). In Figure 7.22, we represent the three components[70] while Figure 7.23 provides an example of the survival function $\mathbf{S}_p(t, u)$ where the mortgage rate drops from 5% to 1% after 6 years. The seasonality component has a small impact on the survival function because it is smoothed when computing the cumulative hazard function. On the contrary, the age and rate components change the prepayment speed.

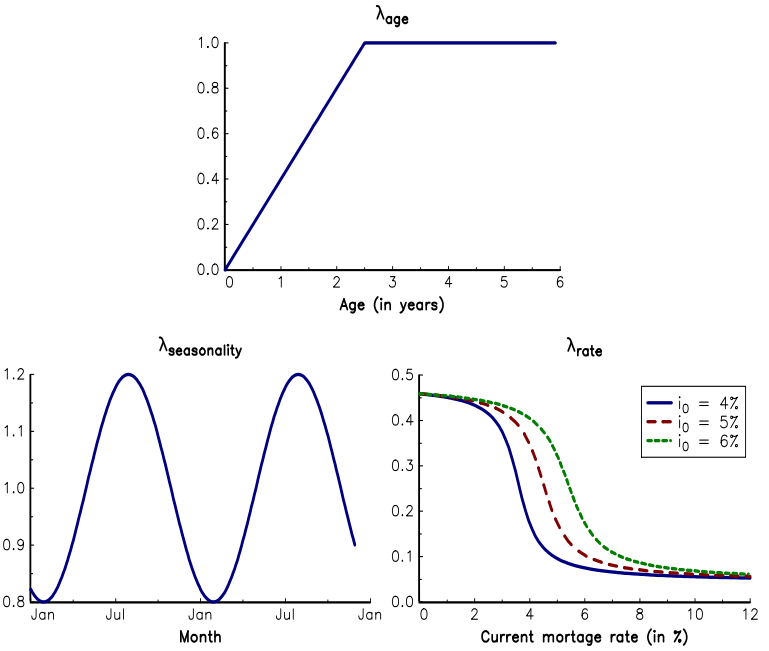


FIGURE 7.22: Components of the OTC model

[70] For the specification of λ_{rate}, we use the default values of OTS (2001, Equation 5.A.7): $\beta_0 = 0.2406$, $\beta_1 = -0.1389$, $\beta_2 = 5.952$, and $\beta_4 = 1.049$. We also assume that $i_0 = 5\%$.

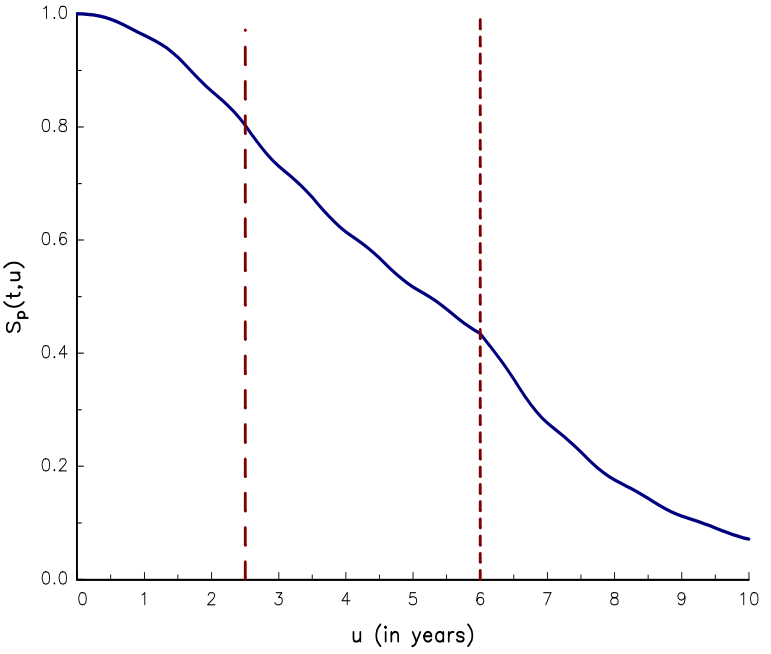


FIGURE 7.23: An example of survival function $\mathbf{S}_p(t,u)$ with a mortgage rate drop

7.3.2.4 Statistical measure of prepayment

In fact, the OTC model doesn't use the concept of hazard rate, but defines the constant prepayment rate CPR, which is the annualized rate of the single monthly mortality:

$$\text{SMM} = \frac{\text{prepayments during the month}}{\text{outstanding amount at the beginning of the month}}$$

The CPR and the SMM are then related by the following equation:

$$\text{CPR} = (1 - (1 - \text{SMM}))^{12}$$

In IRRBB, the CPR is also known as the conditional prepayment rate. It measures prepayments as a percentage of the current outstanding balance for the next year. By definition, it is related to the hazard function as follows:

$$\begin{aligned}
\text{CPR}(u,t) &= \Pr\{u < \boldsymbol{\tau} \leq u+1 \mid \boldsymbol{\tau} \geq u\} \\
&= \frac{\mathbf{S}_p(t,u) - \mathbf{S}_p(t,u+1)}{\mathbf{S}_p(t,u)} \\
&= 1 - \exp\left(-\int_u^{u+1} \lambda_p(t,s)\,\mathrm{d}s\right)
\end{aligned}$$

If $\lambda_p(t,s)$ is constant and equal to λ_p, we can approximate the CPR by the hazard rate λ_p because we have $\text{CPR}(u,t) \approx 1 - e^{-\lambda_p} \approx \lambda_p$.

We use the prepayment monitoring report published by the Federal Housing Finance Agency (FHFA). From 2008 to 2018, the CPR for 30-year mortgages varies between 5% to 35% in the US. The lowest value is reached at the end of 2008. This shows clearly that prepayments depend on the economic cycle. During a crisis, the number of defaults increases

while the number of prepayments decreases. This implies that there is a negative correlation between default and prepayment rates. However, there is a high heterogeneity depending on the coupon rate and the issuance date as shown in Table 7.24. We generally observe that the CPR increases with the coupon rate. For example, in June 2018, the CPR is 7% greater for a 30-year mortgage issued between 2012 and 2016 with a 4.5% coupon than with a 3% coupon. We also verify the ramp effect because the prepayment rate is not of the same magnitude before and after January 2017, which corresponds to the 30-month age after which the prepayment rate can be assumed to be constant. This is why the CPR is only 5.3% and 12.8% for mortgages issued in 2018 and 2017 while it is equal to 17.4% for mortgages issued in 2016 when the coupon rate is 4.5%.

TABLE 7.24: Conditional prepayment rates in June 2018 by coupon rate and issuance date

Year	2012	2013	2014	2015	2016	2017	2018
Coupon = 3%	9.6%	10.2%	10.9%	10.0%	8.7%	5.3%	3.1%
Coupon = 4.5%	16.1%	15.8%	16.6%	17.9%	17.4%	12.8%	5.3%
Difference	6.5%	5.6%	5.7%	8.0%	8.7%	7.6%	2.2%

Source: RiskSpan dataset, FHFA (2018) and author's calculations.

7.3.3 Redemption risk

7.3.3.1 The funding risk of term deposits

A term deposit, also known as time deposit or certificate of deposit (CD), is a fixed-term cash investment. The client deposits a minimum sum of money into a banking account in exchange for a fixed rate over a specified period. A term deposit is then defined by three variables: the deposit or CD rate $i(t)$, the maturity period m and the minimum balance D^-. For example, the minimum deposit is generally \$1 000 in the US, and the typical maturities are 1M, 3M, 6M, 1Y, 2Y and 3Y. In some banks, the deposit rate may depends on the deposit amount[71]. Term deposits are an important source of bank funding with demand deposits and savings accounts. However, they differ from non-maturity deposits because they have a fixed maturity, their rates are higher and they may be redeemed with a penalty. When buying a term deposit, the investor can withdraw their funds only after the term ends. This is why CD rates are generally greater than NMD rates, because term deposits are a most stable funding resource for banks. Moreover, CD rates are generally more sensitive to market interest rates than NMD rates, because a term deposit is more an investment product while a demand deposit is more a transaction account. Under some conditions, the investor may withdraw his term deposit before the maturity date if he pays early redemption costs and fees, which generally correspond to a reduction of the deposit rate. For example, $i(t)$ may be reduced by 80% if the remaining maturity is greater than 50% of the CD maturity and 30% if the remaining maturity is less than 20% of the CD maturity.

According to Gilkeson *et al.* (1999), early time deposit withdrawals may be motivated by two reasons. As for prepayments, the first reason is economic. If market interest rates rise, the investor may have a financial incentive to close his old term deposit and reinvest his

[71] For example, Chase defines six CD rates for a given maturity and considers the following bands: below \$10K, \$10K – \$25K, \$25K – \$50K, \$50K – \$100K, \$100K – \$250K and \$250+ (source: https://www.chase.com/personal/savings/bank-cd).

money into a new term deposit. In this case, the investor is sensitive to the rate differential $i(t) - i_0$ where i_0 is the original CD rate and $i(t)$ is the current CD rate. In this case, early withdrawal risk can be viewed as the opposite of prepayment risk. Indeed, while the economic reason of prepayment risk is a fall of interest rates, the economic reason of redemption risk is a rise of interest rates. Since both risks imply a negative impact on the net interest income, the impact on the liquidity risk is different: the bank receives cash in case of a prepayment, while the funding of the bank is reduced in case of redemption. The second reason is related to negative liquidity shocks of depositors. For example, the client may need to get his money back because of life events: job loss, divorce, revenue decline, etc. In this case, redemption risk is explained by idiosyncratic liquidity shocks that are independent and can be measured by a structural constant rate. But redemption risk can also be explained by systemic liquidity shocks. For example, economic crises increase the likelihood of early withdrawals. In this case, we cannot assume that the redemption rate is constant because it depends on the economic cycle.

7.3.3.2 Modeling the early withdrawal risk

Redemption risk can be measured using the same approach we have used for prepayment risk. This is particularly true for the economic component and the idiosyncratic liquidity component. The systemic component of negative liquidity shocks requires a more appropriate analysis and makes the modeling more challenging. Another difficulty with the early withdrawal risk is the scarcity of academic models, professional publications and data. To our knowledge, there are only five academic publications on this topic and only three articles that give empirical results[72]: Cline and Brooks (2004), Gilkeson *et al.* (1999) and Gilkeson *et al.* (2000).

The redemption-based survival function of time deposits can be decomposed as:

$$\mathbf{S}_r(t,u) = \mathbf{S}_{\text{economic}}(t,u) \cdot \mathbf{S}_{\text{liquidity}}(t,u)$$

where $\mathbf{S}_{\text{economic}}(t,u)$ is the amortization function related to reinvestment financial incentives and $\mathbf{S}_{\text{liquidity}}(t,u)$ is the amortization function due to negative liquidity shocks.

Let us first focus on economic withdrawals. We note t the current date, m the maturity of the time deposit and N_0 the initial investment at time 0. In the absence of redemption, the value of the time deposit at the maturity is equal to $V_0 = N_0(1+i_0)^m$. If we assume that $\boldsymbol{\tau}$ is the withdrawal time, the value of the investment for $\boldsymbol{\tau} = t$ becomes:

$$V_r(t) = N_0 \cdot (1 + (1 - \varphi(t))\, i_0)^t \cdot (1 + i(t))^{m-t} - C(t)$$

where $\varphi(t)$ is the penalty parameter applied to interest paid and $C(t)$ is the break fee. For example, if we specify $\varphi(t) = 1 - t/m$, $\varphi(t)$ is a linear decreasing function between[73] $\varphi(0) = 100\%$ and $\varphi(m) = 0\%$. $C(t)$ may be a flat fee (e.g. $C(t) = \$1\,000$) or $C(t)$ may be a proportional fee: $C(t) = c(t) \cdot N_0$. The rational investor redeems the term deposit if the refinancing incentive is positive:

$$\text{RI}(t) = \frac{V_r(t) - V_0}{N_0} > 0$$

In the case where $C(t) = c(t) N_0$, we obtain the following equivalent condition:

$$i(t) > i^\star(t) = \left(\frac{(1+i_0)^m + c(t)}{(1 + (1 - \varphi(t))\, i_0)^t} \right)^{1/(m-t)} - 1$$

[72] The two other theoretical publications are Stanhouse and Stock (2004), and Gao *et al.* (2018).

[73] $\varphi(t) = 100\%$ if the redemption occurs at the beginning of the contract and $\varphi(m) = 0\%$ when the term deposit matures.

An example of this refinancing incentive rule is given in Figure 7.24. This corresponds to a three-year term deposit whose rate is equal to 2%. The penalty applied to interest paid is given by $\varphi(t) = 1 - t/m$. We show the impact of the fee $c(t)$ on $i^\star(t)$. We observe that the investor has no interest to wait if the interest rate rise is sufficient. Therefore, there is an arbitrage between the current rate $i(t)$ and the original rate i_0. We deduce that the hazard function takes the following form: $\lambda_{\text{economic}}(t,u) = g(i(u) - i_0)$ or $\lambda_{\text{economic}}(t,u) = g(r(u) - i_0)$ where g is a function to estimate. For instance, Gilkeson *et al.* (1999) consider a logistic regression model and explain withdrawal rates by the refinancing incentive variable.

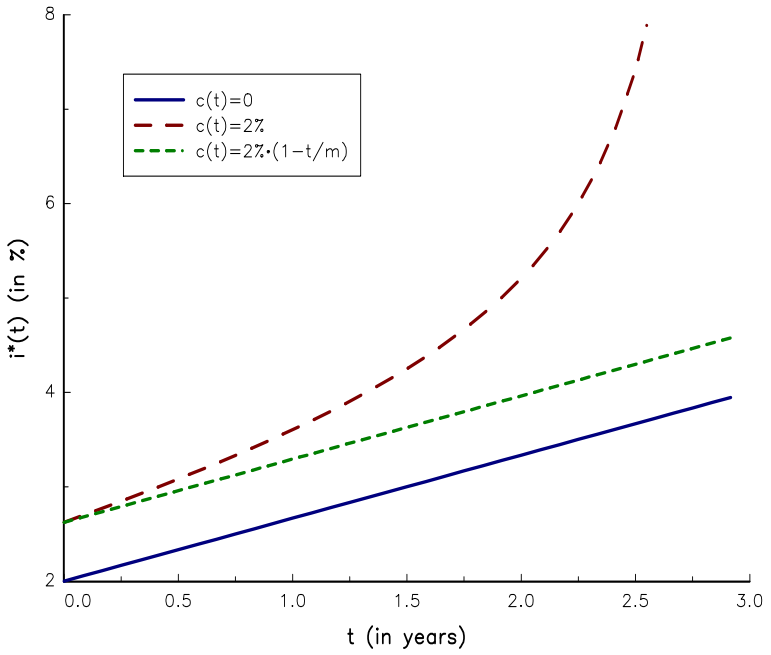


FIGURE 7.24: Refinancing incentive rule of term deposits

For early withdrawals due to negative liquidity shocks, we can decompose the hazard function into two effects:

$$\lambda_{\text{liquidity}}(t,u) = \lambda_{\text{structural}} + \lambda_{\text{cyclical}}(u)$$

where $\lambda_{\text{structural}}$ is the structural rate of redemption and $\lambda_{\text{cyclical}}(u)$ is the liquidity component due to the economic cycle. A simple way to model $\lambda_{\text{cyclical}}(u)$ is to consider a linear function of the GDP growth.

7.4 Exercises

7.4.1 Constant amortization of a loan

We consider a loan that is repaid by annual payments. We assume that the notional of the loan is equal to N_0, the maturity of the loan is n and i is the annual interest rate. We note $N(t)$ the outstanding amount, $I(t)$ the interest payment, $P(t)$ the principal payment at time t and $C(t)$ the present value.

1. Let C_0 be the present value of an annuity A that is paid annually during n years. Calculate C_0 as a function of A, n and i.

2. Determine the constant annuity A of the loan and the corresponding annuity rate $a_{(n)}$.

3. Calculate $I(1)$ and $P(1)$. Show that the outstanding amount $N(1)$ is equal to the present value $C(1)$ of the constant annuity A for the last $n-1$ years.

4. Calculate the general formula of $N(t)$, $I(t)$ and $P(t)$.

7.4.2 Computation of the amortization functions $\mathbf{S}(t,u)$ and $\mathbf{S}^\star(t,u)$

In what follows, we consider a debt instrument, whose remaining maturity is equal to m. We note t the current date and $T = t + m$ the maturity date.

1. We consider a bullet repayment debt. Define its amortization function $\mathbf{S}(t,u)$. Calculate the survival function $\mathbf{S}^\star(t,u)$ of the stock. Show that:

$$\mathbf{S}^\star(t,u) = \mathbb{1}\{t \leq u < t+m\} \cdot \left(1 - \frac{u-t}{m}\right)$$

   in the case where the new production is constant. Comment on this result.

2. Same question if we consider a debt instrument, whose amortization rate is constant.

3. Same question if we assume[74] that the amortization function is exponential with parameter λ.

4. Find the expression of $\mathcal{D}^\star(t)$ when the new production is constant.

5. Calculate the durations $\mathcal{D}(t)$ and $\mathcal{D}^\star(t)$ for the three previous cases.

6. Calculate the corresponding dynamics $\mathrm{d}N(t)$.

7.4.3 Continuous-time analysis of the constant amortization mortgage (CAM)

We consider a constant amortization mortgage, whose maturity is equal to m. We note i the interest rate and A the constant annuity.

1. Let N_0 be the amount of the mortgage at time $t = 0$. Write the equation of $\mathrm{d}N(t)$. Show that the annuity is equal to:

$$A = \frac{i \cdot N_0}{1 - e^{-im}}$$

   Deduce that the outstanding balance at time t is given by:

$$N(t) = \mathbb{1}\{t < m\} \cdot N_0 \cdot \frac{1 - e^{-i(m-t)}}{1 - e^{-im}}$$

2. Find the expression of $\mathbf{S}(t,u)$ and $\mathbf{S}^\star(t,u)$.

3. Calculate the liquidity duration $\mathcal{D}(t)$.

[74] By definition of the exponential amortization, we have $m = +\infty$.

7.4.4 Valuation of non-maturity deposits

This exercise is based on the model of De Jong and Wielhouwer (2003), which is an application of the continuous-time framework of Jarrow and van Deventer (1998). The framework below has been used by de Jong and Wielhouwer to model variable rate savings accounts. However, it is valid for all types of non-maturity deposits (demand deposits and savings accounts). For instance, Jarrow and van Deventer originally develop the approach for all types of demand deposits[75].

1. Let $D(t)$ be the amount of savings accounts. We note $r(t)$ and $i(t)$ the market rate and the interest rate paid to account holders. We define the current market value of liabilities as follows:

$$L_0 = \mathbb{E}\left[\int_0^\infty e^{-r(t)t}\left(i(t)\,D(t) - \partial_t D(t)\right)\,\mathrm{d}t\right]$$

   Explain the expression of L_0, in particular the two components $i(t)\,D(t)$ and $\partial_t D(t)$.

2. By considering that the short rate $r(t)$ is constant, demonstrate that:

$$L_0 = D_0 + \mathbb{E}\left[\int_0^\infty e^{-r(t)t}\left(i(t) - r(t)\right)D(t)\,\mathrm{d}t\right]$$

3. Calculate the current mark-to-market V_0 of savings accounts. How do you interpret V_0?

4. Let us assume that the margin $m(t) = r(t) - i(t)$ is constant and equal to m_0, and $D(t)$ is at the steady state D_∞. Show that:

$$V_0 = m_0 \cdot r_\infty^{-1} \cdot D_\infty$$

   where r_∞ is a parameter to determine.

5. For the specification of the deposit rate $i(t)$ and the deposit balance $D(t)$, De Jong and Wielhouwer (2003) propose the following dynamics:

$$\mathrm{d}i(t) = \left(\alpha + \beta\left(r(t) - i(t)\right)\right)\mathrm{d}t$$

   and:

$$\mathrm{d}D(t) = \gamma\left(D_\infty - D(t)\right)\mathrm{d}t - \delta\left(r(t) - i(t)\right)\mathrm{d}t$$

   where α, $\beta \geq 0$, $\gamma \geq 0$ and $\delta \geq 0$ are four parameters. What is the rationale of these equations? Find the general expression of $i(t)$ and $D(t)$.

6. In the sequel, the market rate $r(t)$ is assumed to be constant and equal to r_0. Deduce the value of $i(t)$ and $D(t)$.

7. Calculate the net asset value V_0 and deduce its sensitivity with respect to the market rate r_0 when $\alpha = 0$.

8. Find the general expression of the sensitivity of V_0 with respect to the market rate r_0 when $\alpha \neq 0$. Deduce the duration $\mathcal{D}_D$ of the deposits.

[75] Janosi *et al.* (1999) provide an empirical analysis of the Jarrow-van Deventer model for negotiable orders of withdrawal accounts (NOW), passbook accounts, statement accounts and demand deposit accounts (DDAs), whereas Kalkbrener and Willing (2004) consider an application to savings accounts. Generally, these different accounts differ with respect to the specification of interest paid $i(t)$ and the dynamics of the deposit amount $D(t)$.

9. We consider a numerical application of the De Jong-Wielhouwer model with the following parameters: $r_0 = 10\%$, $i_0 = 5\%$, $D_0 = 100$, $D_\infty = 150$, $\beta = 0.5$, $\gamma = 0.7$ and $\delta = 0.5$. Make a graph to represent the relationship between the time t and the deposit rate $i(t)$ when α is equal to -1%, 0 and 1%. Why is it natural to consider that $\alpha < 0$? We now assume that $\alpha = -1\%$. Draw the dynamics of $D(t)$. What are the most important parameters that impact $D(t)$? What is the issue if we calculate the duration of the deposits with respect to r_0 when α is equal to zero? Make a graph to represent the relationship between the market rate r_0 and the duration when α is equal to -50 bps, -1% and -2%.

7.4.5 Impact of prepayment on the amortization scheme of the CAM

This is a continuation of Exercise 7.4.3 on page 450. We recall that the outstanding balance at time t is given by:

$$N(t) = \mathbb{1}\{t < m\} \cdot N_0 \cdot \frac{1 - e^{-i(m-t)}}{1 - e^{-im}}$$

1. Find the dynamics $\mathrm{d}N(t)$.
2. We note $\tilde{N}(t)$ the modified outstanding balance that takes into account the prepayment risk. Let $\lambda_p(t)$ be the prepayment rate at time t. Write the dynamics of $\tilde{N}(t)$.
3. Show that $\tilde{N}(t) = N(t) \cdot \mathbf{S}_p(t)$ where $\mathbf{S}_p(t)$ is the prepayment-based survival function.
4. Calculate the liquidity duration $\tilde{\mathcal{D}}(t)$ associated to the outstanding balance $\tilde{N}(t)$ when the hazard rate of prepayments is constant and equal to λ_p.

Chapter 8

Systemic Risk and Shadow Banking System

The financial crisis of 2008 is above all a crisis of the financial system as a whole. This is why it is called the Global Financial Crisis (GFC) and is different than the previous crises (the Great Depression in the 1930s, the Japan crisis in the early 1990s, the Black Monday of 1987, the 1997 Asian financial crisis, etc.). It is a superposition of the 2007 subprime crisis, affecting primarily the mortgage and credit derivative markets, and a liquidity funding crisis following the demise of Lehman Brothers, which affected the credit market and more broadly the shadow banking system. This crisis was not limited to the banking system, but has affected the different actors of the financial sector, in particular insurance companies, asset managers and of course investors. As we have seen in the previous chapters, this led to a strengthening of financial regulation, and not only on the banking sector. The purpose of new regulations in banks, insurance, asset management, pension funds and organization of the financial market is primarily to improve the rules of each sector, but also to reduce the overall systemic risk of the financial sector. In this context, systemic risk is now certainly the biggest concern of financial regulators and the Financial Stability Board (FSB) was created in April 2009 especially to monitor the stability of the global financial system and to manage the systemic risk[1]. It rapidly became clear that the identification of the systemic risk is a hard task and can only be conducted in a gradual manner. This is why some policy responses are not yet finalized, in particular with the emergence of a shadow banking system, whose borders are not well defined.

8.1 Defining systemic risk

The Financial Stability Board defines systemic events in broad terms:

> "*Systemic event is the disruption to the flow of financial services that is (i) caused by an impairment of all or parts of the financial system and (ii) has the potential to have serious negative consequences on the real economy*" (FSB, 2009, page 6).

[1] The FSB is the successor to the Financial Stability Forum (FSF), which was founded in 1999 by the G7 Finance Ministers and Central Bank Governors. With an expanded membership to the G20 countries, the mandate of the FSB has been reinforced with the creation of three Standing Committees:

- the Standing Committee on Assessment of Vulnerabilities (SCAV), which is the FSB's main mechanism for identifying and assessing risks;
- the Standing Committee on Supervisory and Regulatory Cooperation (SRC), which is charged with undertaking further supervisory analysis and framing a regulatory or supervisory policy response to a material vulnerability identified by SCAV;
- the Standing Committee on Standards Implementation (SCSI), which is responsible for monitoring the implementation of agreed FSB policy initiatives and international standards.

Like the Basel Committee on Banking Supervision, the secretariat to the Financial Stability Board is hosted by the Bank for International Settlements and located in Basel.

This definition focuses on three important points. Firstly, systemic events are associated with negative externalities and moral hazard risk, meaning that every financial institution's incentive is to manage its own risk/return trade-off but not necessarily the implications of its risk on the global financial system. Secondly, a systemic event can cause the impairment of the financial system. Lastly, it implies significant spillovers to the real economy and negative effects on economic welfare.

It is clear that the previous definition may appear too large, but also too restrictive. It may be too large, because it is not precise and many events can be classified as systemic events. It is also too restrictive, because it is difficult to identify the event that lies at the origin of the systemic risk. Most of the times, it is caused by the combination of several events. As noted by Zigrand (2014), systemic risk often refers to exogenous shocks, whereas it can also be generated by endogenous shocks:

> "*Systemic risk comprises the risk* ***to*** *the proper functioning of the system as well as the risk created* ***by*** *the system*" (Zigrand, 2014, page 3).

In fact, there are numerous definitions of systemic risk because it is a multifaceted concept.

8.1.1 Systemic risk, systematic risk and idiosyncratic risk

In financial theory, systemic and idiosyncratic risks are generally opposed. Systemic risk refers to the system whereas idiosyncratic risk refers to an entity of the system. For instance, the banking system may collapse, because many banks may be affected by a severe common risk factor and may default at the same time. In economics, we generally make the assumption that idiosyncratic and common risk factors are independent. However, there exist some situations where idiosyncratic risk may affect the system itself. It is the case of large institutions, for example the default of big banks. In this situation, systemic risk refers to the propagation of a single bank distressed risk to the other banks.

Let us consider one of the most famous models in finance, which is the capital asset pricing model (CAPM) developed by William Sharpe in 1964. Under some assumptions, he showed that the expected return of asset i is related to the expected return of the market portfolio in the following way:

$$\mathbb{E}\left[R_{i,t}\right] - r = \beta_i \cdot \left(\mathbb{E}\left[R_{m,t}\right] - r\right) \tag{8.1}$$

where $R_{i,t}$ and $R_{m,t}$ are the asset and market returns, r is the risk-free rate and the coefficient β_i is the beta of the asset i with respect to the market portfolio:

$$\beta_i = \frac{\operatorname{cov}\left(R_{i,t}, R_{m,t}\right)}{\sigma^2\left(R_{m,t}\right)}$$

Contrary to idiosyncratic risks, systematic risk $R_{m,t}$ cannot be diversified, and investors are compensated for taking this risk. This means that the market risk premium is positive ($\mathbb{E}\left[R_{m,t}\right] - r > 0$) whereas the expected return of idiosyncratic risk is equal to zero. By definition, the idiosyncratic risk of asset i is equal to:

$$\epsilon_{i,t} = \left(R_{i,t} - r\right) - \beta_i \cdot \left(\mathbb{E}\left[R_{m,t}\right] - r\right)$$

with $\mathbb{E}\left[\epsilon_{i,t}\right] = 0$. As explained above, this idiosyncratic risk is not rewarded because it can be hedged (or diversified). In this framework, we obtain the one-factor model given by the following equation:

$$R_{i,t} = \alpha_i + \beta_i \cdot R_{m,t} + \varepsilon_i \tag{8.2}$$

where $\alpha_i = (1 - \beta_i)\, r$ and $\varepsilon_{i,t} = \epsilon_{i,t} - \beta_i \cdot (R_{m,t} - \mathbb{E}[R_{m,t}])$ is a white noise process[2]. Because $\varepsilon_{i,t}$ is a new parametrization of the idiosyncratic risk, it is easy to show that this specific factor is independent from the common factor $R_{m,t}$ and the other specific factors $\varepsilon_{j,t}$. If we assume that asset returns are normally distributed, we have $R_{m,t} \sim \mathcal{N}\left(\mathbb{E}[R_{m,t}], \sigma^2_{m,t}\right)$ and:

$$\begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \operatorname{diag}\left(\tilde{\sigma}^2_1, \ldots, \tilde{\sigma}^2_n\right)\right)$$

In the capital asset pricing model, it is obvious that the risk of the system $(R_1, \ldots, R_n)$ is due to the common risk factor also called the systematic risk factor. Indeed, a stress $\mathbb{S}$ can only be transmitted to the system by a shock on R_m:

$$\mathbb{S}(R_m) \implies \mathbb{S}(R_1, \ldots, R_n)$$

This is the traditional form of systemic risk. In the CAPM, idiosyncratic risks are not a source of systemic risk:

$$\mathbb{S}(\varepsilon_i) \not\Longrightarrow \mathbb{S}(R_1, \ldots, R_n)$$

because the specific risk ε_i only affects one component of the system, and not all the components.

In practice, systemic risk can also occur because of an idiosyncratic shock. In this case, we distinguish two different transmission channels:

1. The first channel is the impact of a specific stress on the systematic risk factor:

$$\mathbb{S}(\varepsilon_i) \implies \mathbb{S}(R_m) \implies \mathbb{S}(R_1, \ldots, R_n)$$

   This transmission channel implies that the assumption $\varepsilon_i \perp R_m$ is not valid.

2. The second channel is the impact of a specific stress on the other specific risk factors:

$$\mathbb{S}(\varepsilon_i) \implies \mathbb{S}(\varepsilon_1, \ldots, \varepsilon_n) \implies \mathbb{S}(R_1, \ldots, R_n)$$

   This transmission channel implies that the assumption $\varepsilon_i \perp \varepsilon_j$ is not valid.

Traditional financial models (CAPM, APT) fail to capture these two channels, because they neglect some characteristics of systemic factors: the feedback dynamic of specific risks, the possibility of multiple equilibria and the network density.

The distinction between systematic and idiosyncratic shocks is done by De Bandt and Hartmann (2000). However, as noted by Hansen (2012), systematic risks are aggregate risks that cannot be avoided. A clear example is the equity risk premium. In this case, systematic risks are normal and inherent to financial markets and there is no reason to think that we can prevent them. In the systemic risk literature, common or systematic risks reflect another reality. They are abnormal and are viewed as a consequence of simultaneous adverse shocks that affect a large number of system components (De Bandt and Hartmann, 2000). In this case, the goal of supervisory policy is to prevent them, or at least to mitigate them. In practice, it is however difficult to make the distinction between these two concepts of systematic risk. In what follows, we will use the term *systematic market risk* for normal shocks, even if they are severe and we now reserve the term *systematic risk* for abnormal shocks.

[2] $\varepsilon_{i,t}$ is a new form of the idiosyncratic risk.

8.1.2 Sources of systemic risk

De Bandt and Hartmann (2000) explained that shocks and propagation mechanisms are the two main elements to characterize systemic risk. If we consider our previous analysis, the shock corresponds to the initial stress $\mathbb{S}$ whereas the propagation mechanism indicates the transmission channel $\Longrightarrow$ of this initial shock. It is then useful to classify the several sources of systemic risk depending on the nature of the (systematic) shock or the type of propagation[3].

8.1.2.1 Systematic shocks

Benoit *et al.* (2017) list four main systematic shocks: asset-price bubble risk, correlation risk, leverage risk and tail risk. In what follows, we give their characteristics and some examples. However, even if these risks recover different concepts, they are also highly connected and the boundaries between them are blurred.

Asset-price (or speculative) bubble corresponds to a situation where prices of an asset class rise so sharply that they strongly deviate from their fundamental values[4]. The formation of asset bubbles implies that many financial institutions (banks, insurers, asset managers and asset owners) are exposed to the asset class, because they are momentum investors. They prefer to ride the bubble and take advantage of the situation, because being a contrarian investor is a risky strategy[5]. In this context, the probability of crash occurring increases with investors' belief that "*they can sell the asset at an even higher price in the future*" (Brunnermeier and Oehmke, 2013). Examples of speculative bubbles are Japanese asset bubble in the 1980s, the dot.com bubble between 1997 and 2000 and the United States housing bubble before 2007.

Correlation risk means that financial institutions may invest in the same assets at the same time. They are several reasons to this phenomenon. Herd behavior is an important phenomenon in finance (Grinblatt *et al.*, 1995; Wermers, 1999; Acharya and Yorulmazer, 2008) . It corresponds to the tendency for mimicking the actions of others. According to Devenow and Welch (1996), "*such herding typically arises either from direct payoff externalities (negative externalities in bank runs; positive externalities in the generation of trading liquidity or in information acquisition), principal-agent problems (based on managerial desire to protect or signal reputation), or informational learning (cascades)*". Another reason that explains correlated investments is the regulation, which may have a high impact on the investment behavior of financial institutions. Examples include the liquidity coverage ratio, national regulations of pension funds, Solvency II, etc. Finally, a third reason is the search of diversification or yield. Indeed, we generally notice a strong enthusiasm at the same time for an asset class which is is considered as an investment that helps to diversify portfolios or improve their return.

In periods of expansion, we observe an increase of leverage risk, because financial institutions want to benefit from the good times of the business cycle. As the expansion proceeds, investors becomes then more optimistic and the appetite for risky investments and leverage develops[6]. However, a high leverage is an issue in a stressed period, because of the drop of asset prices. Theoretically, the stressed loss $\mathbb{S}$ cannot be greater than the inverse of the

[3] Concerning idiosyncratic risks, they are several sources of stress, but they can all be summarized by the default of one system's component.

[4] A bubble can be measured by the price-to-earnings (or P/E) ratio, which is equal to the current share price divided by the earnings per share. For instance, stocks of the technology sector had an average price-to-earnings ratio larger than 100 in March 2000.

[5] It is extremely difficult for a financial institution to miss the trend from a short-term business perspective and to see the other financial institutions be successful.

[6] This is known as the Minsky's financial instability hypothesis.

financial institution's leverage ratio $\mathcal{LR}$ in order to maintain its safety:

$$\mathbb{S} \leq \frac{1}{\mathcal{LR}}$$

For instance, in the case where $\mathcal{LR}$ is equal to 5, the financial institution defaults if the loss is greater than 20%. In practice, the stress tolerance depends also on the liquidity constraints. It is then easier to leverage a portfolio in a period of expansion than to deleverage it in a period of crisis, where we generally face liquidity problems. Geanakoplos (2010) explained the downward spiral created by leverage by the amplification mechanism due to the demand of collateral assets[7]. Indeed, decline in asset prices results in asset sales of leveraged investors because of margin call requirements and asset sales results in decline in asset prices. Leverage induces then non-linear and threshold effects that can create systemic risk. The failure of LTCM is a good illustration of leverage risk (Jorion, 2000).

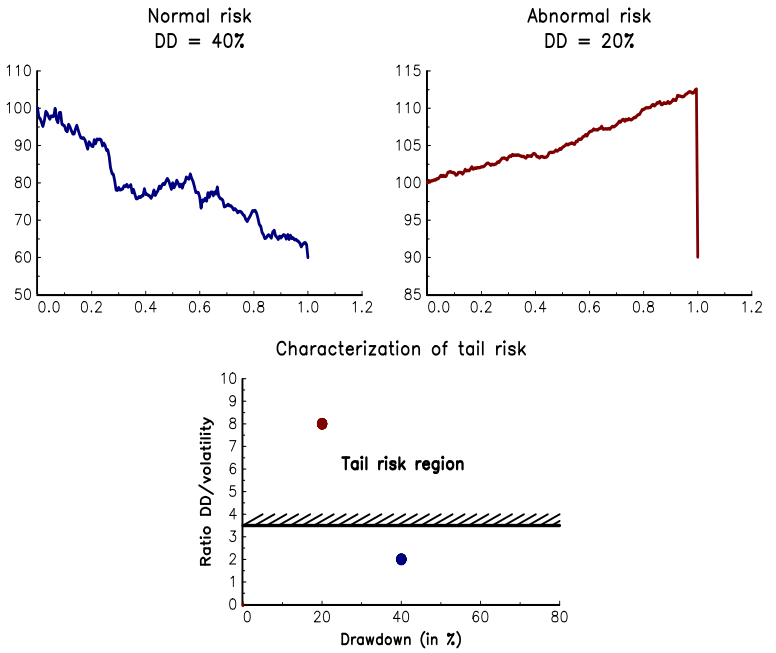


FIGURE 8.1: Illustration of tail risk

The concept of tail risk suggests that the decline in one asset class is abnormal with respect to the normal risk. This means that the probability to observe a tail event is very small. Generally, the normal risk is measured by the volatility. For instance, an order of magnitude is 20% for the long-term volatility of the equity asset class. The probability to observe an annual drop in equities larger than 40% is equal to 2.3%. An equity market crash can therefore not be assimilated to a tail event. By contrast, an asset class whose volatility is equal to 2.5% will experience a tail risk if the prices are 20% lower than before. In this case, the decrease represents eight times the annual volatility. In Figure 8.1, we have reported these two examples of normal and abnormal risks. When the ratio between the drawdown[8] and the volatility is high (e.g. larger than 4), this generally indicates the occurrence of a tail risk. The issue with tail risks is that they are rarely observed and financial institutions

[7] see Section 4.3 on page 293.
[8] It is equal to the maximum loss expressed in percent.

tend to underestimate them. Acharya *et al.* (2010) even suggested that tail risk investments are sought by financial institutions. Such examples are carry or short volatility strategies. For instance, investing in relatively high credit-quality bonds is typically a tail risk strategy. The rationale is to carry the default risk, to capture the spread and to hope that the default will never happen. However, the credit crisis in 2007-2008 showed that very low probability events may occur in financial markets.

The distinction between the four systematic risks is rather artificial and theoretical. In practice, they are highly related. For instance, leverage risk is connected to tail risk. Thus, the carry strategy is generally implemented using leverage. Tail risk is related to bubble risk, which can be partially explained by the correlation risk. In fact, it is extremely difficult to identify a single cause, which defines the zero point of the systemic crisis. Sources of systemic risk are correlated, even between an idiosyncratic event and systematic risks.

8.1.2.2 Propagation mechanisms

As noted by De Bandt and Hartmann (2000), transmission channels of systemic risk are certainly the main element to understand how a systemic crisis happen in an economy. Indeed, propagation mechanisms are more important than the initial (systematic or idiosyncratic) shock, because most of shocks do not produce systemic crisis if they are not spread to the real economy. Among the diversity of propagation mechanisms, academics and regulators have identified three major transmission channels: networks effects, liquidity channel and critical function failure.

Network effects stem from the interconnectedness of financial institutions and can be seen as the system-wide counterpart of an institution's counterparty risk. Network effect is a general term describing the transmission of a systemic shock from one particular entity and market to several entities or markets. In the case of LTCM, systemic risk stemmed from the interconnection between LTCM and the banking system combined with the high leverage strategy pursued by the hedge fund. This created an over sized exposure for the banking system to counterparty credit risk from one single entity. Hence, LTCM's idiosyncratic risk was transferred to the entire financial system and became a source of systemic risk. The early and influential work of Allen and Gale (2000) showed that this source of financial contagion is highly contingent on the network's structure and the size of the shock. Their model also suggests that a fully connected network might be more resilient than an incomplete network, contradicting the idea that systemic risk increases with average interconnectedness. However, interconnectedness of an individual entity is central to the notion of "*being systemically important*". In the banking industry, balance sheet contagion is an important source of systemic risk and is linked to the counterparty credit risk. The complexity of the banking network can create domino effects and feedback loops, because the failure of one bank is a signal on the health of the other banks. This informational contagion is crucial to understand the freeze of the interbank market during the 2008 GFC. Informational contagion is also an important factor of bank runs (Diamond and Dybvig, 1983) . However, network effects are not limited to the banking system. Thus, the subprime crisis showed that they concern the different actors of financial system. It was the case with insurance companies[9] and asset managers. In this last case, money market funds (MMF) were notably impacted, forcing some unprecedented measures as the temporary guarantee of money market funds against losses by the US Treasury:

> "*Following the bankruptcy of Lehman Brothers in 2008, a well-known fund – the Reserve Primary Fund – suffered a run due to its holdings of Lehman's commercial paper. This run quickly spread to other funds, triggering investors' redemp-*

[9] The most famous example is the AIG's bailout by the US government in late 2008.

> *tions of more than USD 300 billion within a few days of Lehman's bankruptcy. Its consequences appeared so dire to financial stability that the U.S. government decided to intervene by providing unlimited deposit insurance to all money market fund deposits. The intervention was successful in stopping the run but it transferred the entire risk of the USD 3 trillion money market fund industry to the government*" (Kacperczyk and Schnabl, 2013).

Liquidity is another important propagation mechanism of systemic risk. For instance, the global financial crisis can be seen as the superposition of the subprime crisis, affecting primarily the mortgage and credit derivative markets and by extension the global banking system, and a liquidity funding crisis following the demise of Lehman Brothers, which affected interbank markets and more broadly the shadow banking system. In this particular case, the liquidity channel caused more stress than the initial systematic event of subprime credit. As shown previously, the concept of liquidity is multi-faceted and recovers various dimensions[10] that are highly connected. In this context, liquidity dry-up events are difficult to predict or anticipate, because they can happen suddenly. This is particularly true for the market liquidity with the recent flash crash/rally events[11]. Brunnermeier and Pedersen (2009) demonstrated that a demand shock can create a flight-to-quality environment in which liquidity and loss spirals can arise simply due to funding requirements on speculators such as margin calls and repo haircuts. In some instances, a liquidity dry-up event resulting from a flight-to-quality environment can result in runs, fire sales, and asset liquidations in general transforming the market into a contagion mechanism. This is particularly true if the market size of the early players affected by the shock is large enough to induce a large increase in price pressure. The likelihood and stringency of these spirals is exacerbated by high leverage ratios.

Besides network effects and liquidity-based amplification mechanisms, the third identified transmission channel for systemic risk relates to the specific function a financial institution may come to play in a specific market, either because of its size relative to the market or because of its ownership of a specific skill which makes its services essential to the functioning of that market. De Bandt and Hartmann (2000) identified payment and settlement systems as the main critical function that can generate systemic risk. The development of central counterparties, which is promoted by the recent financial regulation, is a response to mitigate network and counterparty credit risks, but also to strengthen the critical function of clearing systems. Other examples of critical services concern the entire investment chain from the asset manager to the asset owner, for instance securities lending intermediation chains or custody services.

8.1.3 Supervisory policy responses

The strength of the Global Financial Crisis led to massive government interventions around the world to prop up failing financial institutions, seen as '*too big too fail*'. Public concern about the negative externalities of such interventions called pressingly for structural reforms to prevent whenever possible future similar events. The crisis further brought to light, among other key factors, the failure of regulation to keep up with the complexity of the activities of global financial institutions. In particular, calls for prudential reforms were made around the world to create mechanisms to monitor, prevent and resolve the liquidation

[10] We recall that the main dimensions are market/funding liquidity, idiosyncratic/systematic liquidity, domestic/global liquidity and inside/outside liquidity (see Chapter 6 on page 347).

[11] Examples are the flash crash of 6 May 2010 (US stock markets), the flash rally of 15 October 2014 (US Treasury bonds), the Swiss Franc move of 15 January 2015 (removal of CHF pleg to EUR) and the market dislocation of 24 August 2015 (stock markets and US ETFs).

of financial institutions without the need for government intervention. Consequently, a vast program of financial and institutional reforms was undertaken around the world.

8.1.3.1 A new financial regulatory structure

As explained in the introduction of this chapter, the Financial Stability Board is an international oversight institution created in April 2009 to monitor the stability of the global financial system, and not only the activities of banking and insurance industries[12]. Indeed, the 2008 GFC also highlighted the increasing reliance of large institutions on the shadow banking system. This refers to the broad range of short-term financing products and activities performed by non-bank actors in the financial markets and therefore historically not subject to the same regulatory supervision as banking activities. This explained that the FSB has also the mandate to oversee the systemic risk induced by shadow banking entities[13]. Besides the analysis of the financial system, the main task of the FSB is the identification of systemically important financial institutions (SIFI). FSB (2010) defines them as institutions whose "*distress or disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity*". It distinguishes between three types of SIFIs:

1. G-SIBs correspond to global systemically important banks;
2. G-SIIs designate global systemically important insurers;
3. the third category is defined with respect to the two previous ones; it incorporates other SIFIs than banks and insurers (non-bank non-insurer global systemically important financial institutions or NBNI G-SIFIs).

Every year since 2013, the FSB publishes the list of G-SIFIs. In Tables 8.1 and 8.2, we report the 2018 list of G-SIBs and 2016 list of G-SIIs[14]. At this time, NBNI G-SIFIs are not identified, because the assessment methodology is not achieved[15].

TABLE 8.1: List of global systemically important banks (November 2018)

Agricultural Bank of China	Bank of America	Bank of China
Bank of New York Mellon	Barclays	BNP Paribas
China Construction Bank	Citigroup	Credit Suisse
Deutsche Bank	Goldman Sachs	Crédit Agricole
BPCE	HSBC	ICBC
ING Bank	JPMorgan Chase	Mitsubishi UFJ FG
Mizuho FG	Morgan Stanley	Royal Bank of Canada
Santander	Société Générale	Standard Chartered
State Street	Sumitomo Mitsui FG	UBS
UniCredit	Wells Fargo	

Source: FSB (2018b), 2018 List of Global Systemically Important Banks.

[12] For these two financial sectors, the FSB collaborates with the Basel Committee on Banking Supervision and the International Association of Insurance Supervisors (IAIS).

[13] In this last case, the FSB relies on the works of the International Organization of Securities Commissions (IOSCO).

[14] The list has not been updated since 2016 because FSB and IAIS are in discussion for considering a new framework for the assessment and mitigation of systemic risk in the insurance sector.

[15] See the discussion on page 466.

TABLE 8.2: List of global systemically important insurers (November 2016)

Aegon	Allianz	AIG
Aviva	AXA	MetLife
Ping An Group	Prudential Financial	Prudential plc

Source: FSB (2016), 2016 List of Global Systemically Important Insurers.

Systemic risk is also monitored at the regional level with the European Systemic Risk Board (ESRB) for the European Union and the Financial Stability Oversight Council (FSOC) for the United States. The ESRB was established on 16 December 2010 and is part of the European System of Financial Supervision (ESFS), the purpose of which is to ensure supervision of the EU financial system[16]. As established under the Dodd-Frank reform in July 2010, the FSOC is composed of the Secretary of the Treasury, the Chairman of the Federal Reserve and members of US supervision bodies (CFTC, FDIC, OCC, SEC, etc.).

The global financial crisis had also an impact on the banking supervision structure, in particular in the US and Europe. Since 2010, this is the Federal Reserve Board which is in charge to directly supervise large banks and any firm designated as systemically important by the FSOC (Murphy, 2015). The other banks continue to be supervised by the Federal Deposit Insurance Corporation (FDIC) and the Office of the Comptroller of the Currency (OCC). In Europe, each bank was supervised by its national regulators until the establishment of the Single Supervisory Mechanism (SSM). Starting from 4 November 2014, large European banks are directly supervised by the European Central Bank (ECB), while national supervisors are in a supporting role. This concerns about 120 significant banks and represent 80% of banking assets in the euro area. For each bank supervised by the ECB, a joint supervisory team (JST) is designated. Its main task is to perform the Supervisory Review and Evaluation Process (SREP), propose the supervisory examination programme, implement the approved supervisory decisions and ensure coordination with the on-site inspection teams and liaise with the national supervisors. Public awareness of the systemic risk has also led some countries to reform national supervision structures. For instance in the United Kingdom, the Financial Services Authority (FSA) is replaced in April 2013 by three new supervisory bodies: the Financial Policy Committee (FPC), which is responsible for macro-prudential regulation, the Prudential Regulation Authority (PRA), which is responsible for micro-prudential regulation of financial institutions and the Financial Conduct Authority (FCA), which is responsible for markets regulation.

Remark 88 *The 2008 Global Financial Crisis has also impacted other financial sectors than the banking sector, but not to the same degree. Nevertheless, the powers of existing authorities have been expanded in asset management and markets regulation (ESMA, SEC, CFTC). In 2010, the European Insurance and Occupational Pensions Authority (EIOPA) was established in order to ensure a general supervision at the level of the European Union.*

8.1.3.2 A myriad of new standards

Reforms of the financial regulatory framework were also attempted around the world in order to protect the consumers. Thus, the Dodd-Frank Wall Street Reform and Consumer

[16] Besides the ESRB, the ESFS comprises the European Banking Authority (EBA), the European Insurance and Occupational Pensions Authority (EIOPA), the European Securities and Markets Authority (ESMA) and the Joint Committee of the European Supervisory Authorities.

Protection Act was signed into law in the US in July 2010. It is the largest financial regulation overhaul since 1930. Besides the reform of the US financial regulatory structure, it also concerns investment advisers, hedge funds, insurance, central counterparties, credit rating agencies, derivatives, consumer financial protection, mortgages, etc. One of the most famous propositions is the Volcker rule, which prohibits a bank from engaging in proprietary trading and from owning hedge funds and private equity funds. Another controversial proposition is the Lincoln amendment (or swaps push-out rule), which would prohibit federal assistance to swaps entities.

In Europe, directives on the regulation of markets in financial instruments (MiFID 1 and 2) from 2007 to 2014 as well as regulations on packaged retail and insurance-based investment products (PRIIPS) with the introduction of the key information document (KID) in 2014 came to reinforce the regulation and transparency of financial markets and the protection of investors. European Market Infrastructure Regulation (EMIR) is another important European Union regulation, whose aim is to increase the stability of the OTC derivative markets. It introduces reporting obligation for OTC derivatives (trade repositories), clearing obligation for eligible OTC derivatives, independent valuation of OTC derivatives, common rules for central counterparties and post-trading supervisory.

However, the most important reforms concern the banking sector. Many standards of the Basel III Accord are directly related to systemic risk. Capital requirements have been increased to strengthen the safety of banks. The leverage ratio introduces constraints to limit the leverage of banks. The aim of liquidity ratios (LCR and NSFR) is to reduce the liquidity mismatch of banks. Stress testing programs have been highly developed. Another important measure is the designation of systemically important banks[17], which are subject to a capital surcharge ranging from 1% to 3.5%. All these micro-prudential approaches tend to mitigate idiosyncratic factors. However, common factors are also present in the Basel III Accord. Indeed, the Basel Committee has introduced a countercyclical capital buffer in order to increase the capital of banks during excessive credit growth and to limit the impact of common factors on the systemic risk. Another important change is the careful consideration of counterparty credit risk. This includes of course the 1.25 factor to calculate the default correlation $\rho(\text{PD})$ in the IRB approach[18], but also the CVA capital charge. The promotion of CCPs since 2010 is also another example to limit network effects and reduce the direct interconnectedness between banks. Last but not least, the stressed ES of the Basel III Accord had a strong impact on the capital requirements for market risk.

Remark 89 *Another important reform concerns resolution plans, which describe the banks's strategy for rapid resolution if its financial situation were to deteriorate or if it were to default. In Europe, the Bank Recovery and Resolution Directive (BRRD) applies in all banks and large investment firms since January 2015. In the United States, the orderly liquidation authority (OLA) of the Dodd-Frank Act provides a theoretical framework for bank resolution*[19]*. In Japan, a new resolution regime became effective in March 2014 and ensures that a defaulted bank will be resolved via a bridge bank, where certain assets and liabilities are transferred. More recently, the FSB achieves TLAC standard for global systemically important banks. All these initiatives seek to build a framework to resolve a bank failure without public intervention.*

[17] It concerns both global (G-SIB) and domestic (D-SIB) systemically important banks.

[18] See Footnote 70 on page 184.

[19] Bank resolution plans can be found at the following web page: www.federalreserve.gov/bankinforeg/resolution-plans.htm.

8.2 Systemic risk measurement

They are generally two ways of identify SIFIs. The first one is proposed by supervisors and considers firm-specific information that are linked to the systemic risk, such as the size or the leverage. The second approach has been extensively used by academics and considers market information to measure the impact of the firm-specific default on the entire system.

8.2.1 The supervisory approach

In what follows, we distinguish between the three categories defined by the FSB: banks, insurers and non-bank non-insurer financial institutions.

8.2.1.1 The G-SIB assessment methodology

In order to measure the systemic risk of a bank, BCBS (2014g) considers 12 indicators across five large categories. For each indicator, the score of the bank (expressed in basis points) is equal to the bank's indicator value divided by the corresponding sample total[20]:

$$\text{Indicator Score} = \frac{\text{Bank Indicator}}{\text{Sample Total}} \times 10^4$$

The indicator scores are then averaged to define the category scores and the final score. The scoring system is summarized in Table 8.3. Each category has a weight of 20% and represents one dimension of systemic risk. The size effect (too big too fail) corresponds to the first category, but is also present in all other categories. Network effects are reflected in category 2 (interconnectedness) and category 4 (complexity). The third category measures the degree of critical functions, while the cross-jurisdictional activity tends to identify global banks.

TABLE 8.3: Scoring system of G-SIBs

Category		Indicator		Weight
1	Size	1	Total exposures	1/5
2	Interconnectedness	2	Intra-financial system assets	1/15
		3	Intra-financial system liabilities	1/15
		4	Securities outstanding	1/15
3	Substitutability/financial institution infrastructure	5	Payment activity	1/15
		6	Assets under custody	1/15
		7	Underwritten transactions in debt and equity markets	1/15
4	Complexity	8	Notional amount of OTC derivatives	1/15
		9	Trading and AFS securities	1/15
		10	Level 3 assets	1/15
5	Cross-jurisdictional activity	11	Cross-jurisdictional claims	1/10
		12	Cross-jurisdictional liabilities	1/10

An example of the score computation is given in Table 8.4. It concerns the G-SIB score of BNP Paribas in 2014. Using these figures, the size score is equal to:

$$\text{Score} = \frac{2\,032}{66\,313} = 3.06\%$$

[20]The sample consists of the largest 75 banks defined by the Basel III leverage ratio exposure measure.

TABLE 8.4: An example of calculating the G-SIB score

Category	Indicator	Indicator value[1]	Sample total[1]	Indicator score[2]	Category score[2]
Size	Total exposures	2,032	66,313	306	306
Interconnectedness	Intra-financial system assets	205	7,718	266	370
	Intra-financial system liabilities	435	7,831	556	
	Securities outstanding	314	10,836	290	
Substitutability/financial insitution infrastructure	Payment activity	49,557	1,850,755	268	369
	Assets under custody	4,181	100,012	418	
	Underwritten transactions in debt and equity markets	189	4,487	422	
Complexity	Notional amount of OTC derivatives	39,104	639,988	611	505
	Trading and AFS securities	185	3,311	559	
	Level 3 assets	21	595	346	
Cross-jurisdictional activity	Cross-jurisdictional claims	877	15,801	555	485
	Cross-jurisdictional liabilities	584	14,094	414	
Final score					407

[1] The figures are expressed in billion of EUR.

[2] The figures are expressed in bps.

Source: BCBS (2014), G-SIB Framework: Denominators; BNP Paribas (2014), Disclosure for G-SIBs indicators as of 31 December 2013.

The interconnectedness score is an average of three indicator scores. We obtain:

$$\begin{aligned} \text{Score} &= \frac{1}{3}\left(\frac{205}{7\,718} + \frac{435}{7\,831} + \frac{314}{10\,836}\right) \\ &= \frac{2.656\% + 5.555\% + 2.898\%}{3} \\ &= 3.70\% \end{aligned}$$

The final score is an average of the five category scores:

$$\begin{aligned} \text{Score} &= \frac{1}{5}(3.06\% + 3.70\% + 3.69\% + 5.05\% + 4.85\%) \\ &= 4.07\% \end{aligned}$$

Depending on the score value, the bank is then assigned to a specific bucket, which is used to calculate its specific higher loss absorbency (HLA) requirement. The thresholds used to define the buckets are:

1. 130-229 for Bucket 1 (+1.0% CET1);
2. 230-329 for Bucket 2 (+1.5% CET1);
3. 330-429 for Bucket 3 (+2.0% CET1);
4. 430-529 for Bucket 4 (+2.5% CET1);
5. and 530-629 for Bucket 5 (+3.5% CET1).

For instance, the G-SIB score of BNP Paribas was 407 bps. This implies that BNP Paribas belonged to Bucket 3 and the additional buffer was 2% common equity tier 1 at the end of 2014.

In November 2018, the FSB has published the updated list of G-SIBs and the required level of additional loss absorbency. There are no banks in Bucket 5. The most G-SIB is JPMorgan Chase, which is assigned to Bucket 4 (2.5% of HLA requirement). It is followed by Citigroup, Deutsche Bank and HSBC (Bucket 3 and 2.0% of HLA requirement). Bucket 2 is composed of 8 banks (Bank of America, Bank of China, Barclays, BNP Paribas, Goldman Sachs, ICBC, Mitsubishi UFJ FG and Wells Fargo). The 17 remaining banks given in Table 8.1 on page 460 form Bucket 1. The situation has changed since the first publication in November 2011. Generally, the number of G-SIBs is between 28 and 30 banks. Depending on the year, the list may include BBVA, ING, Nordea and Royal Bank of Scotland. Since 2011, we observe that the number of banks in Buckets 4 and 3 generally decreases, while the number of banks in Bucket 2 increases. For instance, in November 2015, Buckets 4 and 3 were composed of two banks (HSBC and JPMorgan Chase) and four banks (Barclays, BNP Paribas, Citigroup and Deutsche Bank).

Remark 90 *The FSB and the BCBS consider a relative measure of the systemic risk. They first select the universe of the* 75 *largest banks and then defines a G-SIB as a bank which has a total score which is higher than the average score*[21]*. This procedure ensures that there are always systemic banks. Indeed, if the scores are normally distributed, the number of systemic banks is half the number of banks in the universe. This explains that the number of G-SIBs is around* 30*.*

Roncalli and Weisang (2015) reported the average rank correlation (in %) between the five categories for the G-SIBs as of end 2013:

$$\begin{pmatrix} 100.0 & & & & \\ 84.6 & 100.0 & & & \\ 77.7 & 63.3 & 100.0 & & \\ 91.5 & 94.5 & 70.1 & 100.0 & \\ 91.4 & 90.6 & 84.2 & 95.2 & 100.0 \end{pmatrix}$$

We notice the high correlation coefficients[22] between the first (size), second (interconnectedness), fourth (complexity) and fifth categories (cross-jurisdictional activity). This is not surprising that G-SIBs are the largest banks in the world. In fact, the high correlation between the five measures masks the multifaceted reality of systemic risk. This is explained by the homogeneous nature of global systemically important banks in terms of their business model. Indeed, almost all these financial institutions are universal banks mixing both commercial and investment banking.

Besides the HLA requirement, the FSB in consultation with the BCBS has published in November 2015 its proposed minimum standard for '*total loss absorbing capacity*' (TLAC). According to FSB (2015d), "*the TLAC standard has been designed so that failing G-SIBs will have sufficient loss-absorbing and recapitalization capacity available in resolution for authorities to implement an orderly resolution that minimizes impacts on financial stability, maintains the continuity of critical functions, and avoids exposing public funds to loss*". In this context, TLAC requirements would be between 8% to 12%. This means that the total capital would be between 18% and 25% of RWA for G-SIBs[23] as indicated in Figure 8.2.

Remark 91 *Recently, the scoring system has slightly changed with the addition of a trading volume indicator in the third category. The other categories and weights remain unchanged*

[21] It is equal to $10^4/75 \approx 133$.

[22] The highest correlation is between Category 4 and Category 5 (95.2%) whereas the lowest correlation is between Category 2 and Category 3 (63.3%).

[23] Using Table 1.5 on page 21, we deduce that the total capital is equal to 6% of tier 1 plus 2% of tier 2 plus 2.5% of conservation buffer (CB) plus $1\% - 3.5\%$ of systemic buffer (HLA) plus $8\% - 12\%$ of TLAC.

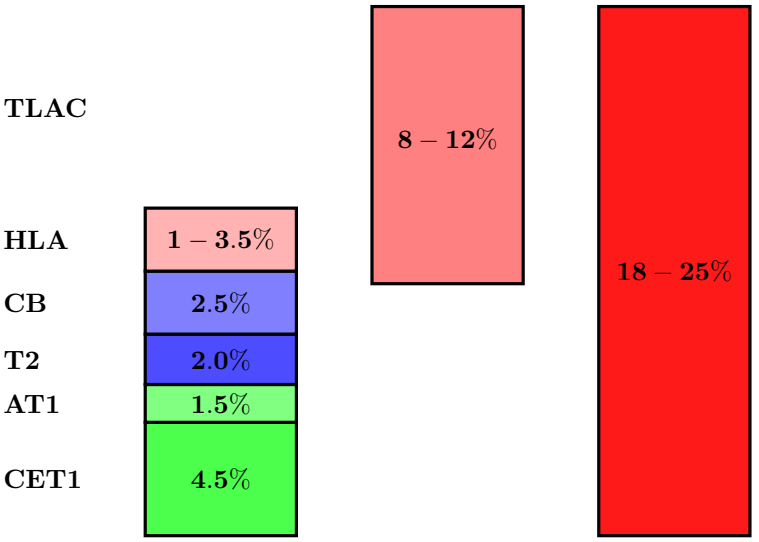


FIGURE 8.2: Impact of the TLAC on capital requirements

except the indicators underwritten transactions in debt and equity markets and trading volume, whose weight is equal to 1/30 *(BCBS, 2018).*

8.2.1.2 Identification of G-SIIs

In the case of insurers, the International Association of Insurance Supervisors (IAIS) has developed an approach similar to the Basel Committee to measure global systemically important insurers (or G-SIIs). The final score is an average of five category scores: size, interconnectedness, substitutability, non-traditional and non-insurance (NTNI) activities and global activity. Contrary to the G-SIB scoring system, the G-SII scoring system does not use an equal weight between the category scores. Thus, a 5% weight is applied to size, substitutability and global activity, whereas interconnectedness and NTNI activities represent respectively 40% and 45% of weighting. In fact, the score highly depends on the banking activities (derivatives trading, short term funding, guarantees, etc.) of the insurance company[24].

8.2.1.3 Extension to NBNI SIFIs

In March 2015, the FSB published a second consultation document, which proposed a methodology for the identification of NBNI SIFIs. The concerned financial sectors were finance companies, market intermediaries, asset managers and their funds. The scoring system was an imitation of the G-SIFI scoring system with the same five categories. As noted by Roncalli and Weisang (2015), this scoring system was not satisfying, because it failed to capture the most important systemic risk of these financial institutions, which is the liquidity risk. Indeed, a large amount of redemptions may create fire sales and affect the liquidity of the underlying market. This liquidity mainly depends on the asset class. For instance, we do not face the same risk when investing in an equity fund and in a bond fund. Finally, the FSB has decided to postpone the assessment framework for NBNI G-SIFIs and to work specifically on financial stability risks from asset management activities.

[24]See IAIS (2013a) on page 20.

8.2.2 The academic approach

Academics propose various methods to measure the systemic risk. Even if they are heterogenous, most of them share a common pattern. They are generally based on publicly market data[25]. Among these different approaches, three prominent measures are particularly popular: the marginal expected shortfall, the delta conditional value-at-risk (Δ CoVaR) and the systemic risk measure (SRISK).

Remark 92 *In what follows, we define the different systemic risk measures and derive their expression in the Gaussian case. Non-Gaussian and non-parametric estimation methods are presented in Chapters 10 and 11.*

8.2.2.1 Marginal expected shortfall

This measure has been proposed by Acharya *et al.* (2017). Let w_i and L_i be the exposure of the system to institution i and the corresponding normalized random loss. We note $w = (w_1, \ldots, w_n)$ the vector of exposures. The loss of the system is equal to:

$$L(w) = \sum_{i=1}^{n} w_i \cdot L_i$$

We recall that the expected shortfall $\mathrm{ES}_\alpha(w)$ with a confidence level α is the expected loss conditional that the loss is greater than the value-at-risk $\mathrm{VaR}_\alpha(w)$:

$$\mathrm{ES}_\alpha(w) = \mathbb{E}[L \mid L \geq \mathrm{VaR}_\alpha(w)]$$

The marginal expected shortfall of institution i is then equal to:

$$\mathrm{MES}_i = \frac{\partial\, \mathrm{ES}_\alpha(w)}{\partial\, w_i} = \mathbb{E}[L_i \mid L \geq \mathrm{VaR}_\alpha(w)] \tag{8.3}$$

In the Gaussian case $(L_1, \ldots, L_n) \sim \mathcal{N}(\mu, \Sigma)$, we have found that[26]:

$$\mathrm{MES}_i = \mu_i + \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{(1-\alpha)\sqrt{w^\top \Sigma w}} \cdot (\Sigma w)_i$$

Another expression of MES is then:

$$\mathrm{MES}_i = \mu_i + \beta_i(w) \cdot (\mathrm{ES}_\alpha(w) - \mathbb{E}(L)) \tag{8.4}$$

where $\beta_i(w)$ is the beta of the institution loss with respect to the total loss:

$$\beta_i(w) = \frac{\mathrm{cov}(L, L_i)}{\sigma^2(L)} = \frac{(\Sigma w)_i}{w^\top \Sigma w}$$

Acharya *et al.* (2017) approximated the MES measure as the expected value of the stock return R_i when the return of the market portfolio R_m is below the 5% quantile:

$$\mathrm{MES}_i = -\mathbb{E}\left[R_i \mid R_m \leq \mathbf{F}^{-1}(5\%)\right]$$

where $\mathbf{F}$ is the cumulative distribution function of the market return R_m. We have:

$$\mathrm{MES}_i = -\frac{1}{\mathrm{card}(\mathbb{T})} \sum_{t \in \mathbb{T}} R_{i,t}$$

[25] The reason is that academics do not have access to regulatory or private data.
[26] See Equation (2.18) on page 107.

where $\mathbb{T}$ represents the set of trading days, which corresponds to the 5% worst days for the market return. Another way of implementing the MES measure is to specify the components of the system and the confidence level α for defining the conditional expectation. For instance, the system can be defined as the set of the largest banks and w_i is the size of Bank i (measured by the market capitalization or the total amount of assets).

Example 76 *We consider a system composed of 3 banks. The total assets managed by these banks are respectively equal to* \$139, \$75 *and* \$81 *bn. We assume that the annual normalized losses are Gaussian. Their means are equal to zero whereas their standard deviations are set equal to* 10%, 12% *and* 15%. *Moreover, the correlations are given by the following matrix:*

$$\mathbb{C} = \begin{pmatrix} 100\% & & \\ 75\% & 100\% & \\ 82\% & 85\% & 100\% \end{pmatrix}$$

By considering a 95% confidence level, the value-at-risk of the system is equal to \$53.86 bn. Using the analytical results given in Section 2.3 on page 104, we deduce that the systemic expected shortfall $\text{ES}_{95\%}$ of the entire system reaches the amount of \$67.55 bn. Finally, we calculate the MES and obtain the values reported in Table 8.5. The MES is expressed in %. This means that if the total assets managed by the first bank increases by \$1 bn, the systemic expected shortfall will increase by \$0.19 bn. In the fourth column of the table, we have indicated the risk contribution $\mathcal{RC}_i$, which is the product of the size w_i and the marginal expected shortfall MES_i. This quantity is also called the systemic expected shortfall of institution i:

$$\text{SES}_i = \mathcal{RC}_i = w_i \cdot \text{MES}_i$$

We have also reported the beta coefficient $\beta_i(w)$ (expressed in bps). Because we have $\mu_i = 0$, we verify that the marginal expected shortfall is equal to the beta times the systemic expected shortfall.

TABLE 8.5: Risk decomposition of the 95% systemic expected shortfall

Bank	w_i (in \$ bn)	MES_i (in %)	SES_i (in \$ bn)	$\beta_i(w)$ (in bps)	$\beta_i(w^\star)$
1	139	19.28	26.80	28.55	0.84
2	75	22.49	16.87	33.29	0.98
3	81	29.48	23.88	43.64	1.29
$\text{ES}_\alpha(w)$			67.55		

The marginal expected shortfall can be used to rank the relative systemic risk of a set of financial institutions. For instance, in the previous example, this is the third bank that is the most risky according to the MES. However, the first bank, which has the lowest MES value, has the highest systemic expected shortfall, because its size is larger than the two other banks. This is why we must not confuse the relative (or marginal) risk and the absolute risk.

The marginal expected shortfall has been criticized because it measures the systematic risk of a financial institution, and not necessarily its systemic risk. In Table 8.5, we give the traditional beta coefficient $\beta_i(w^\star)$, which is calculated with respect to the relative weights $w_i^\star = w_i / \sum_{j=1}^n w_j$. As already shown in Equation (8.4), ranking the financial institutions by their MES is equivalent to rank them by their beta coefficients. In practice, we can nevertheless observe some minor differences because stock returns are not exactly Gaussian.

8.2.2.2 Delta conditional value-at-risk

Adrian and Brunnermeier (2016) define the CoVaR as the value-at-risk of the system conditional on some event $\mathcal{E}_i$ of institution i:

$$\Pr\left\{L(w) \geq \text{CoVaR}_i\left(\mathcal{E}_i\right)\right\} = \alpha$$

Adrian and Brunnermeier determine the risk contribution of institution i as the difference between the CoVaR conditional on the institution being in distressed situation and the CoVaR conditional on the institution being in normal situation:

$$\Delta\,\text{CoVaR}_i = \text{CoVaR}_i\left(\mathcal{D}_i = 1\right) - \text{CoVaR}_i\left(\mathcal{D}_i = 0\right)$$

where $\mathcal{D}_i$ indicates if the bank is in distressed situation or not. Adrian and Brunnermeier use the value-at-risk to characterize the distress situation:

$$\mathcal{D}_i = 1 \Leftrightarrow L_i = \text{VaR}_\alpha\left(L_i\right)$$

whereas the normal situation corresponds to the case when the loss of institution i is equal to its median[27]:

$$\mathcal{D}_i = 0 \Leftrightarrow L_i = m\left(L_i\right)$$

Finally, we obtain:

$$\Delta\,\text{CoVaR}_i = \text{CoVaR}_i\left(L_i = \text{VaR}_\alpha\left(L_i\right)\right) - \text{CoVaR}_i\left(L_i = m\left(L_i\right)\right) \tag{8.5}$$

In the Gaussian case and using the previous notations, we have:

$$\begin{pmatrix} L_i \\ L(w) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_i \\ w^\top \mu \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & (\Sigma w)_i \\ (\Sigma w)_i & w^\top \Sigma w \end{pmatrix}\right)$$

We deduce that[28]:

$$L(w) \mid L_i = \ell_i \sim \mathcal{N}\left(\mu\left(\ell_i\right), \sigma^2\left(\ell_i\right)\right)$$

where:

$$\mu\left(\ell_i\right) = w^\top \mu + \frac{\left(\ell_i - \mu_i\right)}{\sigma_i^2} \cdot (\Sigma w)_i$$

and:

$$\sigma^2\left(\ell_i\right) = w^\top \Sigma w - \frac{(\Sigma w)_i^2}{\sigma_i^2}$$

It follows that:

$$\begin{aligned} \text{CoVaR}_i\left(L_i = \ell\right) &= \mu\left(\ell_i\right) + \Phi^{-1}(\alpha)\,\sigma\left(\ell_i\right) \\ &= w^\top \mu + \frac{\left(\ell_i - \mu_i\right)}{\sigma_i^2}(\Sigma w)_i + \Phi^{-1}(\alpha)\sqrt{w^\top \Sigma w - \frac{(\Sigma w)_i^2}{\sigma_i^2}} \end{aligned}$$

Because $\text{VaR}_\alpha\left(L_i\right) = \mu_i + \Phi^{-1}(\alpha)\,\sigma_i$ and $m\left(L_i\right) = \mathbb{E}\left(L_i\right) = \mu_i$, we obtain:

$$\begin{aligned} \Delta\,\text{CoVaR}_i &= \text{CoVaR}_i\left(L_i = \mu_i + \Phi^{-1}(\alpha)\,\sigma_i\right) - \text{CoVaR}_i\left(L_i = \mu_i\right) \\ &= \Phi^{-1}(\alpha) \cdot \frac{(\Sigma w)_i}{\sigma_i} \\ &= \Phi^{-1}(\alpha) \cdot \sum_{j=1}^{n} w_j \rho_{i,j} \sigma_j \end{aligned}$$

[27] In this case, we have $m\left(L_i\right) = \text{VaR}_{50\%}\left(L_i\right)$.

[28] We use results of the conditional expectation given in Appendix A.2.2.4 on page 1062.

where $\rho_{i,j}$ is the correlation between banks i and j. Another expression of $\Delta\,\text{CoVaR}_i$ is:

$$\Delta\,\text{CoVaR}_i = \Phi^{-1}(\alpha)\cdot\sigma^2(L)\cdot\frac{\beta_i(w)}{\sigma_i} \tag{8.6}$$

The Gaussian case highlights different properties of the CoVaR measure:

- If the losses are independent meaning that $\rho_{i,j}=0$, the Delta CoVaR is the unexpected loss, which is the difference between the nominal value-at-risk and the nominal median (or expected) loss:

$$\begin{aligned}\Delta\,\text{CoVaR}_i &= \Phi^{-1}(\alpha)\cdot w_i\cdot\sigma_i\\ &= w_i\cdot(\text{VaR}_\alpha(L_i)-m(L_i))\\ &= w_i\cdot\text{UL}_\alpha(L_i)\end{aligned}$$

- If the losses are perfectly dependent meaning that $\rho_{i,j}=1$, the Delta CoVaR is the sum of the unexpected losses over all financial institutions:

$$\begin{aligned}\Delta\,\text{CoVaR}_i &= \Phi^{-1}(\alpha)\cdot\sum_{j=1}^{n}w_j\sigma_j\\ &= \sum_{j=1}^{n}w_j\cdot\text{UL}_\alpha(L_j)\end{aligned}$$

  In this case, the Delta CoVaR measure does not depend on the financial institution.

- The sum of all Delta CoVaRs is a weighted average of the unexpected losses:

$$\begin{aligned}\sum_{i=1}^{n}\Delta\,\text{CoVaR}_i &= \Phi^{-1}(\alpha)\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}w_j\rho_{i,j}\sigma_j\\ &= \Phi^{-1}(\alpha)\cdot\sum_{j=1}^{n}w_j\sigma_j\sum_{i=1}^{n}\rho_{i,j}\\ &= n\sum_{j=1}^{n}\bar{\rho}_j\cdot w_j\cdot\text{UL}_\alpha(L_j)\end{aligned}$$

  where $\bar{\rho}_j$ is the average correlation between institution j and the other institutions (including itself). This quantity has no financial interpretation and is not a coherent risk measure satisfying the Euler allocation principle.

Remark 93 *In practice, losses are approximated by stock returns. Empirical results show that MES and CoVaR measures may give different rankings. This can be easily explained in the Gaussian case. Indeed, measuring systemic risk with MES is equivalent to analyze the beta of each financial institution whereas the CoVaR approach consists of ranking them by their beta divided by their volatility. If the beta coefficients are very close, the CoVaR ranking will be highly sensitive to the volatility of the financial institution's stock.*

We consider Example 76 and report in Table 8.6 the calculation of the 95% CoVaR measure. If Bank 1 suffers a loss larger than its 95% value-at-risk ($22.86 bn), it induces a Delta CoVaR of $50.35 bn. This systemic loss includes the initial loss of Bank 1, but also additional losses of the other banks due to their interconnectedness. We notice that CoVaR

and MES produce the same ranking for this example. However, if we define the systemic risk as the additional loss on the other components of the system[29], we find that the stress on Bank 2 induces the largest loss on the other banks[30].

TABLE 8.6: Calculation of the 95% CoVaR measure

Bank	w_i (in \$ bn)	$\mathrm{VaR}_\alpha(L_i)$ (in %)	(in \$ bn)	$\mathrm{CoVaR}_i(\mathcal{E})$ $\mathcal{D}_i = 1$	$\mathcal{D}_i = 0$	$\Delta\,\mathrm{CoVaR}_i$ (in \$ bn)
1	139	16.45	22.86	69.48	19.13	50.35
2	75	19.74	14.80	71.44	22.50	48.94
3	81	24.67	19.98	67.69	16.37	51.32

The dependence function between financial institutions is very important when calculating the CoVaR measure. For instance, we consider again Example 76 with a constant correlation matrix $\mathbb{C}_3(\rho)$. In Figure 8.3, we represent the relationship between $\Delta\,\mathrm{CoVaR}_i$ and the uniform correlation ρ. When losses are independent, we obtain the value-at-risk of each bank. When losses are comonotonic, $\Delta\,\mathrm{CoVaR}_i$ is the sum of the VaRs. Because losses are perfectly correlated, a stress on one bank is entirely transmitted to the other banks.

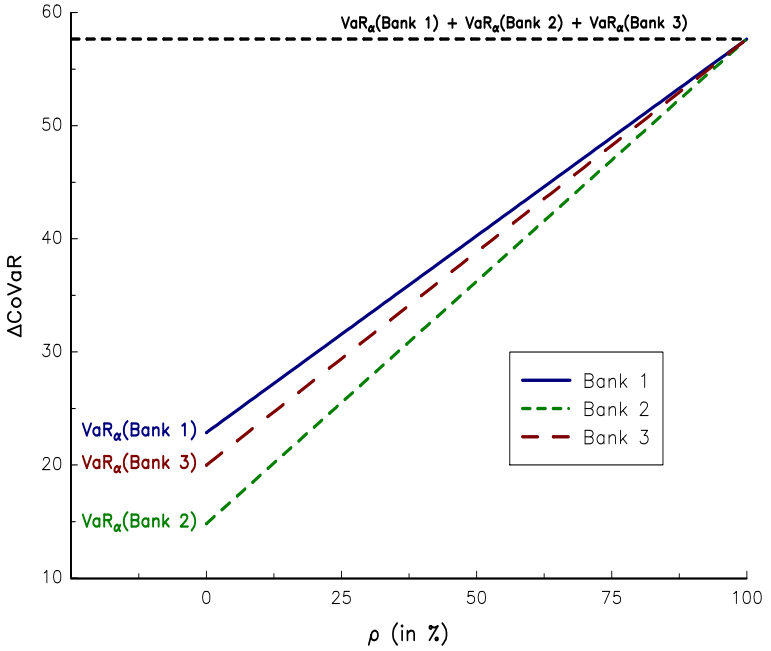


FIGURE 8.3: Impact of the uniform correlation on $\Delta\,\mathrm{CoVaR}_i$

8.2.2.3 Systemic risk measure

Another popular risk measure is the systemic risk measure (SRISK) proposed by Acharya *et al.* (2012), which is a new form of the systemic expected shortfall of Acharya *et al.* (2017) and which was originally developed by Brownlees and Engle (2016) in 2010. Using a stylized

[29] This additional loss is equal to $\mathrm{CoVaR}_i - w_i \cdot \mathrm{VaR}_\alpha(L_i)$.

[30] The additional loss (expressed in \$ bn) is equal to 27.49 for Bank 1, 34.13 for Bank 2 and 31.33 for Bank 3.

balance sheet, the capital shortfall $\text{CS}_{i,t}$ of institution i at time t is the difference between the required capital $\boldsymbol{\mathcal{K}}_{i,t}$ and the market value of equity $V_{i,t}$:

$$\text{CS}_{i,t} = \boldsymbol{\mathcal{K}}_{i,t} - V_{i,t}$$

We assume that $\boldsymbol{\mathcal{K}}_{i,t}$ is equal to $k \cdot A_{i,t}$ where $A_{i,t}$ is the asset value and k is the capital ratio (typically 8% in Basel II). We also have $A_{i,t} = D_{i,t} + V_{i,t}$ where $D_{i,t}$ represents the debt value[31]. We deduce that:

$$\begin{aligned} \text{CS}_{i,t} &= k \cdot (D_{i,t} + V_{i,t}) - V_{i,t} \\ &= k \cdot D_{i,t} - (1-k) \cdot V_{i,t} \end{aligned}$$

We define the capital shortfall of the system as the total amount of capital shortfall $\text{CS}_{i,t}$:

$$\text{CS}_t = \sum_{i=1}^{n} \text{CS}_{i,t}$$

Acharya *et al.* (2012) define the amount of systemic risk as the expected value of the capital shortfall conditionally to a systemic stress $\mathbb{S}$:

$$\begin{aligned} \text{SRISK}_t &= \mathbb{E}\left[\text{CS}_{t+1} | \mathbb{S}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} \text{CS}_{i,t+1} \middle| \mathbb{S}\right] \\ &= \sum_{i=1}^{n} k \cdot \mathbb{E}\left[D_{i,t+1} | \mathbb{S}\right] - (1-k) \cdot \mathbb{E}\left[V_{i,t+1} | \mathbb{S}\right] \end{aligned}$$

They also assume that $\mathbb{E}\left[D_{i,t+1} | \mathbb{S}\right] \approx D_{i,t}$ and:

$$\mathbb{E}\left[V_{i,t+1} | \mathbb{S}\right] = (1 - \text{MES}_{i,t}) \cdot V_{i,t}$$

where $\text{MES}_{i,t}$ is the marginal expected shortfall conditionally to the systemic risk $\mathbb{S}$. By using the leverage ratio $\mathcal{LR}_{i,t}$ defined as the asset value divided by the market value of equity:

$$\mathcal{LR}_{i,t} = \frac{A_{i,t}}{V_{i,t}} = 1 + \frac{D_{i,t}}{V_{i,t}},$$

they finally obtain the following expression of the systemic risk[32]:

$$\text{SRISK}_t = \sum_{i=1}^{n} \left(k \cdot (\mathcal{LR}_{i,t} - 1) - (1-k) \cdot (1 - \text{MES}_{i,t})\right) \cdot V_{i,t}$$

We notice that the systemic risk can be decomposed as the sum of the risk contributions $\text{SRISK}_{i,t}$. We have:

$$\text{SRISK}_{i,t} = \vartheta_{i,t} \cdot V_{i,t} \tag{8.7}$$

with:

$$\vartheta_{i,t} = k \cdot \mathcal{LR}_{i,t} + (1-k) \cdot \text{MES}_{i,t} - 1 \tag{8.8}$$

In these two formulas, k and $\text{MES}_{i,t}$ are expressed in % while $\text{SRISK}_{i,t}$ and $V_{i,t}$ are measured in \$. $\text{SRISK}_{i,t}$ is then a linear function of the market capitalization $V_{i,t}$, which is a proxy of the capital in this model. The scaling factor $\vartheta_{i,t}$ depends on 4 parameters:

[31] Here, we assume that the bank capital is equal to the market value, which is not the case in practice.
[32] We have $D_{i,t} = (\mathcal{LR}_{i,t} - 1) \cdot V_{i,t}$.

1. k is the capital ratio. In the model, we have $\boldsymbol{\mathcal{K}}_{i,t} = k \cdot A_{i,t}$ whereas the capital $\boldsymbol{\mathcal{K}}_{i,t}$ is equal to $k\,\text{RWA}_{i,t}$ in Basel Accords. Under some assumptions, k can be set equal to 8% in the Basel I or Basel II framework. For Basel III and Basel IV, we must use a higher value, especially for SIFIs.

2. $\mathcal{LR}_{i,t}$ is the leverage ratio of institution i. The higher the leverage ratio, the higher the systemic risk.

3. The systemic risk is an increasing function of the marginal expected shortfall. Because we have $\text{MES}_{i,t} \in [0,1]$, we deduce that:

$$(k \cdot \mathcal{LR}_{i,t} - 1) \cdot V_{i,t} \leq \text{SRISK}_{i,t} \leq k \cdot (\mathcal{LR}_{i,t} - 1) \cdot V_{i,t}$$

   A high value of the MES decreases the market value of equity, and then the absorbency capacity of systemic losses.

4. The marginal expected shortfall depends on the stress scenario. In the different publications on the SRISK measure, the stress $\mathbb{S}$ generally corresponds to a 40% drop of the equity market:

$$\text{MES}_{i,t} = -\mathbb{E}\left[R_{i,t+1} \mid R_{m,t+1} \leq -40\%\right]$$

Example 77 *We consider a universe of 4 banks, whose characteristics are given in the table below*[33]*:*

Bank	$V_{i,t}$	$\mathcal{LR}_{i,t}$	μ_i	σ_i	$\rho_{i,m}$
1	57	23	0%	25%	70%
2	65	28	0%	24%	75%
3	91	13	0%	22%	68%
4	120	20	0%	20%	65%

We assume that the expected return μ_m and the volatility σ_m of the equity market are equal to 0% *and* 17%*.*

Using the conditional expectation formula, we have:

$$\mathbb{E}\left[R_{i,t+1} \mid R_{m,t+1} = \mathbb{S}\right] = \mu_i + \rho_{i,m} \cdot \frac{(\mathbb{S} - \mu_m)}{\sigma_m} \cdot \sigma_i$$

We can then calculate the marginal expected shortfall and deduce the scaling factor and the systemic risk contribution thanks to Equations (8.7) and (8.8). Results are given in Table 8.7. In this example, the main contributor is bank 2 because of its high leverage ratio followed by bank 4 because of its high market capitalization. In Table 8.8, we show how the SRISK measure changes with respect to the stress $\mathbb{S}$.

According to Acharya *et al.* (2012), the most important SIFIs in the United States were Bank of America, JPMorgan Chase, Citigroup and Goldman Sachs in 2012. They also noticed that four insurance companies were also in the top 10 (MetLife, Prudential Financial, AIG and Hertford Financial). Engle *et al.* (2015) conducted the same exercise on European institutions with the same methodology. They found that the five most important SIFIs in Europe were Deutsche Bank, Crédit Agricole, Barclays, Royal Bank of Scotland and BNP Paribas. Curiously, HSBC was only ranked at the 15$^{\text{th}}$ place and the first insurance company AXA was 16$^{\text{th}}$. This ranking system is updated in a daily basis by the Volatility Institute at

[33] The market capitalization $V_{i,t}$ is expressed in \$ bn.

TABLE 8.7: Calculation of the SRISK measure ($\mathbb{S} = -40\%$)

Bank	$\text{MES}_{i,t}$ (in %)	$\vartheta_{i,t}$	$\text{SRISK}_{i,t}$ (in \$ bn)	(in %)
1	41.18	1.22	69.47	22.11
2	42.35	1.63	105.93	33.70
3	35.20	0.36	33.11	10.54
4	30.59	0.88	105.77	33.65

TABLE 8.8: Impact of the stress $\mathbb{S}$ on SRISK

Bank	$\mathbb{S} = -20\%$		$\mathbb{S} = -40\%$		$\mathbb{S} = -60\%$	
	(in \$ bn)	(in %)	(in \$ bn)	(in %)	(in \$ bn)	(in %)
1	58.7	22.6	69.5	22.1	80.3	21.7
2	93.3	36.0	105.9	33.7	118.6	32.1
3	18.4	7.1	33.1	10.5	47.8	13.0
4	88.9	34.3	105.8	33.7	122.7	33.2

New York University[34]. In Tables 8.9, 8.10 and 8.11, we report the 10 largest systemic risk contributions by region at the end of November 2015. The ranking within a region seems to be coherent, but the difference in the magnitude of SRISK between American, European and Asian financial institutions is an issue.

Remark 94 *The main drawback of the model is that* $\text{SRISK}_{i,t}$ *is very sensitive to the market capitalization with two effects. The direct effect (*$\text{SRISK}_{i,t} = \vartheta_{i,t} \cdot V_{i,t}$*) implies that the systemic risk is reduced when the equity market is stressed, whereas the indirect effect due to the leverage ratio increases the systemic risk. When we analyze simultaneous the two effects, the first effect is greater. However, we generally observe an increase of the SRISK, because the marginal expected shortfall is much higher in crisis periods.*

8.2.2.4 Network measures

The previous approaches can help to name systemically important financial institutions. However, they cannot help to understand if there is or not a systemic risk. For instance, size is not always the right metric for measuring the systemic risk. If we consider the hedge fund industry, the three most famous bankruptcies are LTCM in 1998 (\$4.6 bn), Amaranth in 2006 (\$6.5 bn) and Madoff in 2008 (\$65 bn). Even if the loss was very large, the Madoff collapse could not produce a systemic risk, because it was a Ponzi scheme, meaning that Madoff assets were not connected to the market. So there were no feedback and spillover effects. In a similar way, the collapse of Amaranth had no impact on the market, except for natural gas futures contracts. Therefore, Amaranth was mainly connected to the market via CCPs. The case of LTCM is completely different, because LCTM was highly leveraged and connected to banks because of interest rate swaps. These three examples show that size is not always a good indicator of systemic risk and interconnectedness is a key parameter for understanding systemic risk. Another issue concerns the sequence of a systemic crisis. In the previous approaches, the origin of a systemic risk is a stress, but there are some events that cannot be explained by such models. This is generally the case of flash crashes, for example the US Stock Market flash crash of 6 May 2010, the US treasury flash crash of

[34]The internet web page is `vlab.stern.nyu.edu`.

TABLE 8.9: Systemic risk contributions in America (2015-11-27)

Rank	institution	$\text{SRISK}_{i,t}$ (in $ bn)	(in %)	$\text{MES}_{i,t}$ (in %)	$\mathcal{LR}_{i,t}$
1	Bank of America	49.7	10.75	2.75	11.42
2	Citigroup	44.0	9.52	3.23	10.83
3	JPMorgan Chase	42.6	9.22	3.09	9.74
4	Prudential Financial	37.6	8.13	3.07	19.64
5	MetLife	33.9	7.33	2.85	15.40
6	Morgan Stanley	28.6	6.20	3.50	12.60
7	Banco do Brasil	24.1	5.22	4.00	29.45
8	Goldman Sachs	20.3	4.38	3.21	10.51
9	Manulife Financial	20.1	4.36	3.43	15.04
10	Power Corp of Canada	16.2	3.50	2.82	26.81

Source: Volatility Institute (2015), vlab.stern.nyu.edu.

TABLE 8.10: Systemic risk contributions in Europe (2015-11-27)

Rank	institution	$\text{SRISK}_{i,t}$ (in $ bn)	(in %)	$\text{MES}_{i,t}$ (in %)	$\mathcal{LR}_{i,t}$
1	BNP Paribas	94.1	8.63	3.42	33.41
2	Crédit Agricole	88.1	8.09	4.22	59.34
3	Barclays	86.3	7.92	4.31	36.60
4	Deutsche Bank	86.1	7.90	4.32	53.61
5	Société Générale	61.3	5.63	3.85	38.74
6	Royal Bank of Scotland	39.5	3.63	3.15	24.23
7	Banco Santander	38.3	3.51	3.79	18.57
8	HSBC	34.5	3.16	2.49	15.96
9	UniCredit	33.1	3.04	3.58	27.21
10	London Stock Exchange	31.3	2.87	2.90	52.67

Source: Volatility Institute (2015), vlab.stern.nyu.edu.

TABLE 8.11: Systemic risk contributions in Asia (2015-11-27)

Rank	institution	$\text{SRISK}_{i,t}$ (in $ bn)	(in %)	$\text{MES}_{i,t}$ (in %)	$\mathcal{LR}_{i,t}$
1	Mitsubishi UFJ FG	121.5	9.45	2.41	24.80
2	China Construction Bank	117.3	9.12	2.61	17.01
3	Bank of China	94.5	7.35	2.53	15.21
4	Mizuho FG	93.7	7.29	2.10	31.84
5	Agricultural Bank of China	92.0	7.16	0.66	19.20
6	Sumitomo Mitsui FG	85.7	6.67	2.71	26.99
7	ICBC	58.4	4.54	0.84	13.80
8	Bank of Communications	45.0	3.50	2.47	16.89
9	Industrial Bank	29.4	2.29	1.38	17.94
10	National Australia Bank	27.4	2.13	3.27	13.48

Source: Volatility Institute (2015), vlab.stern.nyu.edu.

15 October 2014 and the US ETF flash crash of 24 August 2015. These three events have been extensively studied by US regulators (FRB, SEC and CFTC). However, they never found the original cause of these crashes. In fact, such systemic risk events are generally explained by a network risk: small events can propagate in a very dense network in order to produce a large risk because of spillover effects.

Billio *et al.* (2012) and Cont *et al.* (2013) introduce network analysis in order to study the systemic risk of a financial system. In this case, the nodes of the network correspond to financial institutions. Their goal is to measure the connectivity and the centrality of each node in the network. For instance, Figure 8.4 represents the network structure of the Brazilian banking system estimated by Cont *et al.* (2013). The idea is to estimate the contribution of each node to the loss of the system. In this case, the risk contribution depends on the centrality of the node and the density of the network. In particular, they conclude that their results "*emphasize the contribution of heterogeneity in network structure and concentration of counterparty exposures to a given institution in explaining its systemic importance*". The method proposed by Billio *et al.* (2012) is different since it considers Granger-causality networks. However, the two approaches pursue the same goal, which is to propose a measure of connectedness.

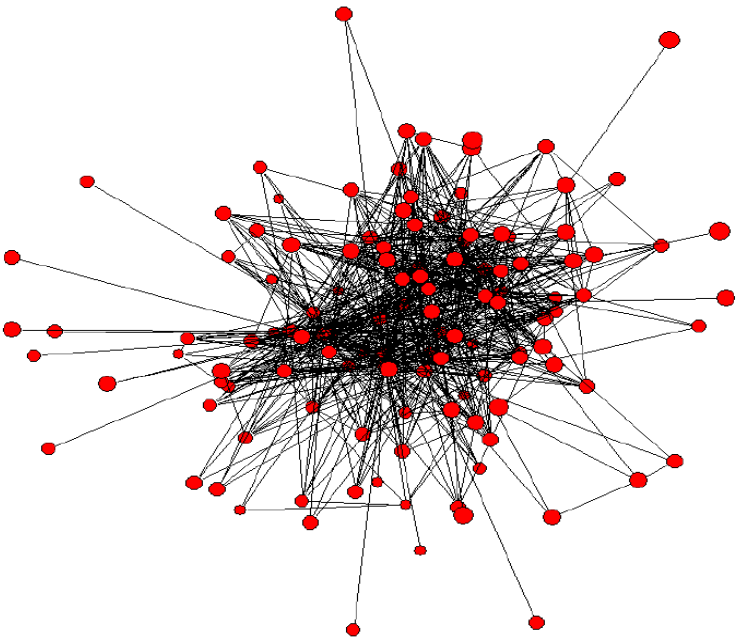


FIGURE 8.4: Network structure of the Brazilian banking system

Source: Cont *et al.* (2013).

Acemoglu *et al.* (2015) have studied the impact of the complexity on the interbank market. They showed that network density can enhance financial stability when (external) shocks are small. But when external shocks are large, a complete network[35] is more risky that a sparse network. This result does not depend on the size of financial institutions.

[35] It corresponds to a network where a financial institution is connected to all the other financial institutions.

These different results illustrate that the systemic risk cannot be reduced to the balance sheet size of financial institutions, but also depends on the connectedness or the density of the network[36]. This is why network risks can be an important component of the systemic risk.

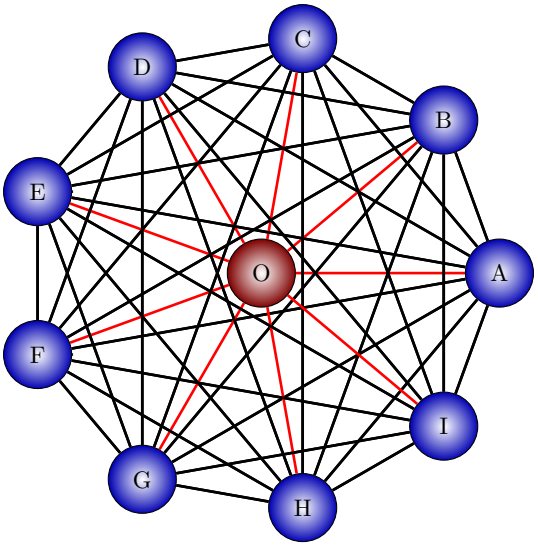


FIGURE 8.5: A completely connected network

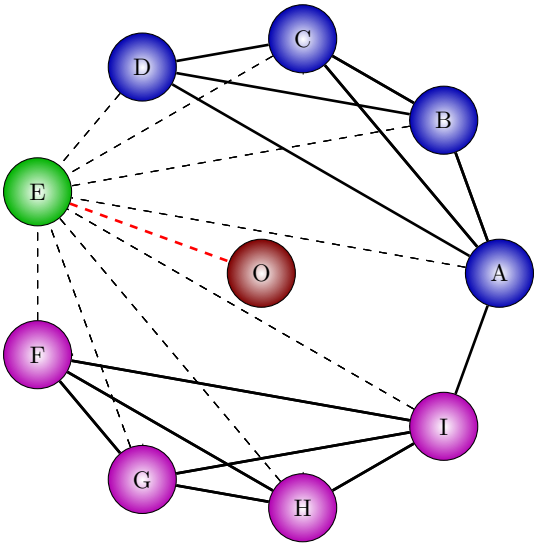


FIGURE 8.6: A sparse network

[36]Figures 8.5 and 8.6 show two examples of networks. The first one is a completely connected network, while the second figure corresponds to a sparse network.

8.3 Shadow banking system

This section on the shadow banking has been included in this chapter together with systemic risk because they are highly connected.

8.3.1 Definition

The shadow banking system (SBS) can be defined as financial entities or activities involved in credit intermediation outside the regular banking system (FSB, 2011; IMF, 2014b). This non-bank credit intermediation complements banking credit, but is not subject to the same regulatory framework. Another important difference is that "*shadow banks are financial intermediaries that conduct maturity, credit, and liquidity transformation without access to central bank liquidity or public sector credit guarantees*" (Pozsar *et al.*, 2013). In this context, shadow banks can raise similar systemic risk issues than regular banks in terms of liquidity, leverage and asset liability mismatch risks.

However, the main characteristic of shadow banking risk is certainly the high interconnectedness within shadow banks and with the banking system. If we describe the shadow banking system in terms of financial entities, it includes finance companies, broker-dealers and asset managers, whose activities are essential for the functioning of the banking system. If we focus on instruments, the shadow banking corresponds to short-term debt securities that are critical for banks' funding. In particular, this concerns money and repo markets. These linkages between the two systems can then create spillover risks, because stress in the shadow banking system may be transmitted to the rest of the financial system (IMF, 2014b). For instance, run risk in shadow banking is certainly the main source of spillover effects and the highest concern of systemic risk. The case of money market funds (MMF) during the 2008 GFC is a good example of the participation of the shadow banking to systemic risk. This dramatic episode also highlights agency and moral hazard problems. Credit risk transfer using asset-backed commercial paper (ABCP) and structured investment vehicles (SIV) is not always transparent for investors of money market funds. This opacity risk increases redemption risk during periods of stress (IMF, 2014b). This led the Federal Reserve to introduce the ABCP money market mutual fund liquidity facility (AMLF) between September 2008 and February 2010 in order to support MMFs.

Concepts of shadow banking and NBNI SIFI are very close. To date, the focus was more on financial entities that can be assimilated to shadow banks or systemic institutions. More recently, we observe a refocusing on instruments and activities. These two approaches go together when measuring the shadow banking system.

8.3.2 Measuring the shadow banking

FSB (2015e) defines two measures of the shadow banking system. The broad measure considers all assets of non-bank financial institutions, while the narrow measure only considers the assets that are part of the credit intermediation chain.

8.3.2.1 The broad (or MUNFI) measure

The broad measure corresponds to the amount of financial assets held by insurance companies, pension funds and other financial intermediaries (OFI). OFIs comprise all financial institutions that are not central banks (CB), banks, insurance companies (IC), pension funds (PF), public financial institutions (PFI) or financial auxiliaries (FA). This broad measure

is also called the MUNFI[37] measure of assets. Table 8.12 shows the amount of assets managed by financial institutions and listed in the 2017 monitoring exercise[38]. Assets rose from \$127.8 tn in 2002 to \$339.9 tn in 2016. This growth is explained by an increase in all financial sectors. In 2016, the MUNFI measure is equal to \$159.3 tn representing 46.9% of the total assets with the following repartition: \$29.1 tn for insurance companies (18.2%), \$31.0 tn for pension funds (19.4%) and \$99.2 tn for other financial intermediaries (62.3%). The MUNFI measure is then larger than banks' assets, which are equal to \$137.8 tn in 2016.

TABLE 8.12: Assets of financial institutions (in \$ tn)

Year	CB	Banks	PFI	IC	PF	OFI	FA	Total	MUNFI	
2002	4.7	52.6	11.2	14.9	11.9	32.4	0.2	127.8	59.2	46.3%
2003	5.5	62.2	12.0	19.3	13.8	39.9	0.3	152.9	73.0	47.7%
2004	6.4	73.1	12.1	22.6	15.3	46.3	0.3	176.0	84.2	47.8%
2005	6.8	76.9	11.9	21.4	16.5	49.9	0.2	183.7	87.8	47.8%
2006	7.7	89.5	11.9	25.3	18.3	60.6	0.3	213.6	104.2	48.8%
2007	10.1	110.7	13.0	29.8	19.8	73.4	0.3	257.1	123.0	47.8%
2008	14.5	123.3	14.2	21.2	19.4	65.8	0.4	258.9	106.5	41.1%
2009	15.1	124.1	14.6	23.7	21.9	70.6	0.6	270.6	116.2	43.0%
2010	16.7	129.8	14.8	25.4	24.4	74.8	0.6	286.5	124.6	43.5%
2011	20.3	139.2	15.0	26.2	25.4	75.7	0.7	302.5	127.3	42.1%
2012	22.4	143.5	15.0	27.9	27.4	83.2	0.8	320.1	138.4	43.3%
2013	23.0	142.0	14.7	28.6	28.9	90.9	0.8	328.9	148.4	45.1%
2014	23.2	138.9	14.7	28.8	29.6	94.9	0.8	330.9	153.3	46.3%
2015	23.6	133.5	15.1	28.2	29.6	94.3	0.7	325.0	152.0	46.8%
2016	26.2	137.8	16.0	29.1	31.0	99.2	0.7	339.9	159.3	46.9%

Source: FSB (2018a) and author's calculations.

Financial assets managed by OFIs are under the scrutiny of the FSB, which has adopted the following classification: money market funds (MMF), hedge funds (HF), other investment funds[39] (IF), real estate investment trusts and real estate funds (REIT), trust companies (TC), finance companies (FC), broker-dealers (BD), structured finance vehicles (SFV), central counterparties (CCP) and captive financial institutions and money lenders[40] (CFI). Table 8.13 gives the repartition of assets by categories. We can now decompose the amount of \$99.2 tn assets reached in 2016 by category of OFIs. 38.1% of these assets concern other investment funds (equity funds, fixed income funds and multi-asset funds). Broker-dealers is an important category of OFIs as they represent 8.8% of assets. It is followed by money market funds (5.1%) and structured finance vehicles (4.5%). We notice that the asset management industry (money market funds, hedge funds, other investments funds and real estate investment companies) represents around 50% of OFIs' assets. The smallest category concerns central counterparties, whose assets are equal to \$404 bn in 2016.

The broad measure suffers from one major shortcoming, because it is an entity-based measure and not an asset-based measure. It then includes both shadow banking assets and other assets. This is particularly true for equity assets, which are not shadow banking

[37] MUNFI is the acronym of '*monitoring universe of non-bank financial intermediation*'.

[38] This exercise covers 29 countries, including for instance BRICS, Japan, the Euro area, the United States and the United Kingdom.

[39] They correspond to equity funds, fixed income funds and multi-asset funds.

[40] They are institutional units that provide financial services, e.g. holding companies used to channel financial flows between group entities or treasury management companies.

TABLE 8.13: Assets of OFIs (in $ tn)

Year	MMF	HF	IF	REIT	TC	FC	BD	SFV	CCP	CFI	Other
2002	3.2	0.0	5.6	0.2	0.0	2.3	3.2	2.3	0.0	2.0	13.6
2003	3.3	0.0	7.5	0.2	0.0	2.7	3.8	2.7	0.0	2.3	17.3
2004	3.4	0.0	8.9	0.4	0.0	2.9	4.6	3.3	0.0	2.7	20.3
2005	3.4	0.0	10.0	0.5	0.0	2.8	5.0	4.2	0.0	3.5	20.6
2006	4.0	0.0	12.6	0.6	0.1	2.9	5.7	5.1	0.0	3.8	26.0
2007	5.1	0.5	16.2	0.6	0.1	3.0	6.5	6.5	0.0	4.4	30.4
2008	6.0	0.6	17.0	0.6	0.3	3.2	9.3	6.4	0.1	4.2	18.1
2009	5.5	0.6	21.9	0.7	0.5	3.6	7.9	9.0	0.5	4.3	16.0
2010	4.8	0.8	25.0	0.8	0.7	3.8	8.7	7.6	0.5	4.6	17.5
2011	4.5	1.5	24.3	1.0	1.0	3.9	9.1	6.7	0.5	4.4	18.9
2012	4.4	2.5	28.8	1.3	1.5	3.5	9.3	6.2	0.5	4.6	20.5
2013	4.5	2.9	33.5	1.4	2.2	3.3	9.1	5.7	0.4	4.7	23.1
2014	4.7	3.5	35.3	1.5	2.7	3.4	9.6	5.1	0.4	4.5	24.1
2015	5.1	3.5	35.1	1.5	2.9	3.4	8.7	4.7	0.4	4.5	24.5
2016	5.0	3.7	37.8	1.6	3.4	3.4	8.7	4.5	0.4	5.1	25.7

Source: FSB (2018a) and author's calculations.

TABLE 8.14: Wholesale funding

		Banks		OFIs	
		2011	2016	2011	2016
Funding (% of balance sheet)	Repo	5.82%	5.52%	6.99%	4.14%
	ST wholesale	4.74%	5.01%	2.91%	4.04%
	LT wholesale	6.94%	7.03%	9.10%	6.45%
Repo (in $ tn)	Assets	3.33	4.16	3.05	4.01
	Liabilities	4.58	4.72	2.92	3.19
	Net position	−0.60	−0.58	0.14	0.83

assets[41]. In this context, the FSB has developed more relevant measures. In Figure 8.7, we have reported the credit assets calculated by the FSB. In 2016, the credit intermediation by banks was equal to $92 tn. At the same time, credit assets by insurance companies and pension funds (ICPF) were equal to $22 tn, whereas the credit intermediation by OFIs peaked at $38 tn. The FSB proposes a sub-decomposition of these credit assets by reporting the lending assets (loans and receivables). The difference between credit and lending assets is essentially composed of investments in debt securities. This decomposition is shown in Figure 8.7. We notice that loans are the main component of banks' credit assets (76%), whereas they represent a small part of the credit intermediation by ICPFs (9%). For OFIs, loans explain 40% of credit assets, but we observe differences between OFIs' sectors. Finance companies and broker-dealers are the main contributors of lending by OFIs.

Remark 95 *Since 2016, the FSB also monitors the funding liquidity, in particular wholesale funding instruments including repurchase agreements (repo). Some figures are given in Table 8.14. Together, short-term wholesale funding and repos represent* 10.5% *and* 8.1% *of the balance sheet of banks and OFIs. We also notice that OFIs are net providers of cash from repos to the financial system, whereas banks are net recipients of cash through repos.*

[41] This concerns for instance equity mutual funds and long/short equity hedge funds.

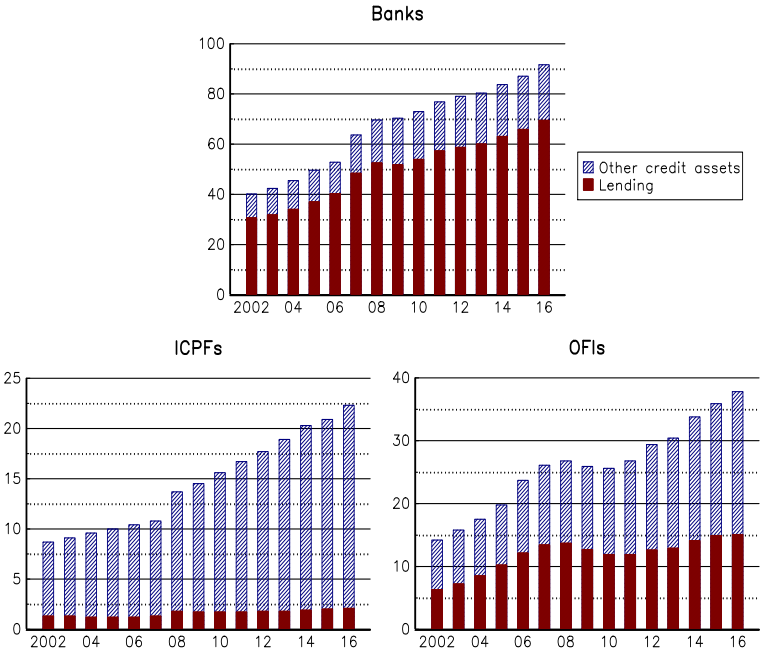


FIGURE 8.7: Credit assets (in $ tn)

Source: FSB (2018a) and author's calculations.

8.3.2.2 The narrow measure

Since 2015, the FSB produces a more relevant measure of the shadow banking system, which is called the narrow measure. The narrow measure is based on the classification of the shadow banking system by economic functions given in Table 8.15. Each of these economic functions involves a shadow banking activity, such as non-bank credit intermediation and/or liquidity/maturity transformation and/or leverage. Moreover, an entity may be classified into two or more economic functions.

The first economic function is related to redemption risks and concerns forced liquidations in an hostile environment. For instance, the lack of liquidity of some fixed income instruments implies a premium for the first investors who unwind their positions on money market and bond funds. In this case, one can observe a run on such funds exactly like a bank run because investors lose confidence in such products and do not want to be the last to move. Run risk can then be transmitted to the entire asset class. This risk mainly concerns collective investment vehicles, whose underlying assets face liquidity issues (fixed income, real estate). The second and fourth economic functions concern lending and credit that are conducted outside of the banking system. The third economic function is related to market intermediation on short-term funding. This includes securities broking services for market making activities and prime brokerage services to hedge funds. Finally, the last economic function corresponds to credit securitization.

The FSB uses these five economic functions in order to calculate the narrow measure defined in Figure 8.8. They consider that pension funds and insurance companies are not participating to the narrow shadow banking system except credit insurance companies. Nevertheless, this last category represents less than $200 bn, implying that the narrow measure principally concerns OFIs. Each OFI is classified or not among the five economic

TABLE 8.15: Classification of the shadow banking system by economic functions

Economic Function	Definition	Typical entity types
EF1	Management of collective investment vehicles with features that are susceptible to runs	Fixed-income funds, mixed funds, credit hedge funds, real estate funds
EF2	Loan provision that is dependent on short-term funding	Finance companies, leasing, factoring and consumer credit companies
EF3	Intermediation of market activities that is dependent on short-term funding or on secured funding of client assets	Broker-dealers, securities finance companies
EF4	Facilitation of credit creation	Credit insurance companies, financial guarantors, monolines
EF5	Securitization-based credit intermediation and funding of financial entities	Securitization vehicles, structured finance vehicles, asset-backed vehicles

Source: FSB (2018a).

functions by the FSB. For instance, equity funds, closed-end funds without leverage and equity REITs are excluded from the shadow banking estimate. Finally, the FSB also removes entities that are subsidiaries of a banking group and consolidated at the group level for prudential purposes[42].

TABLE 8.16: Size of the narrow shadow banking (in $ tn)

Year	2010	2011	2012	2013	2014	2015	2016
Banks	129.8	139.2	143.5	142.0	138.9	133.5	137.8
OFIs	74.8	75.7	83.2	90.9	94.9	94.3	99.2
Shadow banking	28.4	30.2	32.9	35.6	39.0	42.0	45.2

Source: FSB (2018a) and author's calculation.

In Table 8.16, we report the size of the narrow shadow banking and compare it with assets of banks and OFIs. The narrow measure represents 46% of total assets managed by OFIs. These shadow banking assets are located in developed countries, in particular in the United States, Japan, Germany, United Kingdom, Canada and France (see Figure 8.9). We also notice the weight of China (16%) and the importance of three locations: Cayman Islands, Luxembourg and Ireland. These countries are generally used as the domicile of complex mutual funds and alternative investment funds. If we analyze the assets with respect to economic functions, EF1 represents 71.6% of the assets followed by EF5 (9.6%) and EF3 (8.4%), meaning that the shadow banking system involves in the first instance money market and credit funds that are exposed to run risks, securitization vehicles and broker-dealer activities. However, these figures are very different from one country to another country.

[42]This category represents almost 15% of OFIs' assets.

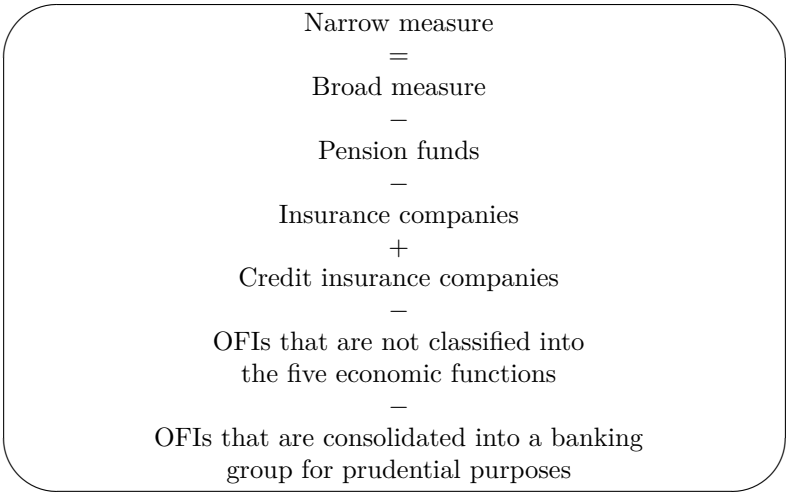


FIGURE 8.8: Calculation of the shadow banking narrow measure

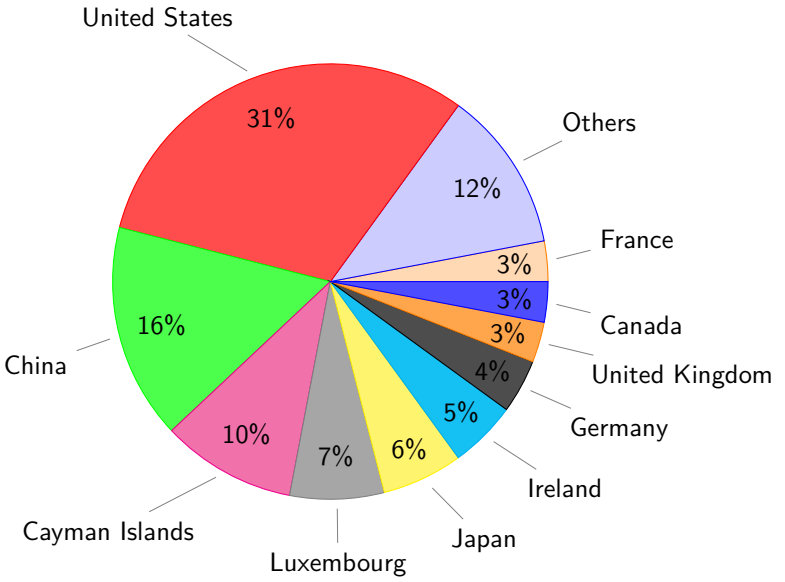


FIGURE 8.9: Breakdown by country of shadow banking assets (2016)

FSB (2018a) provides also network measures between the banking system and OFIs. For that, it estimates the aggregate balance sheet bilateral exposure between the two sectors by considering netting exposures within banking groups that are prudentially consolidated:

- assets of banks to OFIs includes loans to institutions, fixed income securities, reverse repos and investments in money market funds and other investment funds;
- liabilities of banks to OFIs consists of uninsured bank deposits (e.g. certificates of deposit, notes and commercial paper), reverse repos and other short-term debt instruments.

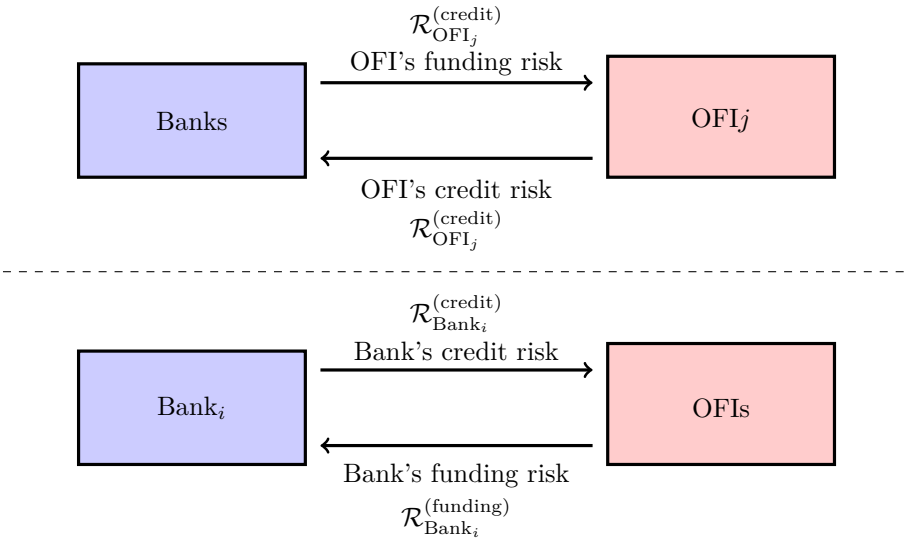


FIGURE 8.10: Interconnectedness between banks and OFIs

Linkages between banks and OFIs are represented in Figure 8.10. These linkages measure the interconnectedness between a set $i \in \mathcal{I}$ of banks and a set $j \in \mathcal{J}$ of OFIs. Let A_{Bank_i} and A_{OFI_j} be the total amount of assets managed by bank i and OFI j. We note $A_{\text{Bank}_i \to \text{OFI}_j}$ and $L_{\text{Bank}_i \to \text{OFI}_j}$ the assets and liabilities of bank i to OFI j, and $A_{\text{OFI}_j \to \text{Bank}_i}$ and $L_{\text{OFI}_j \to \text{Bank}_i}$ the assets and liabilities of OFI j to bank i. By construction, we have $A_{\text{Bank}_i \to \text{OFI}_j} = L_{\text{OFI}_j \to \text{Bank}_i}$ and $L_{\text{Bank}_i \to \text{OFI}_j} = A_{\text{OFI}_j \to \text{Bank}_i}$. In the bottom panel, we have represented the linkage from the bank's perspective. In this case, the credit and funding risks of bank i are equal to:

$$\mathcal{R}^{(\text{credit})}_{\text{Bank}_i} = \frac{A_{\text{Bank}_i \to \text{OFIs}}}{A_{\text{Bank}_i}}$$

and:

$$\mathcal{R}^{(\text{funding})}_{\text{Bank}_i} = \frac{L_{\text{Bank}_i \to \text{OFIs}}}{A_{\text{Bank}_i}}$$

where the aggregate measures are equal to $A_{\text{Bank}_i \to \text{OFIs}} = \sum_{j \in \mathcal{J}} A_{\text{Bank}_i \to \text{OFI}_j}$ and $L_{\text{Bank}_i \to \text{OFIs}} = \sum_{j \in \mathcal{J}} L_{\text{Bank}_i \to \text{OFI}_j}$. In the same way, we can calculate the interconnectedness from the OFI's viewpoint as shown in the top panel. As above, we define the credit and funding risks of OFI j in the following way:

$$\mathcal{R}^{(\text{credit})}_{\text{OFI}_j} = \frac{A_{\text{OFI}_j \to \text{Banks}}}{A_{\text{OFI}_j}}$$

and:

$$\mathcal{R}_{\text{OFI}_j}^{(\text{funding})} = \frac{L_{\text{OFI}_j \to \text{Banks}}}{A_{\text{OFI}_j}}$$

where $A_{\text{OFI}_j \to \text{Banks}} = \sum_{i \in \mathcal{I}} A_{\text{OFI}_j \to \text{Bank}_i}$ and $L_{\text{OFI}_j \to \text{Banks}} = \sum_{i \in \mathcal{I}} L_{\text{OFI}_j \to \text{Bank}_i}$. Using the Shadow Banking Monitoring Dataset 2017, FSB (2018a) finds the following average interconnectedness ratios:

Ratio	$\mathcal{R}_{\text{Banks}}^{(\text{credit})}$	$\mathcal{R}_{\text{Banks}}^{(\text{funding})}$	$\mathcal{R}_{\text{OFIs}}^{(\text{credit})}$	$\mathcal{R}_{\text{OFIs}}^{(\text{funding})}$
2008	6.8%	6.7%	9.5%	9.8%
2016	5.6%	5.4%	6.3%	6.7%

This means that 5.4% of bank's funding depends on the shadow banking system, while the credit risk of banks to OFIs is equal to 5.6% of bank's assets. We also notice that 6.7% of OFIs' assets are provided by banks, while investments of banks into OFIs reaches 6.3%. It is interesting to compare these figures with those during or before the 2008 Global Financial Crisis. We observe that the interconnectedness between banks and OFIs has decreased. For example, the OFI use of funding from banks was 9.8%, while the bank use of funding from OFIs was 6.7%. These figures give an overview of the linkages between banking and OFI sectors. In practice, the interconnectedness is stronger because these ratios are calculated by netting exposures within banking groups. It is obvious that linkages are higher in these entities.

8.3.3 Regulatory developments of shadow banking

The road map for regulating shadow banking, which is presented in FSB (2013), focuses on four key principles:

- measurement and analysis of the shadow banking;
- mitigation of interconnectedness risk between banks and shadow banking entities;
- reduction of the run risk posed by money market funds;
- and improvement of transparency in securitization and more generally in complex shadow banking activities.

8.3.3.1 Data gaps

As seen in the previous section, analyzing the shadow banking system is a big challenge, because it is extremely difficult to measure it. In order to address this issue, FSB and IMF are in charge of the implementation of the G-20 data gaps initiative (DGI). DGI is not specific to shadow banking, but is a more ambitious program for monitoring the systemic risk of the global financial system[43]. However, it is obvious that shadow banking begins to be an important component of DGI. This concerns in particular short-term debt instruments, bonds, securitization and repo markets. Trade repositories, which collect data at the transaction level, complete regulatory reporting to understand shadow banking. They already exist for some OTC instruments in EU and US, but they will certainly be expanded to other markets (e.g. collateralized transactions). Simultaneously, supervisory authorities have strengthened regulatory reporting processes. However, the level of transparency in the shadow banking had still not reach this in banks. Some shadow banking sectors, in particular asset management and pension funds, should then expect new reporting requirements.

[43] For instance, DGI concerns financial soundness indicators (FSI), CDS and securities statistics, banking statistics, public sector debt, real estate prices, etc.

8.3.3.2 Mitigation of interconnectedness risk

BCBS (2013c) has introduced new capital requirements for banks' equity investments in funds that are held in the banking book. They concern investment funds, mutual funds, hedge funds and private equity funds. The framework includes three methods to calculate the capital charge: the fall-back approach (FBA), the mandate-based approach (MBA) and the look-through approach (LTA). In this latter approach, the bank determines the risk-weighted assets of the underlying exposures of the fund. This approach is less conservative than the two others, but requires the full transparency on the portfolio holdings. Under the fall-back approach, the risk weight is equal to 1 250% whatever the risk of the fund. According to the BCBS (2013c), the hierarchy in terms of risk sensitivity between the three approaches was introduced to promote "*due diligence by banks and transparent reporting by the funds in which they invest*". This framework had a significant impact on investment policy of banks and has reduced investments in equity funds and hedge funds.

BCBS (2014c) has developed new standards for measuring and controlling large exposures to single counterparties. This concerns different levels of aggregation from the legal entity to consolidated groups. The large exposures framework is applicable to all international banks, and implies that the exposure of a bank to a consolidated group must be lower than 25% of the bank capital. This figure is reduced to 15% for systemic banks. This framework penalizes then banking groups, which have shadow banking activities (insurance, asset management, brokerage, etc.).

8.3.3.3 Money market funds

Money market funds are under the scrutiny of regulatory authorities since the September 2008 run in the United States. The International Organization of Securities Commissions (2012a) recalled that the systemic risk of these funds is explained by three factors:

1. the illusory perception that MMFs don't have market and credit risks and benefit from capital protection;
2. the first mover advantage, which is pervasive during periods of market distress;
3. and the discrepancy between the published NAV and the asset value.

In order to mitigate these risks, IOSCO (2012a) proposed several recommendations concerning the management of MMFs. In particular, they should be explicitly defined, the investment universe should be restricted to high quality money market and low-duration fixed income instruments, and they should be priced with the fair value approach. Moreover, MMFs that maintain a stable NAV (e.g. 1$ per share) should be converted into floating NAV.

In September 2015, the IOSCO reviewed the implementation progress made by 31 jurisdictions in adopting regulation and policies of MMFs. In particular, this review concerns the five largest jurisdictions (US, Ireland, China, France and Luxembourg), which together account for 90% of global assets under management in MMFs. At that time, only the US reported having final implementation measures in all recommendations, while China and Europe were in the process of finalizing relevant reforms. In July 2014, the US Securities and Exchange Commission adopted final rules for the reform of MMFs. In particular, institutional MMFs will be required to trade at floating NAV. Moreover, all MMFs may impose liquidity fees and redemption gates during periods of stress. In China, the significant growth of the MMF market has forced the Chinese regulator to introduce a number of policy measures in February 2016. This concerns accounting and valuation methods, redefinition of the investment universe, liquidity management and responsibilities of fund managers. In Europe, new rules are applied since July 2017. They distinguish three categories of MMFs:

variable NAV, public debt CNAV and low volatility NAV. These new rules include liquidity management (liquidity fees and redemption gates), prohibit the use of sponsor support and redefine the universe of eligible assets.

8.3.3.4 Complex shadow banking activities

We give here some supervisory initiatives related to some shadow banking activities. In 2011, the European Union has adopted the Alternative Investment Fund Managers Directive (AIFMD), which complements the UCITS directive for asset managers and applies to hedge fund managers, private equity fund managers and real estate fund managers. In particular, it imposes reporting requirements and defines the AIFM passport. In a similar way to MMFs, IOSCO (2012b) published recommendations to improve incentive alignments in securitization, in particular by including issuer risk retention. According to IMF (2014a), Nomura and Daiwa, which are the two largest securities brokerage in Japan, are now subject to Basel III capital requirements and bank-like prudential supervision. New regulation proposals on securities financing transactions (SFT) have been done by the European Commission. They concern reporting, transparency and collateral reuse of SFT activities (repo market, securities lending). These few examples show that the regulation of the shadow banking is in progress and non-bank financial institutions should expect to be better controlled in the future.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part II

Mathematical and Statistical Tools

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 9

Model Risk of Exotic Derivatives

In Chapter 2, we have seen that options and derivative instruments present non-linear risks that are more difficult to assess and measure than for a long-only portfolio of stocks or bonds. Moreover, those financial instruments are traded in OTC markets, meaning that their market value is not known with certainty. These issues imply that the current value is a mark-to-model price and the risk factors depend on the pricing model and the underlying assumptions. The pricing problem is then at the core of the risk management of derivative instruments. However, risk management of such financial products cannot be reduced to a pricing problem. Indeed, the main difficulty lies in managing dynamically the hedging of the option in order to ensure that the replication cost is equal to the option price. In this case, the real challenge is the model risk and concerns three levels: the model risk of pricing the option, the model risk of hedging the option and the discrepancy risk between the pricing model and the hedging model. Therefore, this chapter cannot be just a catalogue of pricing models, but focuses more on pricing errors and hedging uncertainties.

9.1 Basics of option pricing

In this section, we present the basic models that are used for pricing derivatives instruments: the Black-Scholes model, the Vasicek model and the HJM model. While the first one is general and valid for all asset classes, the last two models concern interest rate derivatives.

9.1.1 The Black-Scholes model

9.1.1.1 The general framework

Black and Scholes (1973) assume that the dynamics of the asset price $S(t)$ is given by a geometric Brownian motion:

$$\begin{cases} \mathrm{d}S(t) = \mu S(t)\ \mathrm{d}t + \sigma S(t)\ \mathrm{d}W(t) \\ S(t_0) = S_0 \end{cases} \tag{9.1}$$

where S_0 is the current price, μ is the drift, σ is the volatility of the diffusion and $W(t)$ is a standard Brownian motion. We consider a contingent claim that pays $f(S(T))$ at the maturity T of the derivative contract. For example, if we consider an European option with strike K, we have $f(S(T)) = (S(T) - K)^+$.

Under some conditions, we can show that this contingent claim may be replicated by a hedging portfolio, which is composed of the asset and a risk-free asset, whose instantaneous return is equal to $r(t)$. The price V of the contingent claim is then equal to the cost of the hedging portfolio. In this case, Black and Scholes show that it is the solution of the

following backward equation:

$$\begin{cases} \frac{1}{2}\sigma^2 S^2 \partial_S^2 V(t,S) + (\mu - \lambda(t)\sigma) S\partial_S V(t,S) + \partial_t V(t,S) - r(t) V(t,S) = 0 \\ V(T, S(T)) = f(S(T)) \end{cases}$$

This equation is called the fundamental pricing equation. The function $\lambda(t)$ is interpreted as the risk price of the Wiener process $W(t)$. For an asset whose cost-of-carry is equal to $b(t)$, we have:

$$\lambda(t) = \frac{\mu - b(t)}{\sigma}$$

The previous equation then becomes:

$$\begin{cases} \frac{1}{2}\sigma^2 S^2 \partial_S^2 V(t,S) + b(t) S\partial_S V(t,S) + \partial_t V(t,S) - r(t) V(t,S) = 0 \\ V(T, S(T)) = f(S(T)) \end{cases} \tag{9.2}$$

The current price of the derivatives contract is obtained by solving this partial differential equation (PDE) and to take $V(t_0, S_0)$.

A way to obtain the solution is to apply the Girsanov theorem[1] to the SDE (9.1) with $g(t) = -\lambda(t)$. It follows that:

$$\begin{cases} \mathrm{d}S(t) = b(t) S(t)\, \mathrm{d}t + \sigma S(t)\, \mathrm{d}W^{\mathbb{Q}}(t) \\ S(t_0) = S_0 \end{cases} \tag{9.3}$$

where $W^{\mathbb{Q}}(t)$ is a Brownian motion under the probability $\mathbb{Q}$ defined by:

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^t \lambda(s)\, \mathrm{d}W(s) - \frac{1}{2}\int_0^t \lambda^2(s)\, \mathrm{d}s \right)$$

We may then apply the Feynman-Kac formula[2] with $h(t,x) = r(t)$ and $g(t,x) = 0$ to obtain the martingale solution[3]:

$$V_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(t)\,\mathrm{d}t} f(S(T)) \middle| \mathcal{F}_0 \right] \tag{9.4}$$

Remark 96 *$\mathbb{Q}$ is called the risk-neutral probability (or martingale) measure, because the option price V_0 is the expected discounted value of the payoff*[4].

9.1.1.2 Application to European options

We consider an European call option whose payoff at maturity is equal to:

$$\boldsymbol{\mathcal{C}}(T) = (S(T) - K)^+$$

We assume that the interest rate $r(t)$ and the cost-of-carry parameter $b(t)$ are constant. Then we obtain:

$$\begin{aligned} \boldsymbol{\mathcal{C}}_0 &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r\,\mathrm{d}t} (S(T) - K)^+ \middle| \mathcal{F}_0 \right] \\ &= e^{-rT}\mathbb{E}\left[\left(S_0 e^{\left(b - \frac{1}{2}\sigma^2\right)T + \sigma W^{\mathbb{Q}}(T)} - K \right)^+ \right] \\ &= e^{-rT}\int_{-d_2}^{\infty} \left(S_0 e^{\left(b - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}x} - K \right) \phi(x)\, \mathrm{d}x \\ &= S_0 e^{(b-r)T}\Phi(d_1) - K e^{-rT}\Phi(d_2) \end{aligned} \tag{9.5}$$

[1] See Appendix A.3.5 on page 1072.
[2] See Appendix A.3.4 on page 1070.
[3] We assume that the current date t_0 is equal to 0.
[4] See Exercise 9.4.1 on page 593 for more details.

where:

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T}}\left(\ln\frac{S_0}{K} + bT\right) + \frac{1}{2}\sigma\sqrt{T} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

Let us now consider an European put option with the following payoff:

$$\boldsymbol{\mathcal{P}}(T) = (K - S(T))^+$$

We have:

$$\begin{aligned} \boldsymbol{\mathcal{C}}(T) - \boldsymbol{\mathcal{P}}(T) &= (S(T) - K)^+ - (K - S(T))^+ \\ &= S(T) - K \end{aligned}$$

We deduce that:

$$\begin{aligned} \boldsymbol{\mathcal{C}}_0 - \boldsymbol{\mathcal{P}}_0 &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r\,\mathrm{d}t}(S(T) - K)\middle|\mathcal{F}_0\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}S(T)\middle|\mathcal{F}_0\right] - Ke^{-rT} \\ &= S_0 e^{(b-r)T} - Ke^{-rT} \end{aligned}$$

This equation is known as the put-call parity. It follows that:

$$\begin{aligned} \boldsymbol{\mathcal{P}}_0 &= \boldsymbol{\mathcal{C}}_0 - S_0 e^{(b-r)T} + Ke^{-rT} \\ &= -S_0 e^{(b-r)T}\Phi(-d_1) + Ke^{-rT}\Phi(-d_2) \end{aligned} \tag{9.6}$$

Remark 97 *Equations (9.5) and (9.6) are the famous Black-Scholes formulas. Generally, they are presented with $b = r$, that is for physical assets not paying dividends. The cost-of-carry concept is explained in the next paragraphs.*

We consider a call option on an asset, whose cost-of-carry is equal to 5%. We also assume that the interest rate is equal to 5%. Figure 9.1 represents the option premium with respect to the current value S_0 of the asset. We notice that the price of the call option increases with the current price S_0, the volatility σ and the maturity T. In Figure 9.2, we report the option premium of the put option. In both cases, it may be interesting to decompose the option premium into two components:

- The intrinsic value is the value of exercising the option now:

  $$\mathrm{IV}(t) = f(S_0)$$

  For instance, the intrinsic value of the call option is equal to $(S_0 - K)^+$. If the intrinsic value is positive, the option is said in-the-money (ITM). If the intrinsic value is equal to zero, the option is at-the-money (ATM) or out-of-the-money (OTM).

- The time value is the difference between the option premium and the intrinsic value:

  $$\mathrm{TV}(t) = V(t_0, S_0) - \mathrm{IV}(t)$$

  This quantity is always positive and is related to the risk that the intrinsic value will increase with the time-to-maturity.

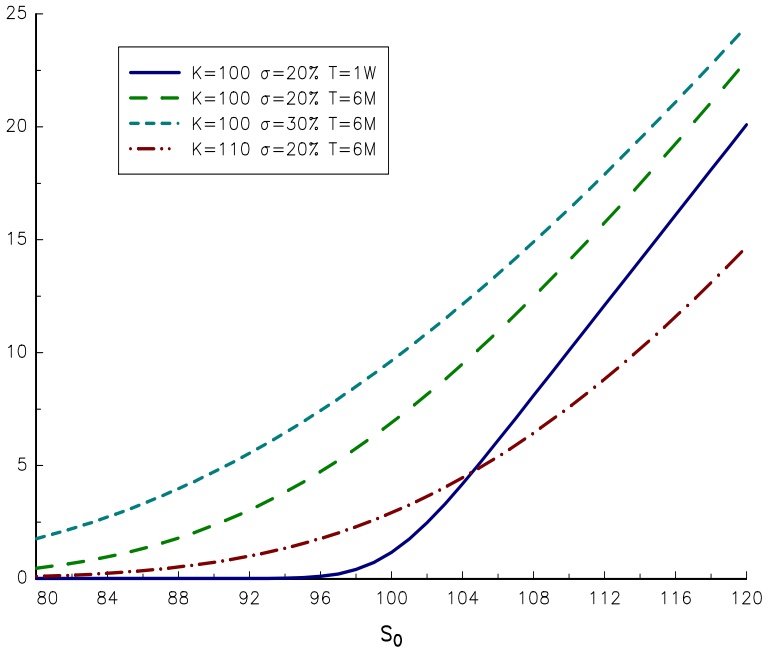


FIGURE 9.1: Price of the call option

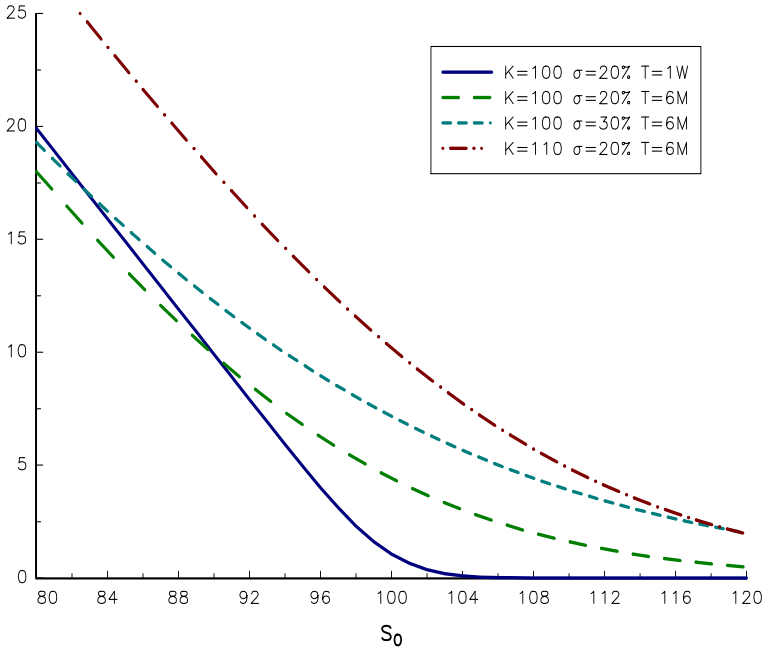


FIGURE 9.2: Price of the put option

9.1.1.3 Principle of dynamic hedging

Self-financing strategy We consider n assets that do not pay dividends or coupons during the period $[0,T]$ and we assume that the price vector $S(t)$ follows a diffusion process. For asset i, we have then:

$$S_i(t) = S_i(0) + \int_0^t \mu_i(u)\,\mathrm{d}u + \int_0^t \sigma_i(u)\,\mathrm{d}W_i(u)$$

We set up a trading portfolio $(\phi_1(t),\ldots,\phi_n(t))$ invested in the assets $(S_1(t),\ldots,S_n(t))$. We note $X(t)$ the value of this portfolio:

$$X(t) = \sum_{i=1}^n \phi_i(t)\,S_i(t)$$

We say that the portfolio is self-financing if the following conditions hold:

$$\begin{cases} \mathrm{d}X(t) - \sum_{i=1}^n \phi_i(t)\,\mathrm{d}S_i(t) = 0 \\ X(0) = 0 \end{cases}$$

The first condition means that all trades are financed by selling or buying assets in the portfolio, whereas the second condition implies that we don't need money to set up the initial portfolio. This implies that:

$$\begin{aligned} X(t) &= X_0 + \sum_{i=1}^n \int_0^t \phi_i(u)\,\mathrm{d}S_i(u) \\ &= \sum_{i=1}^n \phi_i(0)\,S_i(0) + \sum_{i=1}^n \int_0^t \phi_i(u)\,\mathrm{d}S_i(u) \end{aligned}$$

In the Black-Scholes model, we consider a stock that does not pay dividends or coupons during the period $[0,T]$ and we assume that its price process $S(t)$ follows a geometric Brownian motion:

$$\mathrm{d}S(t) = \mu S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t)$$

We also assume the existence of a risk-free asset $B(t)$ that satisfies:

$$\mathrm{d}B(t) = rB(t)\,\mathrm{d}t$$

We set up a trading portfolio $(\phi(t),\psi(t))$ invested in the stock $S(t)$ and the risk-free asset $B(t)$. We note $V(t)$ the value of this portfolio:

$$V(t) = \phi(t)\,S(t) + \psi(t)\,B(t)$$

We now form a strategy $X(t)$ in which we are long the call option $\boldsymbol{\mathcal{C}}(t,S(t))$ and short the trading portfolio $V(t)$:

$$\begin{aligned} X(t) &= \boldsymbol{\mathcal{C}}(t,S(t)) - V(t) \\ &= \boldsymbol{\mathcal{C}}(t,S(t)) - \phi(t)\,S(t) - \psi(t)\,B(t) \end{aligned}$$

Using Itô's lemma, we have:

$$\begin{aligned} \mathrm{d}X(t) &= \partial_S \boldsymbol{\mathcal{C}}(t,S(t))\,\mathrm{d}S(t) + \\ &\quad \left(\partial_t \boldsymbol{\mathcal{C}}(t,S(t)) + \frac{1}{2}\sigma^2 S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t,S(t))\right)\mathrm{d}t - \\ &\quad \phi(t)\,\mathrm{d}S(t) - \psi(t)\,\mathrm{d}B(t) \end{aligned}$$

By assuming that $\phi(t) = \partial_S \boldsymbol{\mathcal{C}}(t, S(t))$, we obtain:

$$\mathrm{d}X(t) = \left(\partial_t \boldsymbol{\mathcal{C}}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t)) - r\psi(t) B(t)\right) \mathrm{d}t$$

$X(t)$ is self-financing if $\mathrm{d}X(t) = 0$ or:

$$\psi(t) = \frac{\partial_t \boldsymbol{\mathcal{C}}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))}{rB(t)}$$

We deduce that:

$$\begin{aligned}
\boldsymbol{\mathcal{C}}(t, S(t)) &= \phi(t) S(t) + \psi(t) B(t) \\
&= \partial_S \boldsymbol{\mathcal{C}}(t, S(t)) S(t) + \\
&\quad \frac{\partial_t \boldsymbol{\mathcal{C}}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))}{rB(t)} B(t)
\end{aligned}$$

This implies that $\boldsymbol{\mathcal{C}}(t, S(t))$ satisfies the following PDE:

$$\frac{1}{2}\sigma^2 S^2 \partial_S^2 \boldsymbol{\mathcal{C}}(t, S) + rS\partial_S \boldsymbol{\mathcal{C}}(t, S) + \partial_t \boldsymbol{\mathcal{C}}(t, S) - r\boldsymbol{\mathcal{C}}(t, S) = 0$$

Since $X(t)$ is self-financing ($X(t) = 0$), we also deduce that the trading portfolio $V(t)$ is the replicating portfolio of the call option:

$$\begin{aligned}
V(t) &= \phi(t) S(t) + \psi(t) B(t) \\
&= \boldsymbol{\mathcal{C}}(t, S(t)) - X(t) \\
&= \boldsymbol{\mathcal{C}}(t, S(t))
\end{aligned}$$

If we define the replicating cost as follows:

$$\begin{aligned}
C(t) &= \int_0^t \phi(u)\,\mathrm{d}S(u) + \int_0^t \psi(u)\,\mathrm{d}B(u) \\
&= \int_0^t (\mu S(u)\phi(u) + rB(u)\psi(u))\,\mathrm{d}u + \int_0^T \sigma S(u)\phi(u)\,\mathrm{d}W(u)
\end{aligned}$$

we have:

$$\begin{aligned}
C(t) &= \int_0^t \mu S(u)\,\partial_S \boldsymbol{\mathcal{C}}(u, S(u))\,\mathrm{d}u + \int_0^T \sigma S(u)\,\partial_S \boldsymbol{\mathcal{C}}(u, S(u))\,\mathrm{d}W(u) \\
&\quad \int_0^t \left(\partial_t \boldsymbol{\mathcal{C}}(u, S(u)) + \frac{1}{2}\sigma^2 S^2(u)\,\partial_S^2 \boldsymbol{\mathcal{C}}(u, S(u))\right) \mathrm{d}u \\
&= \int_0^t \mathrm{d}\boldsymbol{\mathcal{C}}(u, S(u)) \\
&= \boldsymbol{\mathcal{C}}(t, S(t)) - \boldsymbol{\mathcal{C}}(0, S_0)
\end{aligned}$$

We verify that the replicating cost is exactly equal to the P&L of the long exposure on the call option.

Cost-of-carry When the stock does not pay dividends, the cost-of-carry parameter b is equal to the interest rate r. Let us now consider a stock that pays a continuous dividend yield δ, the self-financing portfolio is:

$$X(t) = \mathcal{C}(t, S(t)) - \phi(t) S(t) - \psi(t) B(t)$$

We deduce that the change in the value of this portfolio is:

$$\mathrm{d}X(t) = \mathrm{d}\mathcal{C}(t, S(t)) - \phi(t)\, \mathrm{d}S(t) - \psi(t)\, \mathrm{d}B(t) - \phi(t) \cdot \underbrace{\delta \cdot S(t)\, \mathrm{d}t}_{\text{dividend}}$$

Using the same rationale than previously, we obtain $\phi(t) = \partial_S \mathcal{C}(t, S(t))$ and:

$$\psi(t) = \frac{\partial_t \mathcal{C}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\, \partial_S^2 \mathcal{C}(t, S(t)) - \delta S(t)\, \partial_S \mathcal{C}(t, S(t))}{rB(t)}$$

Finally, we obtain the following PDE:

$$\frac{1}{2}\sigma^2 S^2 \partial_S^2 \mathcal{C}(t, S) + (r - \delta) S \partial_S \mathcal{C}(t, S) + \partial_t \mathcal{C}(t, S) - r\mathcal{C}(t, S) = 0$$

The cost-of-carry parameter b is now equal to $r - \delta$. It is the percentage cost required to carry the asset. Generally, the cost is equal to the interest rate r, but a continuous dividend reduces this cost. In the case of futures or forward contracts, the cost-of-carry is equal to zero. Indeed, the price of such contracts already incorporates the cost-of-carry of the underlying asset. For currency options, the cost-of-carry is the difference between the domestic interest rate r and the foreign interest rate $r^\star$.

TABLE 9.1: Impact of the dividend on the option premium

	Put option				Call option			
S_0 / δ	0.00	0.02	0.05	0.07	0.00	0.02	0.05	0.07
90	1.28	1.44	1.73	1.94	13.50	12.67	11.48	10.72
100	4.42	4.83	5.50	5.97	6.89	6.31	5.50	5.00
110	10.19	10.87	11.91	12.63	2.91	2.59	2.16	1.90

In order to illustrate the impact of the cost-of-carry, we have calculated the option premium in Table 9.1 with the following parameters: $K = 100$, $r = 5\%$ and a six-month maturity. In the case of the put option, the price increases with the dividend yield δ whereas it decreases in the case of the call option. In order to understand these figures, we have to come back to the definition of the replicating portfolio. A call option is replicated using a portfolio that is long on the asset. This implies that the replicating portfolio benefits from the dividends paid by the asset. The self-financing property of the strategy induces that we have to borrow less money. This is why the premium of the call option is lower when the asset pays a dividend. For the put option, this is the contrary. The replicating portfolio is short on the asset. Therefore, it does not receive the dividends, but pays them.

Remark 98 *The value of dividends is an example of model risk. Indeed, future dividends are uncertain, meaning that there is a risk of undervaluation of the option premium. In the case of a call option, the risk is to use expected dividends that are higher than realized values. In the case of put option, the risk is to use low dividends.*

Delta hedging The Black-Scholes model assumes that the replicating portfolio is rebalanced continuously. In practice, it is rebalanced at some fixed dates t_i:

$$0 = t_0 < t_1 < \cdots < t_n = T$$

At the initial date, we have:

$$X(t_0) = \boldsymbol{\mathcal{C}}(t_0, S(t_0)) - V(t_0) = 0$$

where:

$$V(t_0) = \phi(t_0) \cdot S(t_0) + \psi(t_0) \cdot B(t_0)$$

Because we have $\phi(t_0) = \boldsymbol{\Delta}(t_0)$ and $X(t_0) = 0$, we deduce that[5]:

$$\psi(t_0) = \boldsymbol{\mathcal{C}}(t_0, S(t_0)) - \boldsymbol{\Delta}(t_0) S(t_0)$$

At time t_1, the value of the replicating portfolio is then equal to:

$$V(t_1) = \boldsymbol{\Delta}(t_0) S(t_1) + (\boldsymbol{\mathcal{C}}(t_0, S(t_0)) - \boldsymbol{\Delta}(t_0) S(t_0)) \cdot (1 + r(t_0)(t_1 - t_0)) \tag{9.7}$$

It follows that:

$$X(t_1) = \boldsymbol{\mathcal{C}}(t_1, S(t_1)) - V(t_1)$$

Therefore, we are note sure that $X(t_1) = 0$ because it is not possible to hedge the jump $S(t_1) - S(t_0)$. We rebalance the portfolio and we have:

$$V(t_1) = \phi(t_1) \cdot S(t_1) + \psi(t_1) \cdot B(t_1)$$

We deduce that:

$$\phi(t_1) = \boldsymbol{\Delta}(t_1)$$

and:

$$\psi(t_1) = V(t_1) - \boldsymbol{\Delta}(t_1) S(t_1)$$

At time t_2, the value of the replicating portfolio is equal to:

$$V(t_2) = \boldsymbol{\Delta}(t_1) S(t_2) + (V(t_1) - \boldsymbol{\Delta}(t_1) S(t_1)) \cdot (1 + r(t_1)(t_2 - t_1)) \tag{9.8}$$

Equation (9.8) differs from Equation (9.7) because we don't have $V(t_1) = C(t_1, S(t_1))$. More generally, we have:

$$X(t_i) = \boldsymbol{\mathcal{C}}(t_i, S(t_i)) - V(t_i)$$

and:

$$V(t_i) = \underbrace{\boldsymbol{\Delta}(t_{i-1}) S(t_i)}_{V_S(t_i)} + \underbrace{(V(t_{i-1}) - \boldsymbol{\Delta}(t_{i-1}) S(t_{i-1})) \cdot (1 + r(t_{i-1})(t_i - t_{i-1}))}_{V_B(t_i)}$$

where $V_S(t_i)$ is the component due to the delta exposure on the asset and $V_B(t_i)$ is the component due to the cash exposure on the risk-free bond. We notice that:

$$\begin{aligned} V_S(t_i) &= \boldsymbol{\Delta}(t_{i-1}) \cdot S(t_i) \\ &= \boldsymbol{\Delta}(t_{i-1}) \cdot S(t_{i-1}) \cdot (1 + R_S(t_{i-1}; t_i)) \end{aligned}$$

[5] Without any loss of generality, we take the convention that $B(t_i) = 1$.

and:

$$\begin{aligned} V_B(t_i) &= (V(t_{i-1}) - \boldsymbol{\Delta}(t_{i-1}) \cdot S(t_{i-1})) \cdot (1 + r(t_{i-1}) \cdot (t_i - t_{i-1})) \\ &= (V(t_{i-1}) - \boldsymbol{\Delta}(t_{i-1}) \cdot S(t_{i-1})) \cdot (1 + R_B(t_{i-1}; t_i)) \end{aligned}$$

where $R_S(t_{i-1}; t_i)$ and $R_B(t_{i-1}; t_i)$ are the asset and bond returns between t_{i-1} and t_i. At the maturity, we obtain:

$$\begin{aligned} X(T) &= X(t_n) \\ &= (S(T) - K)^+ - V(t_n) \end{aligned}$$

$\Pi(T) = -X(T)$ is the P&L of the delta hedging strategy. To measure its efficiency, we consider the ratio π defined as follows:

$$\pi = \frac{\Pi(T)}{\boldsymbol{\mathcal{C}}(t_0, S(t_0))}$$

Example 78 *We consider the replication of 100 ATM call options. The current price of the asset is* 100 *and the maturity of the option is 20 weeks. We consider the following parameter:* $b = r = 5\%$ *and* $\sigma = 20\%$. *We rebalance the replicating portfolio every week.*

Since the maturity T is equal to $20/52$ and the strike K is equal to 100, the current value $\boldsymbol{\mathcal{C}}(t_0, S(t_0))$ of the call option is equal to \$5.90. The replicating portfolio is rebalanced at times t_i:

$$t_i = \frac{i}{52}$$

In Table 9.2, we have reported a simulated path of the underlying asset. We have $S(t_0) = 100$, $S(t_1) = 95.63$, $S(t_2) = 95.67$, etc. At the maturity date, the price of the underlying asset is equal to 101.83. In the Black-Scholes model, the delta is equal to:

$$\boldsymbol{\Delta}(t) = e^{(b-r)(T-t)} \Phi(d_1)$$

where:

$$d_1 = \frac{1}{\sigma\sqrt{T-t}} \left(\ln \frac{S(t)}{K} + b(T-t) \right) + \sigma\sqrt{T-t}$$

At each rebalancing date t_{i-1}, we compute the delta $\boldsymbol{\Delta}(t_{i-1})$ with respect to the price $S(t_{i-1})$ and the remaining maturity $T - t_{i-1}$. We can then deduce the values of $V_S(t_i)$, $V_B(t_i)$ and $V(t_i)$. We can also calculate the new value $\boldsymbol{\mathcal{C}}(t_i, S(t_i))$ of the call option and compare it with $V(t_i)$ in order to define $X(t_i)$ and $\Pi(t_i) = -X(t_i)$. We obtain $\Pi(T) = -29.76$, implying that:

$$\pi = \frac{-29.76}{100 \times 5.90} = -5.04\%$$

In this case, the delta hedging strategy has produced a negative P&L. If we consider another path of the underlying asset, we can also obtain a positive P&L (see Table 9.3).

We now assume that $S(t)$ is generated by the risk-neutral SDE:

$$\mathrm{d}S(t) = rS(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

We estimate the probability density function of π by simulating 10 000 trajectories of the asset price and calculating the final P&L of the delta hedging strategy. We consider the

TABLE 9.2: An example of delta hedging strategy (negative P&L)

i	t_i	$S(t_i)$	$\boldsymbol{\Delta}(t_{i-1})$	$V_S(t_i)$	$V_B(t_i)$	$V(t_i)$	$\boldsymbol{\mathcal{C}}(t_i, S(t_i))$	$X(t_i)$	$\Pi(t_i)$
0	0.00	100.00	0.00	0.00	590.90	590.90	590.90	0.00	0.00
1	0.02	95.63	58.59	5603.15	−5273.36	329.79	350.22	20.43	−20.43
2	0.04	95.67	43.72	4182.80	−3854.96	327.84	336.15	8.31	−8.31
3	0.06	94.18	43.24	4072.36	−3812.62	259.75	260.57	0.82	−0.82
4	0.08	92.73	37.29	3457.72	−3255.16	202.55	196.22	−6.33	6.33
5	0.10	96.59	31.34	3027.23	−2706.31	320.93	326.47	5.54	−5.54
6	0.12	101.68	44.63	4537.99	−3993.73	544.26	582.71	38.45	−38.45
7	0.13	101.41	63.39	6428.19	−5906.72	521.47	545.64	24.17	−24.17
8	0.15	100.22	62.36	6249.97	−5808.29	441.68	453.62	11.94	−11.94
9	0.17	99.32	57.57	5718.25	−5333.51	384.74	382.58	−2.16	2.16
10	0.19	101.64	53.46	5433.52	−4929.49	504.03	495.99	−8.04	8.04
11	0.21	101.81	63.27	6441.30	−5932.22	509.08	483.87	−25.21	25.21
12	0.23	102.62	64.10	6578.19	−6022.97	555.22	513.53	−41.69	41.69
13	0.25	107.56	67.97	7311.26	−6426.42	884.84	876.68	−8.16	8.16
14	0.27	102.05	86.90	8867.94	−8470.05	397.89	424.07	26.18	−26.18
15	0.29	100.88	66.19	6677.01	−6362.67	314.34	321.76	7.41	−7.41
16	0.31	106.90	59.86	6399.37	−5730.15	669.21	756.02	86.80	−86.80
17	0.33	107.66	90.32	9723.75	−8994.54	729.22	806.47	77.25	−77.25
18	0.35	101.79	94.74	9643.97	−9480.00	163.96	276.24	112.27	−112.27
19	0.37	101.76	69.88	7111.04	−6955.85	155.19	228.08	72.89	−72.89
20	0.38	101.83	75.10	7647.28	−7494.04	153.24	183.00	29.76	−29.76

TABLE 9.3: An example of delta hedging strategy (positive P&L)

i	t_i	$S(t_i)$	$\boldsymbol{\Delta}(t_{i-1})$	$V_S(t_i)$	$V_B(t_i)$	$V(t_i)$	$\boldsymbol{\mathcal{C}}(t_i, S(t_i))$	$X(t_i)$	$\Pi(t_i)$
0	0.00	100.00	0.00	0.00	590.90	590.90	590.90	0.00	0.00
1	0.02	98.50	58.59	5771.31	−5273.36	497.95	489.70	−8.25	8.25
2	0.04	97.00	53.45	5184.51	−4771.31	413.19	396.75	−16.44	16.44
3	0.06	95.47	47.89	4571.99	−4236.14	335.85	311.62	−24.24	24.24
4	0.08	98.17	41.87	4110.19	−3664.81	445.38	419.94	−25.44	25.44
5	0.10	100.48	51.10	5134.88	−4575.85	559.03	528.68	−30.35	30.35
6	0.12	102.92	59.19	6092.33	−5394.04	698.28	664.00	−34.29	34.29
7	0.13	105.50	67.69	7140.94	−6274.05	866.89	829.99	−36.90	36.90
8	0.15	101.81	76.13	7750.53	−7171.44	579.09	550.21	−28.88	28.88
9	0.17	100.65	63.86	6427.97	−5928.66	499.31	457.48	−41.83	41.83
10	0.19	98.86	59.15	5847.59	−5459.40	388.19	337.04	−51.15	51.15
11	0.21	99.26	50.91	5053.11	−4649.03	404.09	335.31	−68.78	68.78
12	0.23	101.78	52.25	5317.65	−4786.50	531.15	458.03	−73.12	73.12
13	0.25	99.28	64.14	6367.78	−6002.74	365.03	288.19	−76.84	76.84
14	0.27	99.19	51.19	5077.96	−4722.07	355.89	257.52	−98.36	98.36
15	0.29	95.53	49.97	4773.36	−4604.77	168.59	92.40	−76.18	76.18
16	0.31	98.02	26.47	2594.85	−2362.61	232.23	148.05	−84.19	84.19
17	0.33	97.03	39.61	3843.35	−3653.84	189.51	83.97	−105.54	105.54
18	0.35	96.64	29.34	2835.17	−2659.65	175.51	44.51	−131.01	131.01
19	0.37	95.01	21.11	2005.37	−1866.05	139.32	3.75	−135.56	135.56
20	0.38	93.67	3.62	338.73	−204.45	134.27	0.00	−134.27	134.27

previous example, but the maturity is now fixed at 130 trading days[6]. Figure 9.3 represents the density function for different fixed rebalancing frequencies[7]. We notice that π is approximately a Gaussian random variable, which is centered around 0. However, the variance depends on the rebalancing frequency. In Figure 9.4, we have reported the relationship between the hedging efficiency $\sigma(\pi)$ and the rebalancing frequency. We confirm that we can perfectly replicate the option with a continuous rebalancing.

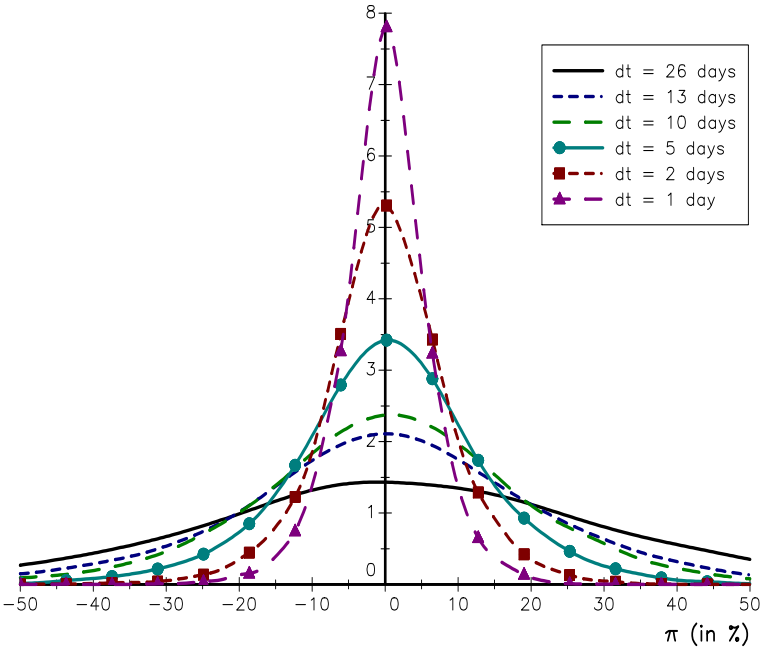


FIGURE 9.3: Probability density function of the hedging ratio π

Let us now understand how the hedging ratio is impacted by the dynamics of the underlying asset. We consider again the previous example and simulate one trajectory (see the first panel in Figure 9.5). We hedge the call option every half an hour. At the maturity, the hedging ratio is equal to 1.8%. The maximum is reached at time $t = 0.466$ and is equal to 3.5%. We now introduce a jump at time $t = 0.25$. This jump induces a large negative P&L for the trader, whatever the sign of the jump (see the second and third panels in Figure 9.5). If we introduce a jump later at time $t = 0.40$, the cost depends on the magnitude and the sign of the jump (Figure 9.6). A positive jump has no impact on the cost of the replicating portfolio, whereas a negative jump has an impact only if the jump is very large. To understand these results, we have to analyze the delta coefficient. At time $t = 0.40$, the option is in-the-money and the delta is close to 1. This implies that a positive jump has low impact on the delta hedging, because the delta is bounded by one. If there is a negative jump, the impact is also limited because the delta is lowly reduced. However, in the case of a high negative jump, the impact may be important because the delta can be dramatically reduced. We also observe the same results when the option is highly out-of-the-money and the delta is close to zero. In this case, a negative jump has no impact, because it decreases

[6] We assume that a year corresponds to 260 trading days. This implies that the maturity of the option is exactly one-half year.

[7] We note $t_i - t_{i-1} = \mathrm{d}t$.

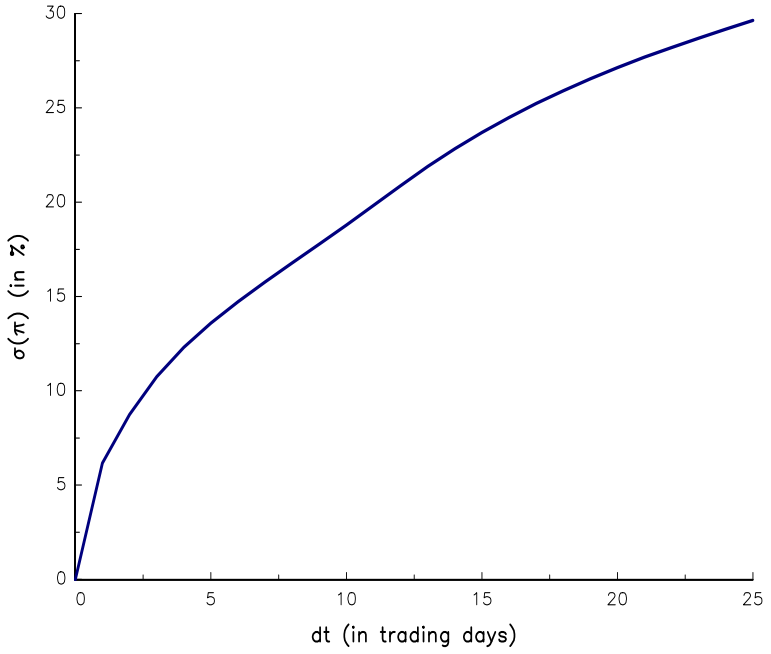


FIGURE 9.4: Relationship between the hedging efficiency $\sigma(\pi)$ and the hedging frequency

the delta but the delta is bounded by zero. Conversely, a positive jump may have an impact if the magnitude is enough sufficiently large to increase the delta.

In the case of liquid markets with low transaction costs, a delta neutral hedging may be efficiently implemented in a high frequency basis (daily or intra-day rebalancing). This is not the case of less liquid markets. Moreover, we observe an asymmetry between call and put options. The delta of call options is positive, implying that the replicating portfolio is long on the asset. For put option, the delta is negative and the replicating portfolio is short on the asset. We know that it is easier to implement a long position than a short position. Sometimes, it is even impossible to be short. For instance, this explains that there exist call options on mutual funds, but not put options on mutual funds. We understand that model risk of derivatives does not only concern the right values of model parameters. In fact, model risk also concerns the hedging management of the option including the feasibility and efficiency of the delta hedging strategy. A famous example is the difference between a put option on S&P 500 index and Eurostoxx 50 index. We know that the returns of the Eurostoxx 50 index present more discontinuous patterns than those of the S&P 500 index. The reason is that European markets react more strongly to American markets than the opposite. This explains that the difference between the closing price and the opening price is more higher in European markets than in American markets. Therefore, a put option on the Eurostoxx 50 index contains an additional premium compared to a put option on the S&P 500 index in order to take into account these stylized facts.

Greek sensitivities We have seen that the delta of the call option is defined by:

$$\boldsymbol{\Delta}(t) = \frac{\partial \boldsymbol{\mathcal{C}}(t, S(t))}{\partial S(t)}$$

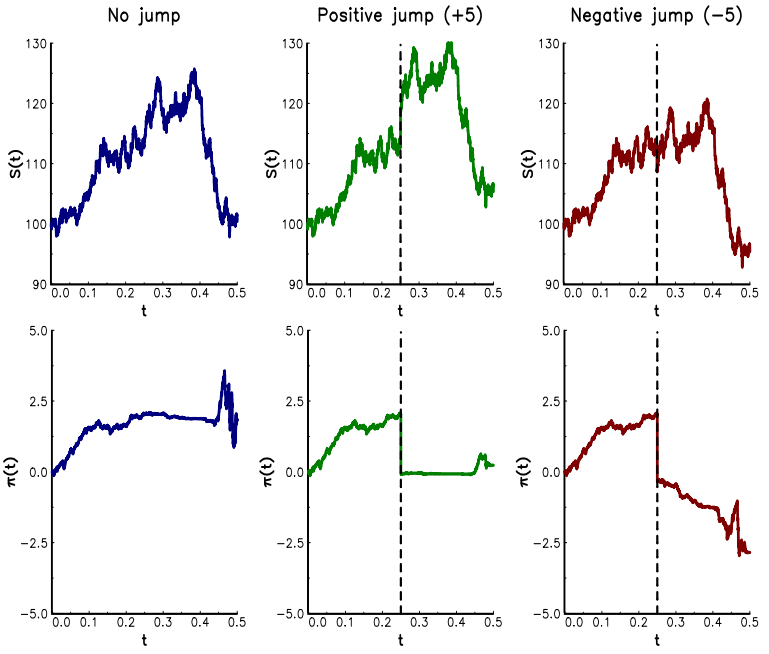


FIGURE 9.5: Impact of a jump on the hedging ratio $\pi(t)$

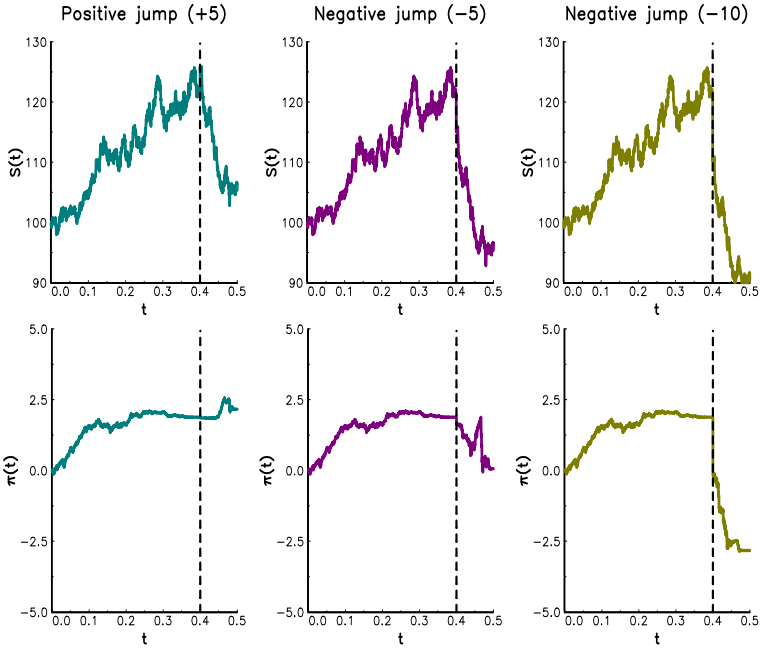


FIGURE 9.6: Impact of a jump on the hedging ratio $\pi(t)$

We have then:

$$\boldsymbol{\mathcal{C}}(t+\mathrm{d}t, S(t+h)) - \boldsymbol{\mathcal{C}}(t, S(t)) \approx \boldsymbol{\Delta}(t) \cdot (S(t+\mathrm{d}t) - S(t))$$

This Taylor expansion can be extended to other orders and other parameters. For instance, the delta-gamma-theta approximation is:

$$\begin{aligned}\boldsymbol{\mathcal{C}}(t+\mathrm{d}t, S(t+h)) - \boldsymbol{\mathcal{C}}(t, S(t)) \quad \approx \quad & \boldsymbol{\Delta}(t) \cdot (S(t+\mathrm{d}t) - S(t)) + \\ & \frac{1}{2}\boldsymbol{\Gamma}(t) \cdot (S(t+\mathrm{d}t) - S(t))^2 + \\ & \boldsymbol{\Theta}(t) \cdot ((t+\mathrm{d}t) - t)\end{aligned}$$

where the gamma is the second-order derivative of the call option price with respect to the underlying asset price:

$$\boldsymbol{\Gamma}(t) = \frac{\partial^2 \boldsymbol{\mathcal{C}}(t, S(t))}{\partial S(t)^2} = \frac{\partial \boldsymbol{\Delta}(t)}{\partial S(t)}$$

and the theta is the derivative of the call option price with respect to the time:

$$\boldsymbol{\Theta}(t) = \frac{\partial \boldsymbol{\mathcal{C}}(t, S(t))}{\partial t} = -\frac{\partial \boldsymbol{\mathcal{C}}(t, S(t))}{\partial T}$$

A positive theta coefficient implies that the option value increases if nothing changes, in particular the price of the underlying asset. By construction, the theta is related to the time value of the option. This is why the theta is generally low for options with a short maturity. In fact, understanding theta effects is complicated, because the theta coefficient is not monotonic in any of the parameters (underlying price, volatility and maturity). We recall that the option price satisfies the PDE:

$$\frac{1}{2}\sigma^2 S^2 \boldsymbol{\Gamma} + bS\boldsymbol{\Delta} + \boldsymbol{\Theta} - r\boldsymbol{\mathcal{C}} = 0$$

We deduce that the theta of the option can be calculated as follows:

$$\boldsymbol{\Theta} = r\boldsymbol{\mathcal{C}} - \frac{1}{2}\sigma^2 S^2 \boldsymbol{\Gamma} - bS\boldsymbol{\Delta}$$

This equation shows that the different coefficients are highly related.

Example 79 *We consider a call option, whose strike K is equal to 100. The risk-free rate and the cost-of-carry parameter are equal to 5%. For the volatility coefficient, we consider two cases: (a) $\sigma = 20\%$ and (b) $\sigma = 50\%$.*

In Figure 9.7, we have reported the option delta for different values of the asset price S_0 and different values of the maturity T. We have $\boldsymbol{\Delta}(t) \in [0, 1]$. The delta is close to zero when the asset price is far below the option strike, whereas it is close to one when the option is highly in-the-money. We also notice that the coefficient $\boldsymbol{\Delta}$ is an increasing function of the price of the underlying asset. The relationship between the option delta and the maturity parameter is not monotonous and depends whether the option is in-the-money or out-of-the-money. In a similar way, the impact of the volatility is not obvious, and may be different if the option maturity is long or short.

Figure 9.8 represents the option gamma[8]. It is close to zero when the current price of the underlying asset is far from the option strike. In this case, the option trader does not

[8] See Exercise 2.4.7 on page 121 for the analytical expression of the different sensitivity coefficients of the call option.

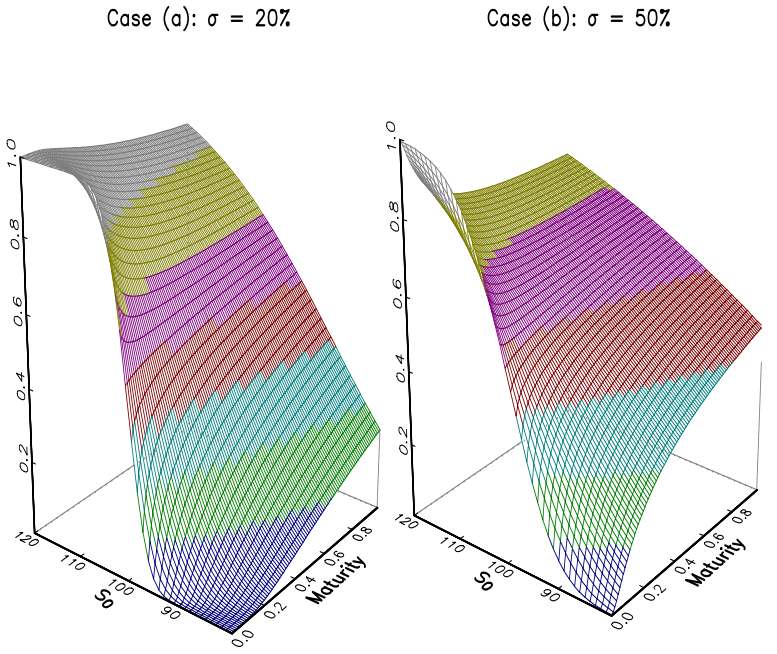


FIGURE 9.7: Delta coefficient of the call option

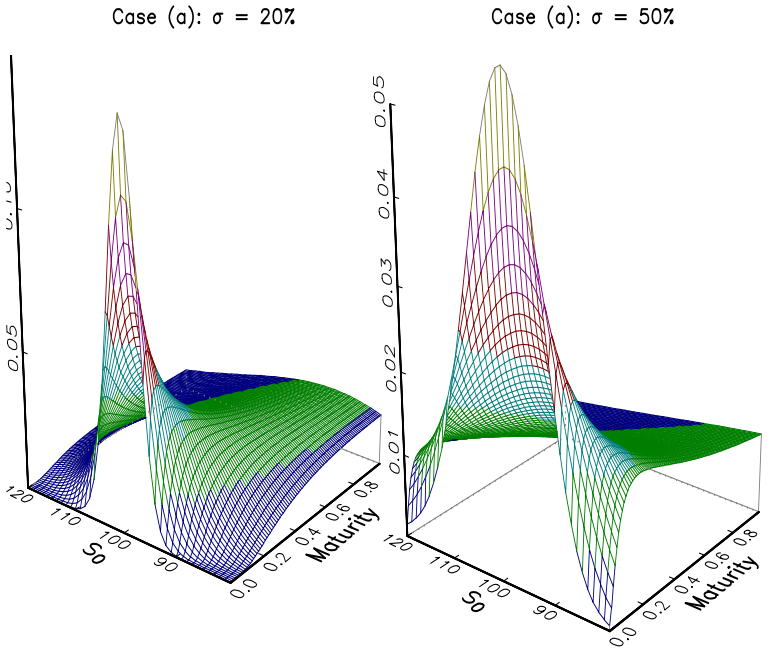


FIGURE 9.8: Gamma coefficient of the call option

need to revise its delta exposure frequently. The gamma coefficient is maximum in the at-the-money region or when the delta is close to 50%. In this situation, the delta can highly vary and the trader must rebalance the replicating portfolio more frequently in order to reduce the residual risk.

Let us assume a delta neutral hedging portfolio. The trader can face four configurations of residual risk given by the following table:

		$\boldsymbol{\Gamma}$	
		$-$	$+$
$\boldsymbol{\Theta}$	$-$		✓
	$+$	✓	

The configuration $(\boldsymbol{\Gamma} < 0, \boldsymbol{\Theta} < 0)$ is not realistic, because the trader will not accept to build a portfolio, whose P&L is almost surely negative. The configuration $(\boldsymbol{\Gamma} > 0, \boldsymbol{\Theta} > 0)$ is also not realistic, because it would mean that the P&L is always positive whatever the market. Therefore, two main configurations are interesting:

(a) a negative gamma exposure with a positive theta;

(b) a positive gamma exposure with a negative theta.

We have represented these two cases in Figure 9.9, and we notice that they lead to different P&L profiles[9]:

(a) If the gamma is negative, the best situation is obtained when the asset price does not move. Any changes in the asset price reduce the P&L, which can be negative if the gamma effect is more important than the theta effect. We also notice that the gain is bounded and the loss is unbounded in this configuration.

(b) If the theta is negative, the loss is bounded and maximum when the asset price does not move. Any changes in the asset price increase the P&L because the gamma is positive. In this configuration, the gain is unbounded.

In order to understand these P&L profiles, we have represented the gamma and theta effects in Figure 9.10 for the case (b). The portfolio is long on a call option and short on the delta neutral hedging strategy. The parameters are the following: $S_0 = 98$, $K = 100$, $\sigma = 10\%$, $b = 5\%$, $r = 5\%$ and $T = 0.25$. The value of the option is equal to 1.601 and we have $\boldsymbol{\Delta}(t_0) = 44.87\%$. In the first panel in Figure 9.10, we have reported the option price (solid curve) and the delta hedging strategy (dashed line) at the current date t_0 when the asset price moves. The area between the two curves represents the gamma effect. We notice that it is positive. For instance, we have $\boldsymbol{\Gamma}(t_0) = 11.55\%$. We do not rebalance the portfolio until time $t = t_0 + \mathrm{d}t$ where $\mathrm{d}t = 0.15$. The dashed curve indicates the value of the option price[10] at the date t. The area between $\boldsymbol{\mathcal{C}}(t, S(t))$ (dashed curve) and $\boldsymbol{\mathcal{C}}(t_0, S(t))$ (solid curve) represents the theta effect. We notice that it is negative[11]. In the second panel, we have reported the resulting P&L. This is the difference between the first area (positive gamma effect) and the second area (negative theta effect). We retrieve the results given in the second panel in Figure 9.9.

[9] We have also indicated the case (a') where the gamma is equal to zero. In this case, we obtain a gamma neutral hedging portfolio and it is not necessary to adjust frequently the hedging portfolio.

[10] We use the same parameters, except that the maturity is now equal to 0.10.

[11] We have $\boldsymbol{\Theta}(t_0, S_0) = -7.09$.

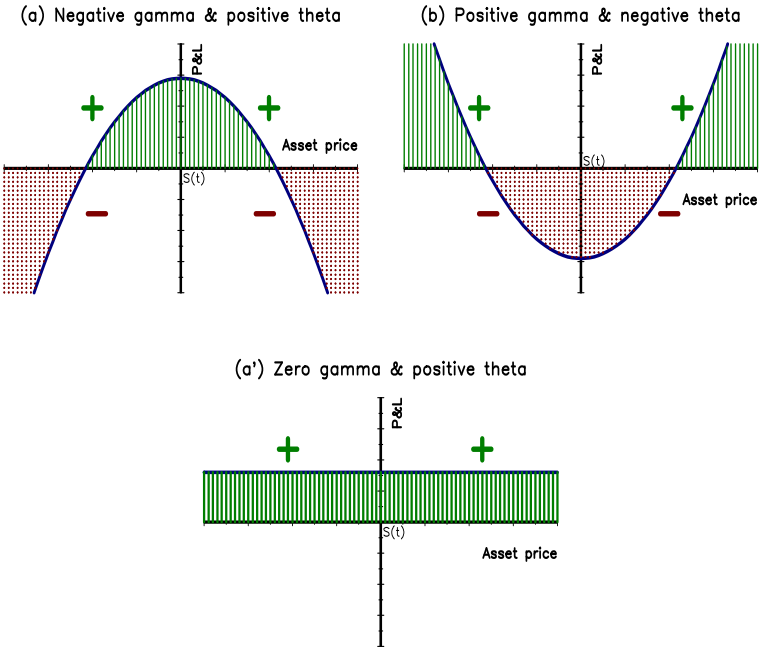


FIGURE 9.9: P&L of the delta neutral hedging portfolio

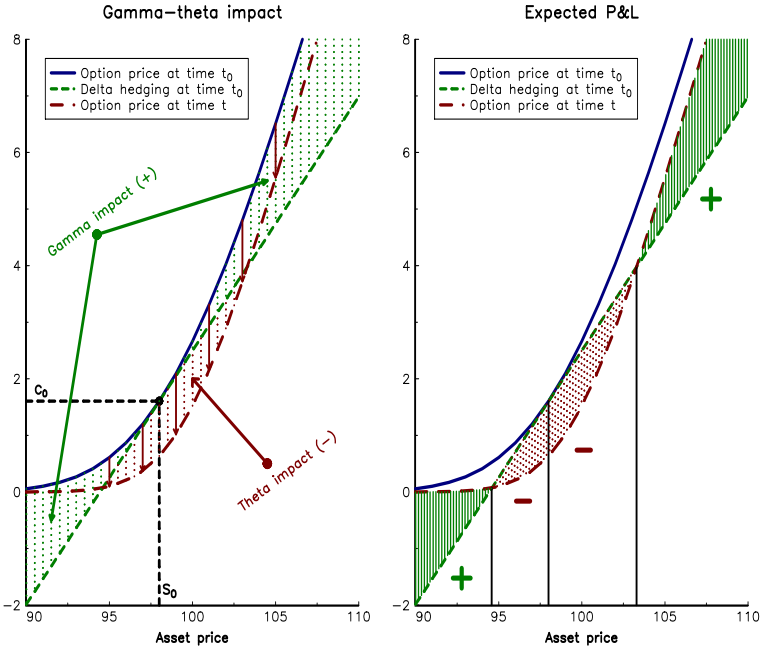


FIGURE 9.10: Illustration of the configuration ($\mathbf{\Gamma} > 0, \mathbf{\Theta} < 0$)

9.1.1.4 The implied volatility

Definition In the Black-Scholes formula, all the parameters are objective except the volatility σ. To calibrate this parameter, we can use a historical estimate $\hat{\sigma}$. However, the option prices computed with the historical volatility $\hat{\sigma}$ do not fit the option prices observed in the market. In practice, we use the Black-Scholes formula to deduce the implied volatility that gives the market prices:

$$f_{\text{BS}}\left(S_0, K, \sigma_{\text{implied}}, T, b, r\right) = V\left(T, K\right)$$

where f_{BS} is the Black-scholes formula and $V\left(T, K\right)$ is the market price of the option, whose maturity date is T and whose strike is K. By convention, the implied volatility is denoted by Σ, and is a function of the parameters[12] T and K:

$$\sigma_{\text{implied}} = \Sigma\left(T, K\right)$$

Example 80 *We consider a call option, whose maturity is one year. The current price of the underlying asset is normalized and is equal to* 100*. Moreover, the risk-free rate and the cost-of-carry parameter are equal to* 5%*. Below, we report the market price of European call options of three assets for several strikes:*

K	90	95	98	100	101	102	105	110
$\boldsymbol{\mathcal{C}}_1\left(T, K\right)$	16.70	13.35	11.55	10.45	9.93	9.42	8.02	6.04
$\boldsymbol{\mathcal{C}}_2\left(T, K\right)$	18.50	14.50	12.00	10.45	9.60	9.00	7.50	5.70
$\boldsymbol{\mathcal{C}}_3\left(T, K\right)$	18.00	14.00	11.80	10.45	9.90	9.50	8.40	7.40

TABLE 9.4: Implied volatility $\Sigma\left(T, K\right)$

K	90	95	98	100	101	102	105	110
$\Sigma_1\left(T, K\right)$	20.00	20.01	19.99	20.0	20.01	19.99	20.00	20.00
$\Sigma_2\left(T, K\right)$	26.18	23.41	21.24	20.0	19.14	18.90	18.69	19.14
$\Sigma_3\left(T, K\right)$	24.53	21.95	20.68	20.0	19.93	20.20	20.95	23.43

For each asset and each strike, we calculate $\Sigma\left(T, K\right)$ and report the results in Table 9.4 and Figure 9.11. For the first set $\boldsymbol{\mathcal{C}}_1$ of options, the implied volatility is constant. In the case of the options $\boldsymbol{\mathcal{C}}_2$, the implied volatility is decreasing with respect to the strike K. In the third case, the implied volatility is decreasing for in-the-money options and increasing for out-of-the-money options.

Remark 99 *When the curve of implied volatility is decreasing and increasing, the curve is called a volatility smile. When the curve of implied volatility is just decreasing, it is called a volatility skew. If we consider the maturity dimension, the term structure of implied volatility is known as the volatility surface.*

Relationship between the implied volatility and the risk-neutral density Breeden and Litzenberger (1978) showed that volatility smile and risk-neutral density are related. Let $\boldsymbol{\mathcal{C}}_t\left(T, K, \right)$ be the market price of the European call option at time t, whose maturity is

[12] $\Sigma\left(T, K\right)$ also depends on the other parameters S_0, b and r, but they are fixed values at the current date t_0.

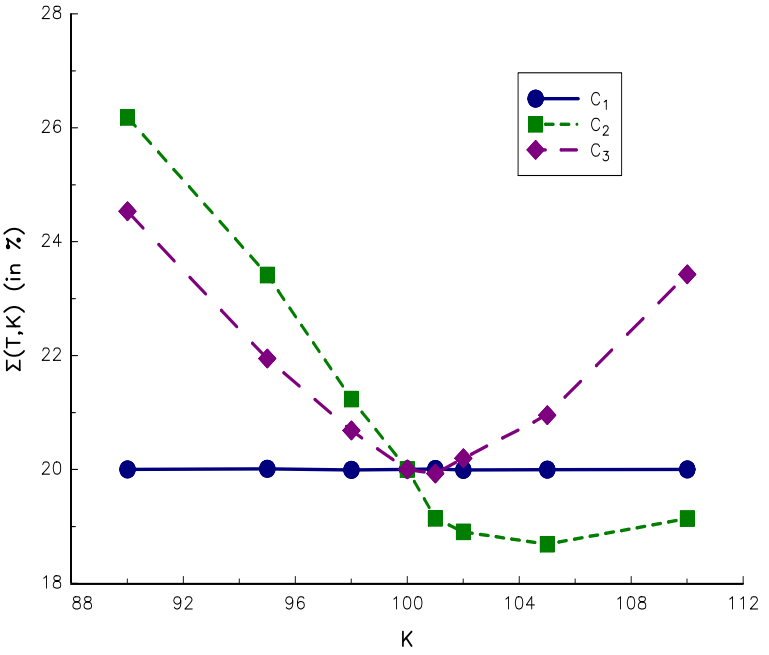


FIGURE 9.11: Volatility smile

T and strike is K. We have:

$$\begin{aligned}
\boldsymbol{\mathcal{C}}_t(T,K) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r\,\mathrm{d}s}\left(S(T)-K\right)^+\middle|\mathcal{F}_t\right] \\
&= e^{-r(T-t)}\int_{-\infty}^{\infty}(S-K)^+\,q_t(T,S)\,\mathrm{d}S \\
&= e^{-r(T-t)}\int_{K}^{\infty}(S-K)\,q_t(T,S)\,\mathrm{d}S
\end{aligned}$$

where $q_t(T,S)$ is the risk-neutral probability density function of $S(T)$ at time t. By definition, the risk-neutral cumulative distribution function $\mathbb{Q}_t(T,S)$ is equal to[13]:

$$\mathbb{Q}_t(T,S)=\int_{-\infty}^{S}q_t(T,x)\,\mathrm{d}x$$

We deduce that:

$$\begin{aligned}
\frac{\partial\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial K} &= -e^{-r(T-t)}\int_{K}^{\infty}q_t(T,S)\,\mathrm{d}S \\
&= -e^{-r(T-t)}\left(1-\mathbb{Q}_t(T,K)\right)
\end{aligned}$$

and:

$$\frac{\partial^2\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial K^2}=e^{-r(T-t)}q_t(T,K)$$

[13] We use the notations $\mathbb{Q}_t(T,S)$ and $q_t(T,S)$ instead of $\mathbb{Q}(S)$ and $q(S)$ because they will be convenient when considering the local volatility model.

It follows that the risk-neutral cumulative distribution function is related to the first derivative of the call option price:

$$\begin{aligned}\mathbb{Q}_t(T,K) &= \Pr\{S(T) \leq K \mid \mathcal{F}_t\} \\ &= 1 + e^{r(T-t)} \cdot \partial_K \boldsymbol{\mathcal{C}}_t(T,K)\end{aligned}$$

We note $\Sigma_t(T,K)$ the volatility surface and $\boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma)$ the Black-Scholes formula. It follows that:

$$\begin{aligned}\mathbb{Q}_t(T,K) &= 1 + e^{r(T-t)} \cdot \partial_K \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t(T,K)) + \\ &\quad e^{r(T-t)} \cdot \partial_\Sigma \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t(T,K)) \cdot \partial_K \Sigma_t(T,K)\end{aligned}$$

where:

$$\partial_K \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma) = -e^{-r(T-t)} \cdot \Phi(d_2)$$

and:

$$\partial_\Sigma \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma) = S(t) \cdot e^{(b-r)(T-t)} \cdot \sqrt{T-t} \cdot \phi\left(d_2 + \Sigma\sqrt{T-t}\right)$$

If we are interested in the risk-neutral probability density function, we obtain:

$$\begin{aligned}q_t(T,K) &= \partial_K \mathbb{Q}_t(T,K) \\ &= e^{r(T-t)} \cdot \partial_K^2 \boldsymbol{\mathcal{C}}_t(T,K)\end{aligned}$$

where:

$$\begin{aligned}\partial_K^2 \boldsymbol{\mathcal{C}}_t(T,K) &= \partial_K^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t) + \\ &\quad 2 \cdot \partial_{K,\Sigma}^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t) \cdot \partial_K \Sigma_t(T,K) + \\ &\quad \partial_\Sigma \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t) \cdot \partial_K^2 \Sigma_t(T,K) + \\ &\quad \partial_\Sigma^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma_t) \cdot (\partial_K \Sigma_t(T,K))^2\end{aligned}$$

and:

$$\begin{aligned}\partial_K^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma) &= e^{-r(T-t)} \frac{\phi(d_2)}{K\Sigma\sqrt{T-t}} \\ \partial_{K,\Sigma}^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma) &= e^{(b-r)(T-t)} \frac{S(t)\, d_1 \phi(d_1)}{\Sigma K} \\ \partial_\Sigma^2 \boldsymbol{\mathcal{C}}_t^\star(T,K,\Sigma) &= e^{(b-r)(T-t)} \frac{S(t)\, d_1 d_2 \sqrt{T-t}\phi(d_1)}{\Sigma}\end{aligned}$$

Example 81 *We assume that* $S(t) = 100$, $T - t = 10$, $b = r = 5\%$ *and:*

$$\Sigma_t(T,K) = 0.25 + \ln\left(1 + 10^{-6}(K-90)^2 + 10^{-6}(K-180)^2\right)$$

In Figure 9.12, we have represented the volatility surface and the associated risk-neutral probability density function. In fact, they both contain the same information, but professionals are more familiar with implied volatilities than risk-neutral probabilities. We have also reported the Black-Scholes risk-neutral distribution by considering the at-the-money implied volatility. We notice that the Black-Scholes model underestimates the probability of extreme events in this example.

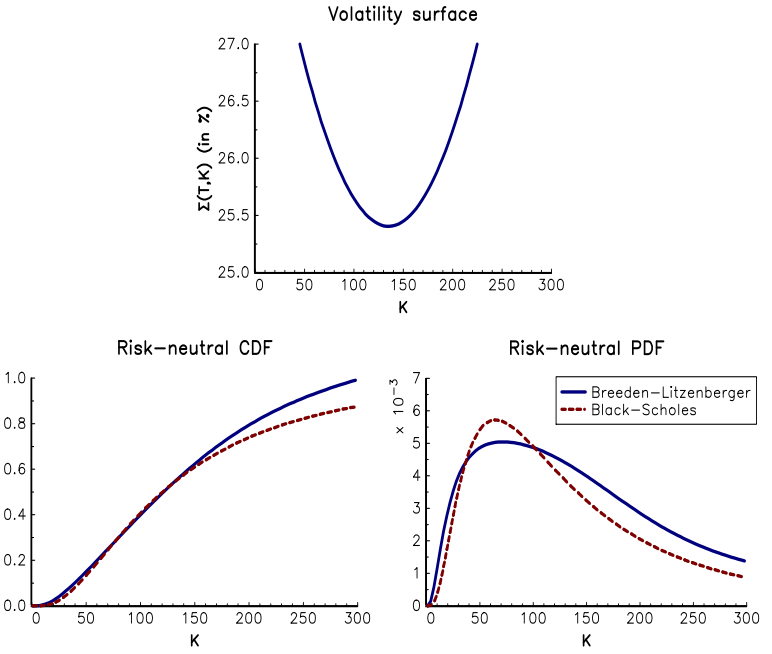


FIGURE 9.12: Risk-neutral probability density function

Robustness of the Black-Scholes formula El Karoui *et al.* (1998) assume that the underlying price process is given by:

$$\mathrm{d}S(t) = \mu(t) S(t)\ \mathrm{d}t + \sigma(t) S(t)\ \mathrm{d}W(t) \tag{9.9}$$

whereas the trader hedges the call option with the implied volatility $\Sigma(T,K)$, meaning that the risk-neutral process is:

$$\mathrm{d}S(t) = rS(t)\ \mathrm{d}t + \Sigma(T,K) S(t)\ \mathrm{d}W^{\mathbb{Q}}(t) \tag{9.10}$$

We reiterate that the dynamics of the replicating portfolio is:

$$\begin{aligned}
\mathrm{d}V(t) &= \phi(t)\ \mathrm{d}S(t) + \psi(t)\ \mathrm{d}B(t) \\
&= \phi(t)\ \mathrm{d}S(t) + \frac{(V(t) - \phi(t) S(t))}{B(t)} rB(t)\ \mathrm{d}t \\
&= \phi(t)\ \mathrm{d}S(t) + r(V(t) - \phi(t) S(t))\ \mathrm{d}t \\
&= rV(t)\ \mathrm{d}t + \phi(t)\ (\mathrm{d}S(t) - rS(t)\ \mathrm{d}t)
\end{aligned}$$

Since $\boldsymbol{\mathcal{C}}(t) = \boldsymbol{\mathcal{C}}(t, S(t))$, we also have:

$$\begin{aligned}
\mathrm{d}\boldsymbol{\mathcal{C}}(t) &= \left(\partial_t \boldsymbol{\mathcal{C}}(t, S(t)) + \frac{1}{2}\sigma^2(t) S^2(t) \partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))\right) \mathrm{d}t + \\
&\partial_S \boldsymbol{\mathcal{C}}(t, S(t))\ \mathrm{d}S(t)
\end{aligned}$$

Using the PDE (9.2), we notice that:

$$\begin{aligned}
\partial_t \boldsymbol{\mathcal{C}}(t, S(t)) &= r\boldsymbol{\mathcal{C}}(t, S(t)) - rS(t) \partial_S \boldsymbol{\mathcal{C}}(t, S(t)) - \\
&\frac{1}{2}\Sigma^2(T,K) S^2(t) \partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))
\end{aligned}$$

We deduce that:

$$\begin{aligned} \mathrm{d}\boldsymbol{\mathcal{C}}(t) &= r\boldsymbol{\mathcal{C}}(t, S(t))\,\mathrm{d}t + \\ &\quad \partial_S \boldsymbol{\mathcal{C}}(t, S(t))\,(\mathrm{d}S(t) - rS(t)\,\mathrm{d}t) + \\ &\quad \frac{1}{2}\left(\sigma^2(t) - \Sigma^2(T,K)\right) S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))\,\mathrm{d}t \end{aligned}$$

We consider the hedging error defined by:

$$e(t) = V(t) - \boldsymbol{\mathcal{C}}(t)$$

Since $\phi(t) = \partial_S \boldsymbol{\mathcal{C}}(t, S(t))$, we have:

$$\begin{aligned} \mathrm{d}e(t) &= \mathrm{d}V(t) - \mathrm{d}\boldsymbol{\mathcal{C}}(t) \\ &= rV(t)\,\mathrm{d}t + \phi(t)\,(\mathrm{d}S(t) - rS(t)\,\mathrm{d}t) - r\boldsymbol{\mathcal{C}}(t, S(t))\,\mathrm{d}t - \\ &\quad \partial_S \boldsymbol{\mathcal{C}}(t, S(t))\,(\mathrm{d}S(t) - rS(t)\,\mathrm{d}t) + \\ &\quad \frac{1}{2}\left(\Sigma^2(T,K) - \sigma^2(t)\right) S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))\,\mathrm{d}t \\ &= re(t)\,\mathrm{d}t + \frac{1}{2}\left(\Sigma^2(T,K) - \sigma^2(t)\right) S^2(t)\,\partial_S^2 \boldsymbol{\mathcal{C}}(t, S(t))\,\mathrm{d}t \end{aligned}$$

We deduce that[14]:

$$V(T) - \boldsymbol{\mathcal{C}}(T) = \frac{1}{2}\int_0^T e^{r(T-t)} \boldsymbol{\Gamma}(t)\left(\Sigma^2(T,K) - \sigma^2(t)\right) S^2(t)\,\mathrm{d}t \tag{9.11}$$

This equation is know as the robustness formula of Black-Scholes hedging (El Karoui *et al.*, 1998). Formula (9.11) is one of the most important results of this chapter. Indeed, since the gamma coefficient of a call option is always positive, we can obtain an almost sure P&L if the implied volatility is larger than the realized volatility and if there is no jump. More generally, the previous result is valid for all types of European options:

$$V(T) - f(S(T)) = \frac{1}{2}\int_0^T e^{r(T-t)} \boldsymbol{\Gamma}(t)\left(\Sigma^2(T,K) - \sigma^2(t)\right) S^2(t)\,\mathrm{d}t \tag{9.12}$$

where $f(S(T))$ is the payoff of the option. We obtain the following results:

- if $\boldsymbol{\Gamma}(t) \geq 0$, a positive P&L is achieved by overestimating the realized volatility:

$$\Sigma(T,K) \geq \sigma(t) \Longrightarrow V(T) \geq f(S(T))$$

- if $\boldsymbol{\Gamma}(t) \leq 0$, a positive P&L is achieved by underestimating the realized volatility:

$$\Sigma(T,K) \leq \sigma(t) \Longrightarrow V(T) \geq f(S(T))$$

- the variance of the hedging error is an increasing function of the absolute value of the gamma coefficient:

$$|\boldsymbol{\Gamma}(t)| \nearrow \Longrightarrow \operatorname{var}(V(T) - f(S(T))) \nearrow$$

In terms of model risk, the robustness formula highlights the role of the implied volatility, the realized volatility and the gamma coefficient. An important issue concerns the case when the gamma can be positive and negative and changes sign during the life of the option. We cannot then control the P&L by using a lower or an upper bound for the implied volatility[15].

[14] Because we have $e(0) = V(0) - \boldsymbol{\mathcal{C}}(0) = 0$.

[15] This issue is solved on page 530.

Example 82 *We consider the replication of 100 ATM call options. The current price of the asset is* 100 *and the maturity of the option is* 6 *months (or* 130 *trading days). We consider the following parameters:* $b = r = 5\%$*. We rebalance the delta hedging portfolio every trading day. Moreover, we assume that the option is priced and hedged with a* 20% *implied volatility.*

Figure 9.13 represents the density function of the hedging ratio π. In the case where the realized volatility $\sigma(t)$ is equal to the implied volatility, we retrieve the previous results: π is centered around zero. However, if the realized volatility $\sigma(t)$ is below (or above) the implied volatility, π is shifted to the right (or the left). If $\sigma(t) < \Sigma$, then there is a higher probability that the trader makes a profit. In our example, we obtain:

$$\Pr\{\pi > 0 \mid \Sigma = 20\%, \sigma = 15\%\} = 99.04\%$$

and:

$$\Pr\{\pi > 0 \mid \Sigma = 20\%, \sigma = 25\%\} = 0.09\%$$

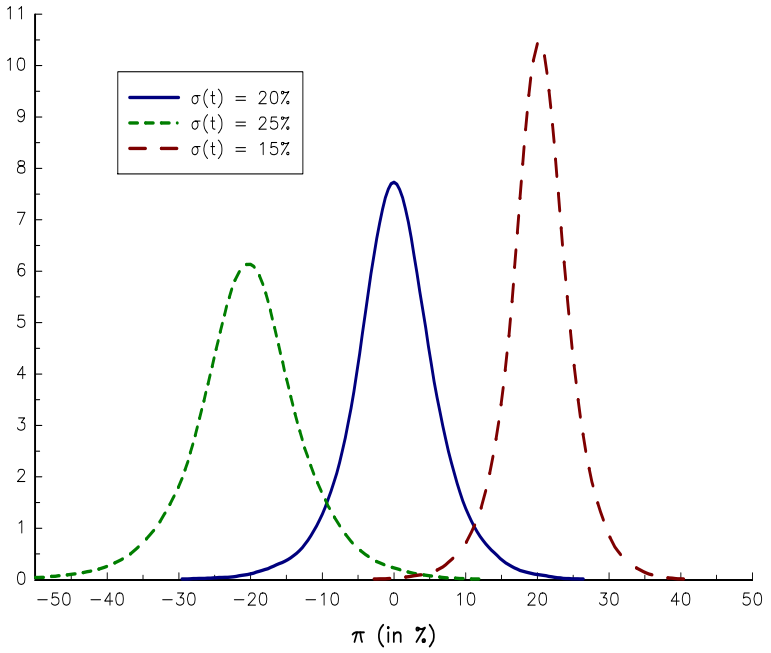


FIGURE 9.13: Hedging error when the implied volatility is 20%

9.1.2 Interest rate risk modeling

Even if the Vasicek model is not used today by practitioners, it is interesting to study it in order to understand the calibration challenge when considering fixed income derivatives. Indeed, in the Black-Scholes model, the calibration consists in estimating a few number of parameters and the main issue concerns the implied volatility. We will see that pricing exotic fixed income derivatives is a more difficult task, because the choice of the risk factors is not obvious and may depend on the tractability of the pricing model[16].

[16] We invite the reader to refer to the book of Brigo and Mercurio (2006) for a more comprehensive presentation on the pricing of fixed income derivatives.

9.1.2.1 Pricing zero-coupon bonds with the Vasicek model

Vasicek (1977) assumes that the state variable is the instantaneous interest rate and follows an Ornstein-Uhlenbeck process:

$$\begin{cases} \mathrm{d}r(t) = a(b - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \\ r(t_0) = r_0 \end{cases}$$

We recall that a zero-coupon bond is a bond that pays \$1 at the maturity date T. Therefore, we have $V(T,r) = 1$ if we note $V(t,r)$ the price of the zero-coupon bond at time t when the interest rate $r(t)$ is equal to r. The corresponding partial differential equation becomes then:

$$\frac{1}{2}\sigma^2\frac{\partial^2 V(t,r)}{\partial r^2} + (a(b - r(t)) - \lambda(t)\sigma)\frac{\partial V(t,r)}{\partial r} + \frac{\partial V(t,r)}{\partial t} - r(t)V(t,r) = 0$$

By applying the Feynman-Kac representation theorem, we deduce that:

$$V(0,r_0) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(t)\,\mathrm{d}t}\middle|\,\mathcal{F}_0\right] \tag{9.13}$$

where the risk-neutral dynamic of $r(t)$ is:

$$\begin{cases} \mathrm{d}r(t) = (a(b - r(t)) - \lambda(t)\sigma)\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t) \\ r(t_0) = r_0 \end{cases}$$

Vasicek (1977) assumes that the risk price of the Wiener process is constant: $\lambda(t) = \lambda$. It follows that the risk-neutral dynamic of $r(t)$ is an Ornstein-Uhlenbeck process:

$$\begin{cases} \mathrm{d}r(t) = a(b' - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t) \\ r(t_0) = r_0 \end{cases}$$

where:

$$b' = b - \frac{\lambda\sigma}{a}$$

We note $Z = \int_0^T r(t)\,\mathrm{d}t$. In Exercise 9.4.2 on page 593, we show that Z is a Gaussian random variable where:

$$\mathbb{E}[Z] = bT + (r_0 - b)\left(\frac{1 - e^{-aT}}{a}\right)$$

and:

$$\mathrm{var}(Z) = \frac{\sigma^2}{a^2}\left(T - \left(\frac{1 - e^{-aT}}{a}\right) - \frac{1}{2a}\left(1 - e^{-aT}\right)^2\right)$$

We deduce that:

$$\begin{aligned} V(0,r_0) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-Z}\middle|\,\mathcal{F}_0\right] \\ &= \exp\left(-\mathbb{E}^{\mathbb{Q}}[Z] + \frac{1}{2}\mathrm{var}^{\mathbb{Q}}(Z)\right) \\ &= \exp\left(-r_0\beta - \left(b' - \frac{\sigma^2}{2a^2}\right)(T - \beta) - \frac{\sigma^2\beta^2}{4a}\right) \end{aligned}$$

where:

$$\beta = \frac{1 - e^{-aT}}{a}$$

If we use the standard notation $B(t,T)$, we have $B(t,T) = V(T-t, r(t))$. We recall that the zero-coupon rate $R(t,T)$ is defined by:

$$B(t,T) = e^{-(T-t)R(t,T)}$$

We deduce that:

$$\begin{aligned} R(t,T) &= -\frac{1}{T-t}\ln B(t,T) \\ &= \frac{r_t\beta}{T-t} + \left(b' - \frac{\sigma^2}{2a^2}\right)\left(\frac{T-t-\beta}{T-t}\right) + \frac{\sigma^2\beta^2}{4a(T-t)} \\ &= \left(b' - \frac{\sigma^2}{2a^2}\right) + \left(r_t - b' + \frac{\sigma^2}{2a^2}\right)\frac{\beta}{T-t} + \frac{\sigma^2\beta^2}{4a(T-t)} \end{aligned}$$

Since we have:

$$R_\infty = \lim_{T\to\infty} R(t,T) = b' - \frac{\sigma^2}{2a^2}$$

the zero-coupon rate has the following expression:

$$R(t,T) = R_\infty + (r_t - R_\infty)\left(\frac{1-e^{-a(T-t)}}{a(T-t)}\right) + \frac{\sigma^2\left(1-e^{-a(T-t)}\right)^2}{4a^3(T-t)} \tag{9.14}$$

The yield curve can take three different forms (Figure 9.14). Vasicek (1977) shows that the curve is increasing if $r_t \leq R_\infty - \dfrac{\sigma^2}{4a^2}$ and decreasing if $r_t \geq R_\infty + \dfrac{\sigma^2}{2a^2}$. Otherwise, it is a bell curve.

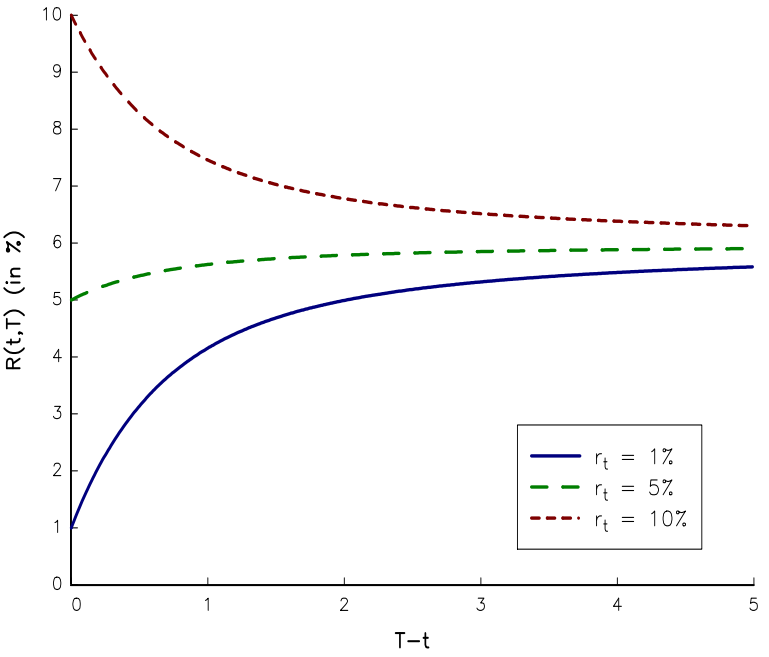


FIGURE 9.14: Vasicek model ($a = 2.5$, $b = 6\%$ and $\sigma = 5\%$)

Let $F(t,T_1,T_2)$ be the forward rate at time t for the period $[T_1,T_2]$. It verifies the following relationship:

$$B(t,T_2) = e^{-(T_2-T_1)F(t,T_1,T_2)} B(t,T_1)$$

We deduce that the expression of $F(t,T_1,T_2)$ is:

$$F(t,T_1,T_2) = -\frac{1}{(T_2-T_1)} \ln \frac{B(t,T_2)}{B(t,T_1)}$$

It follows that the instantaneous forward rate is given by this equation[17]:

$$f(t,T) = F(t,T,T) = -\frac{\partial \ln B(t,T)}{\partial T}$$

Using Equation (9.14), we deduce another expression of the price of the zero-coupon bond:

$$B(t,r_t) = \exp\left(-(T-t)R_\infty - (r_t - R_\infty)\left(\frac{1-e^{-a(T-t)}}{a}\right) - \frac{\sigma^2\left(1-e^{-a(T-t)}\right)^2}{4a^3}\right)$$

Therefore, the instantaneous forward rate in the Vasicek model is:

$$f(t,T) = R_\infty + (r_t - R_\infty)e^{-a(T-t)} + \frac{\sigma^2\left(1-e^{-a(T-t)}\right)e^{-a(T-t)}}{2a^2}$$

Remark 100 *Forward rates are interest rates that are locked in forward rate agreements (FRA). It involves two dates: T_1 is the start of the period the rate will be fixed for, and T_2 is the maturity date of the FRA. $T_2 - T_1$ is the maturity of the locked interest rate. It is also called the tenor of the interest rate that is being fixed. Therefore, $F(t,T_1,T_2)$ is the forward value of the spot rate $R(t,T_2-T_1)$.*

9.1.2.2 The calibration issue of the yield curve

Hull and White (1990) propose to extend the Vasicek model by considering that the three parameters a, b and σ are deterministic functions of time. Under the risk-neutral probability measure, the dynamics of the interest rate is then:

$$\mathrm{d}r(t) = a(t)(b(t) - r(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

The underlying idea is to fit the term structure of interest rates and other quantities, such as the term structure of spot volatilities. However, the generalized Vasicek model produces unrealistic volatility term structures. Therefore, Hull and White (1994) focused on this extension:

$$\begin{aligned}\mathrm{d}r(t) &= a(b(t) - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t) \\ &= (\theta(t) - ar(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t)\end{aligned}$$

[17] We also notice that $B(t,T)$ can be expressed in terms of instantaneous forward rates:

$$B(t,T) = e^{-\int_t^T f(t,u)\,\mathrm{d}u}$$

where $\theta(t) = a \cdot b(t)$. If we want to fit exactly the yield curve, we can consider arbitrary values for the parameters a and σ, because the calibration of the yield curve is done by the time-varying mean-reverting parameter:

$$\theta(t) = \frac{\partial f(0,t)}{\partial t} + af(0,t) + \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)$$

or:

$$b(t) = f(0,t) + \frac{1}{a}\partial_t f(0,t) + \frac{\sigma^2}{2a^2}\left(1 - e^{-2at}\right) \tag{9.15}$$

We notice that $b(t)$ depends on the instantaneous forward rate, which is the first derivative of the price of the zero-coupon bond.

Example 83 *We assume that the zero-coupon rates are given by the Nelson-Siegel model with* $\theta_1 = 5.5\%$, $\theta_2 = 0.5\%$, $\theta_3 = -4.5\%$ *and* $\theta_4 = 1.8$.

We reiterate that the spot rate $R(t,T)$ in the Nelson-Siegel model is equal to:

$$R(t,T) = \theta_1 + \theta_2\left(\frac{1 - e^{-(T-t)/\theta_4}}{(T-t)/\theta_4}\right) + \theta_3\left(\frac{1 - e^{-(T-t)/\theta_4}}{(T-t)/\theta_4} - e^{-(T-t)/\theta_4}\right)$$

We deduce that the instantaneous forward rate corresponds to the following expression:

$$\begin{aligned} f(t,T) &= \frac{\partial\,(T-t)\,R(t,T)}{\partial T} \\ &= \theta_1 + \theta_2 e^{-(T-t)/\theta_4} + \frac{\theta_3(T-t)}{\theta_4}e^{-(T-t)/\theta_4} \end{aligned}$$

For the slope, we have:

$$\frac{\partial f(t,T)}{\partial T} = \left(\frac{(\theta_3 - \theta_2)}{\theta_4} - \frac{\theta_3(T-t)}{\theta_4^2}\right)e^{-(T-t)/\theta_4}$$

Fitting exactly the Nelson-Siegel yield curve is then equivalent to define the time-varying mean-reverting parameter $b(t)$ of the extended Vasicek model as follows:

$$\begin{aligned} b(t) &= \theta_1 + \theta_2 e^{-t/\theta_4} + \frac{\theta_3 t}{\theta_4}e^{-t/\theta_4} + \frac{\sigma^2}{2a^2}\left(1 - e^{-2at}\right) + \\ &\quad \frac{1}{a}\left(\frac{(\theta_3 - \theta_2)}{\theta_4} - \frac{\theta_3 t}{\theta_4^2}\right)e^{-t/\theta_4} \\ &= \theta_1 + \left(\left(\theta_2 + \frac{\theta_3 t}{\theta_4}\right)\left(1 - \frac{1}{a\theta_4}\right) + \frac{\theta_3}{a\theta_4}\right)e^{-t/\theta_4} + \\ &\quad \frac{\sigma^2}{2a^2}\left(1 - e^{-2at}\right) \end{aligned}$$

In Figure 9.15, we have represented the yield curve obtained with the Nelson-Siegel model in the top/left panel. We have also reported the curve of instantaneous forward rates in the top/right panel. The bottom/left panel corresponds to the time-varying mean-reverting parameter $b(t)$. We have used three set of parameters (a, σ). Finally, we have recalculated the yield curve of the extended Vasicek model in the bottom/right panel. We retrieve the original yield curve. We can compare this solution with those obtained by minimizing the sum of the squared residuals:

$$\left(\hat{r}_0, \hat{a}, \hat{b}, \hat{\sigma}\right) = \arg\min \sum_i \left(R^{\text{NS}}(t,T_i) - R(t,T_i; r_0, a, b, \sigma)\right)^2$$

where $R^{\text{NS}}(t, T_i)$ is the Nelson-Siegel spot rate, $R(t, T_i; r_0, a, b, \sigma)$ is the theoretical spot rate of the Vasicek model and i denotes the i^{th} observation. By considering all the maturities between zero and twenty years with a step of one month, we obtain $\hat{r}_0 = 6\%$, $\hat{a} = 16.88$, $\hat{b} = 7.47\%$ and $\hat{\sigma} = 3.91\%$. Unfortunately, the fitted Vasicek model (curve #2) does not reproduce the original yield curve contrary to the fitted extended Vasicek model (curve #1).

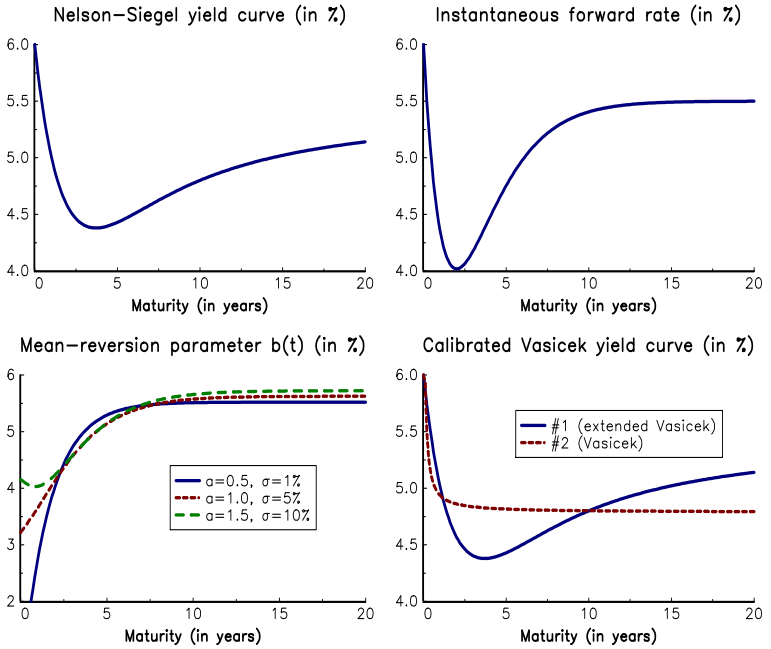


FIGURE 9.15: Calibration of the Vasicek model

The yield curve is not the only market information to calibrate. More generally, the calibration set of an interest rate model also includes caplets, floorlets and swaptions (Brigo and Mercurio, 2006). This explains that pricing interest rate exotic options is more difficult than pricing equity exotic options, and one-factor models based on the short rate are not sufficient, because it is not possible to calibrate caps, floors and swaptions.

9.1.2.3 Caps, floors and swaptions

We consider a number of future dates $T_0, T_1, \ldots, T_n$, and we assume that the period between two dates T_i and T_{i-1} is approximately constant (e.g. 3M or 6M). A caplet is the analog of a call option, whose underlying asset is a forward rate. It is defined by the payoff $(T_i - T_{i-1})(F(T_{i-1}, T_{i-1}, T_i) - K)^+$, where K is the strike of the caplet and $F(T_{i-1}, T_{i-1}, T_i)$ is the forward rate at the future date T_{i-1}. $\delta_{i-1} = T_i - T_{i-1}$ is then the tenor of the caplet, T_{i-1} is the resetting date (or the fixing date) of the forward rate whereas T_i is the maturity date of the caplet. A cap is a portfolio of successive caplets[18]:

$$\text{Cap}(t) = \sum_{i=1}^{n} \text{Caplet}(t, T_{i-1}, T_i)$$

[18] We have $t \leq T_0$.

Similarly, a floor is a portfolio of successive floorlets:

$$\text{Floor}(t) = \sum_{i=1}^{n} \text{Floorlet}(t, T_{i-1}, T_i)$$

where the payoff of the floorlet is $(T_i - T_{i-1})(K - F(T_{i-1}, T_{i-1}, T_i))^+$.

A par swap rate is the fixed rate of an interest rate swap[19]:

$$\text{Sw}(t) = \frac{B(t, T_0) - B(t, T_n)}{\sum_{i=1}^{n} (T_i - T_{i-1}) \cdot B(t, T_i)}$$

Then, we define the payoff of a payer swaption as[20]:

$$(\text{Sw}(T_0) - K)^+ \sum_{i=1}^{n} (T_i - T_{i-1}) B(T_0, T_i)$$

where $\text{Sw}(T_0)$ is the forward swap rate.

Remark 101 *Generally, caps, floors and swaptions are written on the Libor rate, which is defined as a simple forward rate:*

$$L(t, T_{i-1}, T_i) = \frac{1}{T_i - T_{i-1}} \left(\frac{B(t, T_{i-1})}{B(t, T_i)} - 1 \right)$$

In order to price these interest rate products, we can use the risk-neutral probability measure $\mathbb{Q}$, and we have[21]:

$$\text{Caplet}(t, T_{i-1}, T_i) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^{T_i} r(s)\,\mathrm{d}s} \delta_{i-1} (L(T_{i-1}, T_{i-1}, T_i) - K)^+ \middle| \mathcal{F}_t \right]$$

and:

$$\text{Swaption}(t) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^{T_n} r(s)\,\mathrm{d}s} (\text{Sw}(T_0) - K)^+ \sum_{i=1}^{n} \delta_{i-1} B(T_0, T_i) \middle| \mathcal{F}_t \right]$$

We face here a problem because the discount factor is stochastic and is not independent from the forward rate $L(T_{i-1}, T_{i-1}, T_i)$ or the forward swap rate $\text{Sw}(T_0)$. Therefore, the risk-neutral transform does not help to price interest rate derivatives.

9.1.2.4 Change of numéraire and equivalent martingale measure

We recall that the price of the contingent claim, whose payoff is $V(T) = f(S(T))$ at time T, is given by:

$$V(0) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_0^T r(s)\,\mathrm{d}s} \cdot V(T) \middle| \mathcal{F}_0 \right]$$

where $\mathbb{Q}$ is the risk-neutral probability measure. We can rewrite this equation as follows:

$$\frac{V(0)}{M(0)} = \mathbb{E}^{\mathbb{Q}} \left[\frac{V(T)}{M(T)} \middle| \mathcal{F}_0 \right] \tag{9.16}$$

[19] $T_0 = t$ corresponds to a spot swap, whereas $T_0 > t$ corresponds to a forward start swap.

[20] The payoff of a receiver swaption is:

$$(K - \text{Sw}(T_0))^+ \sum_{i=1}^{n} (T_i - T_{i-1}) B(T_0, T_i)$$

[21] We recall that δ_{i-1} is equal to $T_i - T_{i-1}$.

where[22]:

$$M(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right)$$

Under the probability measure $\mathbb{Q}$, we know that $\tilde{V}(t) = V(t)/M(t)$ is an $\mathcal{F}_t$-martingale. The money market account $M(t)$ is then the numéraire when the martingale measure is the risk-neutral probability measure[23], but other numéraires can be used in order to simplify pricing problems:

> "*The use of the risk-neutral probability measure has proved to be very powerful for computing the prices of contingent claims [...] We show here that many other probability measures can be defined in the same way to solve different asset-pricing problems, in particular option pricing. Moreover, these probability measure changes are in fact associated with numéraire changes*" (Geman *et al.*, 1995, page 443).

Let us consider another numéraire $N(t) > 0$ and the associated probability measure given by the Radon-Nikodym derivative:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} &= \frac{N(T)/N(0)}{M(T)/M(0)} \\ &= e^{-\int_0^T r(s)\,\mathrm{d}s} \cdot \frac{N(T)}{N(0)}\end{aligned}$$

We have:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}^\star}\left[\frac{V(T)}{N(T)}\middle|\mathcal{F}_0\right] &= \mathbb{E}^{\mathbb{Q}}\left[\frac{V(T)}{N(T)}\cdot\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}}\middle|\mathcal{F}_t\right] \\ &= \frac{M(0)}{N(0)}\cdot\mathbb{E}^{\mathbb{Q}}\left[\frac{V(T)}{M(T)}\middle|\mathcal{F}_0\right] \\ &= \frac{M(0)}{N(0)}\cdot V(0)\end{aligned}$$

We deduce that:

$$\frac{V(0)}{N(0)} = \mathbb{E}^{\mathbb{Q}^\star}\left[\frac{V(T)}{N(T)}\middle|\mathcal{F}_0\right] \tag{9.17}$$

We notice that Equation (9.17) is similar to Equation (9.16), except that we have changed the numéraire ($M(t) \to N(t)$) and the probability measure ($\mathbb{Q} \to \mathbb{Q}^\star$). More generally, we have:

$$V(t) = N(t)\cdot\mathbb{E}^{\mathbb{Q}^\star}\left[\frac{V(T)}{N(T)}\middle|\mathcal{F}_t\right]$$

Thanks to Girsanov theorem, we also notice that $e^{-\int_0^t r(s)\,\mathrm{d}s}N(t)$ is an $\mathcal{F}_t$-martingale.

Example 84 *The forward numéraire is the zero-coupon bond price of maturity T:*

$$N(t) = B(t,T)$$

In this case, the probability measure is called the forward probability and is denoted by $\mathbb{Q}^\star(T)$. This martingale measure has been originally used by Jamshidian (1989) for pricing bond options with the Vasicek model. Another important result is that forward rates are martingales under the forward probability measure (Brigo and Mercurio, 2006).

[22] We note that $M(0) = 1$.
[23] $M(t)$ is also called the spot numéraire.

By noticing that $N(T) = B(T,T) = 1$, Equation (9.17) becomes:

$$V(t) = B(t,T)\,\mathbb{E}^{\mathbb{Q}^\star(T)}\left[V(T)\middle|\mathcal{F}_t\right]$$

For instance, in the case of a caplet, we obtain:

$$\begin{aligned}
\text{Caplet}(t,T_{i-1},T_i) &= \delta_{i-1}\mathbb{E}^{\mathbb{Q}}\left[\frac{M(t)}{M(T_i)}\left(L(T_{i-1},T_{i-1},T_i)-K\right)^+\middle|\mathcal{F}_t\right] \\
&= \delta_{i-1}\mathbb{E}^{\mathbb{Q}^\star(T_i)}\left[\frac{N(t)}{N(T_i)}\left(L(T_{i-1},T_{i-1},T_i)-K\right)^+\middle|\mathcal{F}_t\right] \\
&= \delta_{i-1}B(t,T_i)\,\mathbb{E}^{\mathbb{Q}^\star(T_i)}\left[\left(L(T_{i-1},T_{i-1},T_i)-K\right)^+\middle|\mathcal{F}_t\right]
\end{aligned}$$

where $L(t,T_{i-1},T_i)$ is an $\mathcal{F}_t$-martingale under the forward probability measure $\mathbb{Q}^\star(T_i)$. If we use the standard Black model, we obtain:

$$\text{Caplet}(t,T_{i-1},T_i) = \delta_{i-1}B(t,T_i)\left(L(t,T_{i-1},T_i)\,\Phi(d_1)-K\Phi(d_2)\right) \tag{9.18}$$

where[24]:

$$d_1 = \frac{1}{\sigma_{i-1}\sqrt{T_{i-1}-t}}\ln\frac{L(t,T_{i-1},T_i)}{K}+\frac{1}{2}\sigma_{i-1}\sqrt{T_{i-1}-t}$$

and:

$$d_2 = d_1-\sigma_{i-1}\sqrt{T_{i-1}-t}$$

If we consider other models, the general formula of the caplet price is[25]:

$$\text{Caplet}(t,T_{i-1},T_i) = B(t,T_i)\,\mathbb{E}^{\mathbb{Q}^\star(T_i)}\left[\left(\frac{1}{B(T_{i-1},T_i)}-(1+\delta_{i-1}K)\right)^+\middle|\mathcal{F}_t\right]$$

Example 85 *The annuity numéraire is equal to:*

$$N(t) = \sum_{i=1}^{n}(T_i-T_{i-1})\,B(t,T_i)$$

While the forward swap rate is a martingale under the annuity probability measure $\mathbb{Q}^\star$, the annuity numéraire is used to price a swaption (Brigo and Mercurio, 2006).

[24] σ_{i-1} is the volatility of the Libor rate $L(t,T_{i-1},T_i)$.

[25] We have:

$$\begin{aligned}
\delta_{i-1}\left(L(t,T_{i-1},T_i)-K\right)^+ &= \left(\frac{B(t,T_{i-1})}{B(t,T_i)}-(1+\delta_{i-1}K)\right)^+ \\
&= \frac{\left(B(t,T_{i-1})-(1+\delta_{i-1}K)\,B(t,T_i)\right)^+}{B(t,T_i)}
\end{aligned}$$

and:

$$\delta_{i-1}\left(L(T_{i-1},T_{i-1},T_i)-K\right)^+ = \left(\frac{1}{B(T_{i-1},T_i)}-(1+\delta_{i-1}K)\right)^+$$

We deduce the following pricing formula for the swaption:

$$\begin{aligned}
\text{Swaption}(t) &= \mathbb{E}^{\mathbb{Q}}\left[\frac{M(t)}{M(T_n)}\left(\text{Sw}(T_0)-K\right)^+\sum_{i=1}^{n}\delta_{i-1}B(T_0,T_i)\middle|\mathcal{F}_t\right] \\
&= \mathbb{E}^{\mathbb{Q}^\star}\left[\frac{N(t)}{N(T_0)}\left(\text{Sw}(T_0)-K\right)^+\sum_{i=1}^{n}\delta_{i-1}B(T_0,T_i)\middle|\mathcal{F}_t\right] \\
&= N(t)\,\mathbb{E}^{\mathbb{Q}^\star}\left[\left(\text{Sw}(T_0)-K\right)^+\middle|\mathcal{F}_t\right] \qquad (9.19)\\
&= N(t)\,\mathbb{E}^{\mathbb{Q}^\star}\left[\left(\frac{1-B(T_0,T_n)}{N(T_0)}-K\right)^+\middle|\mathcal{F}_t\right]
\end{aligned}$$

Using Equation (9.19), we can also find a Black formula for the swaption, in exactly the same way as caps and floors. However, we face here an issue. Indeed, it is equivalent to assume that all the forward rates are log-normal under the different forward probability measures $\mathbb{Q}^\star(T_i)$ and the swap rates are also log-normal under the annuity probability measures $\mathbb{Q}^\star$. The problem is that these different forward and swap rates are related, and their dynamics are not independent.

9.1.2.5 The HJM model

Until the beginning of the nineties, the state variable of fixed income models is the instantaneous interest rate $r(t)$. For instance, it is the case of the models of Vasicek (1977) and Cox *et al.* (1985). However, we have seen that we face some calibration issues when considering such framework. Heath *et al.* (1992) propose then that the state variables are forward rates, and not spot rates. Under the risk-neutral probability measure $\mathbb{Q}$, the dynamics of the instantaneous forward rate for the maturity T is given by:

$$f(t,T) = f(0,T) + \int_0^t \alpha(s,T)\,\mathrm{d}s + \int_0^t \sigma(s,T)\,\mathrm{d}W^{\mathbb{Q}}(s)$$

where $f(0,T)$ is the current forward rate. Therefore, the stochastic differential equation is:

$$\mathrm{d}f(t,T) = \alpha(t,T)\,\mathrm{d}t + \sigma(t,T)\,\mathrm{d}W^{\mathbb{Q}}(t) \qquad (9.20)$$

Bond pricing We recall that:

$$B(t,T) = e^{-\int_t^T f(t,u)\,\mathrm{d}u}$$

If we note $X(t) = -\int_t^T f(t,u)\,\mathrm{d}u$, we have:

$$\begin{aligned}
\mathrm{d}X(t) &= f(t,t)\,\mathrm{d}t - \int_t^T \mathrm{d}f(t,u)\,\mathrm{d}u \\
&= f(t,t)\,\mathrm{d}t - \left(\int_t^T \alpha(t,u)\,\mathrm{d}u\right)\mathrm{d}t - \left(\int_t^T \sigma(t,u)\,\mathrm{d}u\right)\mathrm{d}W^{\mathbb{Q}}(t) \\
&= \left(f(t,t) + a(t,T)\right)\mathrm{d}t + b(t,T)\,\mathrm{d}W^{\mathbb{Q}}(t)
\end{aligned}$$

where:

$$a(t,T) = -\int_t^T \alpha(t,u)\,\mathrm{d}u$$

and:

$$b(t,T) = -\int_t^T \sigma(t,u)\,\mathrm{d}u$$

We deduce that:

$$\begin{aligned}\mathrm{d}B(t,T) &= e^{X(t)}\,\mathrm{d}X(t) + \frac{1}{2}e^{X(t)}\left\langle \mathrm{d}X(t), \mathrm{d}X(t)\right\rangle \\ &= \left(f(t,t) + a(t,T) + \frac{1}{2}b^2(t,T)\right) B(t,T)\,\mathrm{d}t + \\ &\quad b(t,T)\,B(t,T)\,\mathrm{d}W^{\mathbb{Q}}(t)\end{aligned}$$

Since $f(t,t)$ is equal to the spot rate $r(t)$, the HJM model implies the following restriction[26]:

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,\mathrm{d}u \tag{9.21}$$

Equation (9.21) is known as the '*drift restriction*' and is necessary to ensure no-arbitrage opportunities. In this case, we verify that the discounted zero-coupon bond is a martingale under the risk-neutral probability measure $\mathbb{Q}$:

$$\mathrm{d}B(t,T) = r(t)\,B(t,T)\,\mathrm{d}t + b(t,T)\,B(t,T)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

Dynamics of spot and forward rates The drift restriction implies that the dynamics of the instantaneous forward rate $f(t,T)$ is given by:

$$\mathrm{d}f(t,T) = \left(\sigma(t,T)\int_t^T \sigma(t,u)\,\mathrm{d}u\right)\mathrm{d}t + \sigma(t,T)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

Therefore, we have:

$$f(t,T) = f(0,T) + \int_0^t \left(\sigma(s,T)\int_s^T \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}s + \int_0^t \sigma(s,T)\,\mathrm{d}W^{\mathbb{Q}}(s)$$

If we are interested in the instantaneous spot rate $r(t)$, we obtain:

$$\begin{aligned} r(t) &= f(t,t) \\ &= r(0) + \int_0^t \left(\sigma(s,t)\int_s^t \sigma(s,u)\,\mathrm{d}u\right)\mathrm{d}s + \int_0^t \sigma(s,t)\,\mathrm{d}W^{\mathbb{Q}}(s)\end{aligned}$$

Forward probability measure We now consider the dynamics of the forward rate $f(t,T_1)$ under the forward probability measure $\mathbb{Q}^\star(T_2)$ with $T_2 \geq T_1$. We reiterate that the new numéraire $N(t)$ is given by:

$$N(t) = B(t,T_2) = e^{-\int_t^{T_2} f(t,u)\,\mathrm{d}u}$$

[26] Indeed, we must have:

$$a(t,T) + \frac{1}{2}b^2(t,T) = 0$$

or:

$$\partial_T a(t,T) = -b(t,T)\cdot\partial_T b(t,T)$$

In Exercise 9.4.5 on page 596, we show that:

$$\mathrm{d}f(t,T_1) = -\left(\sigma(t,T_1)\int_{T_1}^{T_2}\sigma(t,u)\,\mathrm{d}u\right)\mathrm{d}t + \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}^\star(T_2)}(t)$$

It follows that $f(t,T_1)$ is a martingale under the forward probability measure $\mathbb{Q}^\star(T_1)$:

$$\mathrm{d}f(t,T_1) = \sigma(t,T_1)\,\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(t)$$

We can also show that $B(t,T_2)/B(t,T_1)$ is a martingale under $\mathbb{Q}^\star(T_1)$ and we have:

$$B(T_1,T_2) = \frac{B(t,T_2)}{B(t,T_1)}\exp\left(\int_t^{T_1} g(u)\,\mathrm{d}W^{\mathbb{Q}^\star(T_1)}(u) - \frac{1}{2}\int_t^{T_1} g^2(u)\,\mathrm{d}u\right)$$

where:

$$g(t) = b(t,T_2) - b(t,T_1)$$

Some examples If we assume that $\sigma(t,T)$ is constant and equal to σ, we obtain:

$$f(t,T) = f(0,T) + \sigma^2\left(T - \frac{t}{2}\right)t + \sigma W^{\mathbb{Q}}(t)$$

and:

$$r(t) = f(0,t) + \sigma^2\frac{t^2}{2} + \sigma W^{\mathbb{Q}}(t)$$

This case corresponds to the Gaussian model of Ho and Lee (1986).

Brigo and Mercurio (2006) consider the case of separable volatility:

$$\sigma(t,T) = \xi(t)\,\psi(T)$$

We have:

$$\begin{aligned}\mathrm{d}r(t) &= \left(\partial_t f(0,t) + \psi^2(t)\int_0^t \xi^2(s)\,\mathrm{d}s + \frac{(r(t) - f(0,t))}{\psi(t)}\psi'(t)\right)\mathrm{d}t + \\ &\quad \xi(t)\,\psi(t)\,\mathrm{d}W^{\mathbb{Q}}(t)\end{aligned}$$

For example, if we set $\sigma(t,T) = \sigma e^{-a(T-t)}$, we have $\xi(t) = \sigma e^{at}$, $\psi(T) = e^{-aT}$ and[27]:

$$\mathrm{d}r(t) = \left(\partial_t f(0,t) + \sigma^2\left(\frac{1-e^{-2at}}{2a}\right) + a\left(f(0,t) - r(t)\right)\right)\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t)$$

We retrieve the generalized Vasicek model proposed by Hull and White (1994):

$$\mathrm{d}r(t) = a\left(b(t) - r(t)\right)\mathrm{d}t + \sigma\,\mathrm{d}W^{\mathbb{Q}}(t)$$

where $b(t)$ is given by Equation (9.15) on page 517.

[27] We have:

$$\begin{aligned}\psi^2(t)\int_0^t \xi^2(s)\,\mathrm{d}s &= \sigma^2 e^{-2at}\int_0^t e^{2as}\,\mathrm{d}s \\ &= \sigma^2\left(\frac{1-e^{-2at}}{2a}\right)\end{aligned}$$

and:

$$\frac{\psi'(t)}{\psi(t)} = -a$$

Ritchken and Sankarasubramanian (1995) have identified necessary and sufficient conditions on the functions ξ and ψ in order to obtain a Markovian short-rate process. They showed that they must satisfy the following conditions:

$$\xi(t) = \sigma(t) e^{\int_0^t \kappa(s)\,\mathrm{d}s}$$

and:

$$\psi(T) = e^{-\int_0^T \kappa(s)\,\mathrm{d}s}$$

where $\sigma(t)$ and $\kappa(t)$ are two $\mathcal{F}_t$-adapted processes. In this case, we obtain:

$$\sigma(t,T) = \sigma(t) e^{-\int_t^T \kappa(s)\,\mathrm{d}s}$$

For instance, the generalized Vasicek model is a special case of this framework where the two functions $\sigma(t)$ and $\kappa(t)$ are constant[28].

Extension to multi-factor models We can show that the previous results can be extended when we assume that the instantaneous forward rate is given by the following SDE:

$$\mathrm{d}f(t,T) = \alpha(t,T)\,\mathrm{d}t + \sigma(t,T)^\top\,\mathrm{d}W^{\mathbb{Q}}(t)$$

where $W^{\mathbb{Q}}(t) = \left(W_1^{\mathbb{Q}}(t), \ldots, W_n^{\mathbb{Q}}(t)\right)$ is a n-dimensional Brownian motion and ρ is the correlation matrix of $W^{\mathbb{Q}}(t)$. For instance, the drift restriction (9.21) becomes:

$$\alpha(t,T) = \sigma(t,T)^\top \rho \int_t^T \sigma(t,u)\,\mathrm{d}u$$

In the two-dimensional case, we obtain:

$$\begin{aligned}
\mathrm{d}f(t,T) &= \left(\sigma_1(t,T)\int_t^T \sigma_1(t,u)\,\mathrm{d}u\right)\mathrm{d}t + \left(\sigma_2(t,T)\int_t^T \sigma_2(t,u)\,\mathrm{d}u\right)\mathrm{d}t \\
&\quad + \rho_{1,2}\left(\sigma_1(t,T)\int_t^T \sigma_2(t,u)\,\mathrm{d}u + \sigma_1(t,T)\int_t^T \sigma_2(t,u)\,\mathrm{d}u\right)\mathrm{d}t \\
&\quad \sigma_1(t,T)\,\mathrm{d}W_1^{\mathbb{Q}}(t) + \sigma_2(t,T)\,\mathrm{d}W_2^{\mathbb{Q}}(t)
\end{aligned}$$

For example, Heath *et al.* (1992) extend the Vasicek model by assuming that $\sigma_1(t,T) = \sigma_1$, $\sigma_2(t,T) = \sigma_2 e^{-a_2(T-t)}$ and $\rho_{1;2} = 0$. In this case, we obtain:

$$\begin{aligned}
r(t) &= f(0,t) + \sigma_1^2\frac{t^2}{2} + \frac{\sigma_2^2}{a_2^2}\left(\left(1-e^{-a_2 t}\right) - \frac{1}{2}\left(1-e^{-2a_2 t}\right)\right) + \\
&\quad \sigma_1 W_1^{\mathbb{Q}}(t) + \sigma_2\int_0^t e^{-a_2(t-s)}\,\mathrm{d}W_2^{\mathbb{Q}}(s)
\end{aligned}$$

9.1.2.6 Market models

One of the disadvantages of short-rate and HJM models is that they focus on instantaneous spot or forward interest rates. However, these quantities are unobservable. At the end of the nineties, academics have developed two families of models in order to bypass these disadvantages: the Libor market model (LMM) and the swap market model (SMM).

[28] We have $\sigma(t) = \sigma$ and $\kappa(t) = a$.

The Libor market model The Libor market model has been introduced by Brace *et al.* (1997) and is also known as the BGM model in reference to the names of Brace, Gatarek and Musiela. We recall that the Libor rate is defined as a simple forward rate:

$$L(t, T_i, T_{i+1}) = \frac{1}{T_{i+1} - T_i}\left(\frac{B(t, T_i)}{B(t, T_{i+1})} - 1\right)$$

In order to simplify the notation, we write $L_i(t) = L(t, T_i, T_{i+1})$. Under the forward probability measure $\mathbb{Q}^\star(T_{i+1})$, the Libor rate $L_i(t)$ is a martingale:

$$\mathrm{d}L_i(t) = \gamma_i(t) L_i(t) \, \mathrm{d}W_i^{\mathbb{Q}^\star(T_{i+1})}(t) \tag{9.22}$$

Then, we can use the Black formula (9.18) on page 521 to price caplets and floorlets where the volatility σ_i is defined by:

$$\sigma_i^2 = \frac{1}{T_i - t}\int_t^{T_i} \gamma_i^2(s) \, \mathrm{d}s$$

Therefore, we can price caps and floors because they are just a sum of caplets and floorlets.

Flat or spot implied volatility We can define two surfaces of implied volatilities. Since we observe the market prices of caps and floors, we can deduce the corresponding implied volatilities by assuming that the volatility in the Black model is constant. Thus, we have:

$$\begin{aligned}
\mathrm{Cap}_n(t) &= \mathrm{Cap}(t, T_0, T_1, \ldots, T_n) \\
&= \sum_{i=1}^{n} \mathrm{Caplet}(t, T_{i-1}, T_i) \\
&= \sum_{i=1}^{n} \mathrm{Caplet}_i(t)
\end{aligned}$$

where $\mathrm{Caplet}_i(t) = \boldsymbol{\mathcal{C}}(L_{i-1}(t), K, \sigma_{i-1}, T_i)$ and $\boldsymbol{\mathcal{C}}(L, K, \sigma, T)$ is the Black formula with volatility σ. The implied volatility $\Sigma(K, T)$ is then obtained by solving the following equation:

$$\sum_{i=1}^{n} \boldsymbol{\mathcal{C}}(L_{i-1}(t), K, \Sigma, T_i) = \mathrm{Cap}_n(t)$$

The implied volatility is also called the '*flat*' volatility and is denoted by $\Sigma^{\mathrm{flat}}(K, T_n)$. In this case, there is a flat implied volatility for each strike K and each maturity T_n of caps/floors. However, we can also compute an implied volatility $\Sigma(K, T)$ for each caplet. We have:

$$\begin{aligned}
\mathrm{Cap}_n(t) &= \mathrm{Cap}(t, T_0, T_1, \ldots, T_n) \\
&= \sum_{i=1}^{n} \mathrm{Caplet}(t, T_{i-1}, T_i) \\
&= \sum_{i=1}^{n} \boldsymbol{\mathcal{C}}(L_{i-1}(t), K, \Sigma(K, T_{i-1}), T_i)
\end{aligned}$$

The estimation of the implied volatility surface is obtained by minimizing the sum of squared residuals between observed and theoretical prices. In this case, the implied volatility is called the '*spot*' volatility and is denoted by $\Sigma^{\mathrm{spot}}(K, T_{i-1})$.

Example 86 *We consider* 6 *caplets on the 3M Libor rate, whose strike is equal to* 3%*. The tenor structures are respectively (3M,6M), (6M,9M), (9M,12M), (12M,15M), (15M,18M) and (18M,21M). In the following table, we indicate the price of the six caps*[29]*, whose notional is equal to \$1 m.*

Maturity of the cap	6M	9M	12M	15M	18M	21M
Cap price	151.50	529.74	1259.38	2221.82	3295.31	4594.40

We indicate below the current value of the forward Libor rate, and also the value of the zero-coupon rate.

Start date T_{i-1}	3M	6M	9M	12M	15M	18M
Maturity T_i	6M	9M	12M	15M	18M	21M
Forward Libor rate	3.05%	3.15%	3.30%	3.40%	3.45%	3.55%
Zero-coupon rate	3.05%	3.10%	3.15%	3.20%	3.25%	3.30%

Given the term structure of the volatility, we can price the caplets and the caps[30]. Since we have the price of the caps, we can calibrate the flat and spot implied volatilities. We obtain the results given in Table 9.5.

TABLE 9.5: Calibration of $\Sigma^{\text{flat}}(K,T_n)$, $\Sigma^{\text{spot}}(K,T_i)$ and γ_i

T_n	$\Sigma^{\text{flat}}(K,T_n)$	T_i	$\Sigma^{\text{spot}}(K,T_i)$	T_i	γ_i
6M	5.000%	3M	5.000%	3M	5.000%
9M	5.083%	6M	5.199%	6M	5.391%
12M	5.130%	9M	5.449%	9M	5.918%
15M	5.158%	12M	5.497%	12M	5.637%
18M	5.192%	15M	5.557%	15M	5.794%
21M	5.214%	18M	5.616%	18M	5.899%

We consider that the functions $\gamma_i(t)$ are the same and are equal to $\gamma(t)$. If we assume that $\gamma(t)$ is a piecewise constant function, we have:

$$\gamma(t) = \begin{cases} \gamma_0 & \text{if } t \in [0, T_0[\\ \gamma_i & \text{if } t \in [T_{i-1}, T_i[\end{cases}$$

It follows that:

$$\int_0^{T_i} \gamma^2(s)\,\mathrm{d}s = \int_0^{T_i-1} \gamma^2(s)\,\mathrm{d}s + \int_{T_i-1}^{T_i} \gamma^2(s)\,\mathrm{d}s$$

or:

$$T_i \Sigma^{\text{spot}}(K,T_i)^2 = T_{i-1}\Sigma^{\text{spot}}(K,T_{i-1})^2 + (T_i - T_{i-1})\,\gamma_i^2$$

We deduce that:

$$\gamma_0 = \Sigma^{\text{spot}}(K,T_0)$$

[29] The i^{th} cap is the sum of the first i caplets.

[30] For instance, if we assume that the volatility σ_i for the second caplet is 5%, we obtain:

$$\text{Caplet}(0,6\text{M},9\text{M}) = 10^6 \times 0.25 \times e^{-0.75\times 3.05\%} \times (3.15\% \times \Phi(d_1) - 3\% \times \Phi(d_2)) = \$394.48$$

where:

$$d_1 = \frac{1}{5\% \times \sqrt{0.5}} \ln\left(\frac{3.15\%}{3\%}\right) + \frac{1}{2} \times 5\% \times \sqrt{0.5} = 1.3977$$

and:

$$d_2 = d_1 - 5\% \times \sqrt{0.5} = 1.3623$$

and:

$$\gamma_i = \sqrt{\frac{T_i \Sigma^{\text{spot}}(K, T_i)^2 - T_{i-1} \Sigma^{\text{spot}}(K, T_{i-1})^2}{T_i - T_{i-1}}}$$

Therefore, we can use the spot volatilities to calibrate the function $\gamma(t)$ (see Table 9.5 and Figure 9.16).

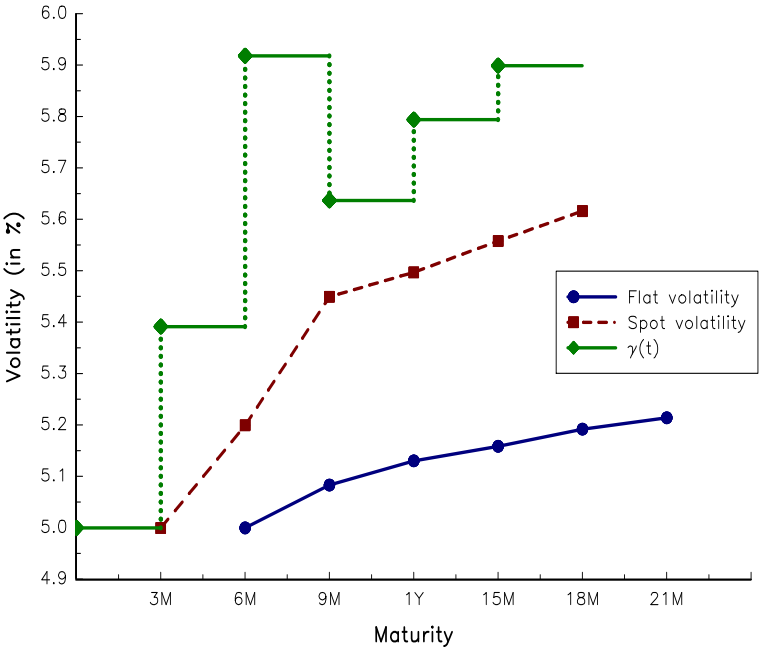


FIGURE 9.16: Flat and spot implied volatilities

Remark 102 *There is a lag between the flat volatility and the spot volatility, because we use the convention that the flat volatility is measured at the maturity date of the cap while the spot volatility is measured at the fixing date. In the previous example, the first flat volatility corresponds to the 6-month maturity date of the cap, whereas the first spot volatility corresponds to the 3-month fixing date of the caplet.*

Dynamics under other probability measures The dynamics (9.22) is valid for the Libor forward rate $L(t, T_i, T_{i+1})$. Then, we have:

$$\begin{cases} \mathrm{d}L_0(t) = \gamma_0(t) L_0(t) \, \mathrm{d}W_0^{\mathbb{Q}^\star(T_1)}(t) \\ \vdots \\ \mathrm{d}L_{n-1}(t) = \gamma_{n-1}(t) L_{n-1}(t) \, \mathrm{d}W_{n-1}^{\mathbb{Q}^\star(T_n)}(t) \end{cases}$$

It is obvious that the Wiener processes $(W_0, \ldots, W_{n-1})$ are correlated. We can show that the dynamics of $L_i(t)$ under the probability measure $\mathbb{Q}^\star(T_{k+1})$ is equal to:

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \mu_{i,k}(t) \, \mathrm{d}t + \gamma_i(t) \, \mathrm{d}W_k^{\mathbb{Q}^\star(T_{k+1})}(t)$$

where[31]:

$$\mu_{i,k}(t) = -\gamma_i(t) \sum_{j=i+1}^{k} \rho_{i,j}\gamma_j(t) \frac{(T_{j+1} - T_j) L_j(t)}{1 + (T_{j+1} - T_j) L_j(t)} \qquad \text{if } k > i$$

and $\rho_{i,j}$ is the correlation between $W_i^{\mathbb{Q}^\star(T_{i+1})}$ and $W_j^{\mathbb{Q}^\star(T_{j+1})}$.

Brigo and Mercurio (2006) derive the risk-neutral dynamics of the forward Libor rate $L_i(t)$ when we use the spot numéraire $M(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right)$. However, the expression is complicated and it is not very useful from a practical point of view. This is why they define another version of the spot numéraire, when the money market account is rebalanced only on the resetting dates $T_0, T_1, \ldots, T_{n-1}$. Let $\varphi(t)$ be the next resetting date index after time t, meaning that $\varphi(t) = i$ if $T_{i-1} < t < T_i$. The spot Libor numéraire is then defined as follows:

$$M^\dagger(t) = B\left(t, T_{\varphi(t)-1}\right) \prod_{j=0}^{\varphi(t)-1} (1 + \delta_j L_j(T_j))$$

and we have:

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \left(\gamma_i(t) \sum_{j=\varphi(t)}^{i} \rho_{i,j}\gamma_j(t) \frac{\delta_j L_j(t)}{1 + \delta_j L_j(t)}\right) \mathrm{d}t + \gamma_i(t)\,\mathrm{d}W_k^{\mathbb{Q}}(t)$$

where $W_k^{\mathbb{Q}}(t)$ is a Brownian motion when the numéraire is $M^\dagger(t)$.

The swap market model Since forward Libor rates $L_i(t)$ are log-normal distributed, the forward swap rate $\mathrm{Sw}(t)$ cannot be log-normal. Then, the Black formula cannot be applied to price swaptions[32]. However, we can always price swaptions using Monte Carlo methods by considering the spot measure (Glasserman, 2003). To circumvent this issue, Jamshidian (1997) proposed a model where the swap rate is a martingale under the annuity probability measure $\mathbb{Q}^\star$:

$$\mathrm{d}\,\mathrm{Sw}(t) = \eta(t)\,\mathrm{Sw}(t)\,\mathrm{d}W^{\mathbb{Q}^\star}(t)$$

Again, we can use the Black formula for pricing swaptions. However, we face the same problem as previously, because forward swap and Libor rates cannot be both log-normal.

9.2 Volatility risk

In the first section of this chapter, we have seen canonical models (Black-Scholes, Black, HJM and LMM) used to price options. In fact, they are not really '*option*' pricing models in the sense that European options such as calls, puts, caps, floors and swaptions are observed in the market. Indeed, they are more '*volatility*' pricing models, because they give a price to the implied volatility of European options. Knowing the implied volatility surface, the trader can then price exotic or OTC derivatives, and more importantly, define corresponding hedging portfolios.

[31] If $k < i$, we have:

$$\mu_{i,k}(t) = \gamma_i(t) \sum_{j=k+1}^{i} \rho_{i,j}\gamma_j(t) \frac{(T_{j+1} - T_j) L_j(t)}{1 + (T_{j+1} - T_j) L_j(t)}$$

[32] Nevertheless, there exist several approximations for pricing swaptions (Rebonato, 2002).

9.2.1 The uncertain volatility model

On page 512, we have seen that the P&L of the replicating strategy is given by the formula of El Karoui *et al.* (1998):

$$V(T) - f(S(T)) = \frac{1}{2}\int_0^T e^{r(T-t)}\mathbf{\Gamma}(t)\left(\Sigma^2(T,K) - \sigma^2(t)\right)S^2(t)\,\mathrm{d}t$$

If we assume that $\sigma(t) \in [\sigma^-, \sigma^+]$, we obtain a simple rule for achieving a positive P&L:

- if $\mathbf{\Gamma}(t) \geq 0$, we have to hedge the portfolio by considering an implied volatility that is equal to the upper bound σ^+;
- if $\mathbf{\Gamma}(t) \leq 0$, we set the implied volatility to the lower bound σ^-.

This rule is valid if the gamma of the option is always positive or negative, that is when the payoff is convex. Avellaneda *et al.* (1995) extend this rule when the gamma can change its sign during the life of the option. This is the case of many exotic options, which depend on conditional events (butterfly, barrier, call spread, ratchet, etc.).

9.2.1.1 Formulation of the partial differential equation

We assume that the dynamics of the underlying price is given by the following SDE:

$$\mathrm{d}S(t) = r(t)S(t)\,\mathrm{d}t + \sigma(t)S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) \tag{9.23}$$

where:

$$\sigma^- \leq \sigma(t) \leq \sigma^+ \tag{9.24}$$

Let $V(t, S(t))$ be the option price, whose payoff is $f(S(T))$. Avellaneda *et al.* (1995) show that $V(t, S(t))$ is bounded:

$$V^-(t, S(t)) \leq V(t, S(t)) \leq V^+(t, S(t))$$

where $V^-(t, S(t)) = \inf_{\mathbb{Q}(\sigma)} \mathbb{E}^{\mathbb{Q}(\sigma)}\left[\exp\left(-\int_t^T r(s)\,\mathrm{d}s\right) f(S(T))\right]$, $V^+(t, S(t)) = \sup_{\mathbb{Q}(\sigma)} \mathbb{E}^{\mathbb{Q}(\sigma)}\left[\exp\left(-\int_t^T r(s)\,\mathrm{d}s\right) f(S(T))\right]$ and $\mathbb{Q}(\sigma)$ denotes all the probability measures such that Equations (9.23) and (9.24) hold. We can then show that V^- and V^+ satisfy the HJB equation:

$$\begin{aligned}\sup_{\sigma^- \leq \sigma(t) \leq \sigma^+} \left(\frac{1}{2}\sigma^2(t)S^2\frac{\partial^2 V(t,S)}{\partial S^2} + b(t)S\frac{\partial V(t,S)}{\partial S}\right) + \\ \frac{\partial V(t,S)}{\partial t} - r(t)V(t,S) = 0\end{aligned}$$

Solving the HJB equation is equivalent to solve the modified Black-Scholes PDE:

$$\begin{cases} \frac{1}{2}\sigma^2(\mathbf{\Gamma}(t,S))S^2\partial_S^2 V(t,S) + b(t)S\partial_S V(t,S) + \partial_t V(t,S) - r(t)V(t,S) = 0 \\ V(T, S(T)) = f(S(T)) \end{cases}$$

where:

$$\sigma(x) = \begin{cases} \sigma^+ & \text{if } x \geq 0 \\ \sigma^- & \text{if } x < 0 \end{cases} \qquad \text{for } V(t,S(t)) = V^+(t,S(t))$$

and:

$$\sigma(x) = \begin{cases} \sigma^- & \text{if } x > 0 \\ \sigma^+ & \text{if } x \leq 0 \end{cases} \qquad \text{for } V(t,S(t)) = V^-(t,S(t))$$

Since $\mathbf{\Gamma}(t,S) = \partial_S^2 V(t,S)$ may change its sign during the time interval $[t,T]$, we have to solve the PDE numerically. A solution consists in using finite difference methods described in Appendix A.1.2.4 on page 1041.

Let u_i^m be the numerical solution of $V(t_m, S_i)$. At each iteration m, we approximate the gamma coefficient by the central difference method:

$$\mathbf{\Gamma}(t_m, S_i) \simeq \frac{u_{i+1}^m - 2u_i^m + u_{i+1}^m}{h^2}$$

By assuming that:

$$\operatorname{sign}(\mathbf{\Gamma}(t_m, S_i)) \approx \operatorname{sign}(\mathbf{\Gamma}(t_{m+1}, S_i))$$

we can compute the values taken by $\sigma(\mathbf{\Gamma}(t,S))$ and solve the PDE for the next iteration $m+1$.

9.2.1.2 Computing lower and upper pricing bounds

If we consider the European call option, we have $\mathbf{\Gamma}(t,S) > 0$, meaning that:

$$V^+(t, S(t)) = C_{\text{BS}}\left(t, S(t), \sigma^+\right)$$

and:

$$V^-(t, S(t)) = C_{\text{BS}}\left(t, S(t), \sigma^-\right)$$

where $C_{\text{BS}}(t,S,\sigma)$ is the Black-Scholes price at time t when the underlying price is equal to S and the implied volatility is equal to Σ. Then, the worst-case scenario occurs when the volatility $\sigma(t)$ reaches the upper bound σ^+.

This result is obtained because the delta of the option is a monotone function with respect to the underlying price. However, this property does not hold for many derivative contracts, in particular when the payoff is path dependent. In this case, the payoff depends on the trajectory of the underlying asset. For instance, the payoff of a barrier option depends on whether a certain barrier level was touched (or not touched) at some time during the life of the option. We give here the payoff associated to the four main types of single barrier[33]:

- down-and-in call and put options (DIC/DIP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\left\{S_0 > L, \min_{t\in\mathcal{T}} S(t) \leq L\right\} \cdot f_{\text{Vanilla}}(S(T))$$

- down-and-out call and put option (DOC/DOP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\left\{S_0 > L, \min_{t\in\mathcal{T}} S(t) > L\right\} \cdot f_{\text{Vanilla}}(S(T))$$

- up-and-in call and put options (UIC/UIP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\left\{S_0 < H, \max_{t\in\mathcal{T}} S(t) \geq H\right\} \cdot f_{\text{Vanilla}}(S(T))$$

[33] We have:

$$f_{\text{Vanilla}}(S(T)) = \begin{cases} (S(T)-K)^+ & \text{for the call option} \\ (K-S(T))^+ & \text{for the put option} \end{cases}$$

- up-and-out call and put options (UOC/UOP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\left\{S_0 < H, \max_{t\in\mathcal{T}} S(t) < H\right\} \cdot f_{\text{Vanilla}}(S(T))$$

In the case of knocked-out barrier payoffs (DOC/DOP, UOC/UOP), the option terminates the first time the barrier is crossed, whereas knocked-in barrier options (DIC/DIP, UIC/UIP), the payoff is paid only if the underlying asset crosses the barrier. These barriers can also be combined in order to obtain double barrier options:

- double knocked-in call and put options (KIC/KIP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\{S(t) \notin [L, H], t \in \mathcal{T}\} \cdot f_{\text{Vanilla}}(S(T))$$

- double knocked-out call and put option (KOC/KOP):

$$f_{\text{Barrier}}(S(T)) = \mathbb{1}\{S(t) \in [L, H], t \in \mathcal{T}\} \cdot f_{\text{Vanilla}}(S(T))$$

These options also depend on the time monitoring $t \in \mathcal{T}$ of the barriers. In particular, we distinguish continuous ($\mathcal{T} = [0, T]$), window ($\mathcal{T} \subset [0, T]$) and discrete ($\mathcal{T} = \{t_1, t_2, \ldots, t_n\}$) barriers.

Example 87 *We consider a double KOC barrier option with the following parameters:* $K = 100$, $L = 80$, $H = 120$, $T = 1$, $b = 5\%$ *and* $r = 5\%$. *We assume that the volatility* $\sigma(t)$ *lies in the range of* 15% *and* 25%.

In the first and second panels of Figure 9.17, we report the price $V(T, S)$ of the call option for the continuous barrier ($\mathcal{T} = [0, 1]$). If we use the Black-Scholes model[34], the upper bound is reached when $\sigma(t) = \sigma^- = 15\%$ whereas the lower bound is reached when $\sigma(t) = \sigma^+ = 25\%$. We have the feeling that the barrier price is a decreasing function of the volatility. However, this is not true. Indeed, a high volatility increases the time value of the final payoff $(S(T) - K)^+$, but also decreases the probability to remain within the barrier interval $[L, H]$. Therefore, there is a trade-off between these two opposite effects. If we consider the uncertain volatility model (UVM), the upper bound is larger than this obtained with the BS model, because the worst-case scenario is to have a low volatility when the asset price is close to one barrier and a high volatility when the asset price is far way from the barriers. Therefore, the worst-case scenario at time t depends on the relative position of $S(t)$ with respect to L, H and K. If we consider a window barrier with $\mathcal{T} = [0.25, 0.75]$, we obtain the third and fourth panels of Figure 9.17. We notice that the BS price is not monotone with respect to the volatility. When the current asset price S_0 is equal to the strike K, the BS price is higher when $\sigma(t) = \sigma^- = 15\%$. This is not the case when $S_0 = 150$. The reason is that a high volatility increases the probability than the asset price is below the up barrier H when the window is triggered. A high volatility is also good when the window ends.

9.2.1.3 Application to ratchet options

Ratchet or cliquet options are financial derivatives that provide a minimum return in exchange for capping the maximum return. They are used by investors because they may

[34] Prices can be computed by numerically solving the PDE, or using the closed-form formulas of Rubinstein and Reiner (1991).

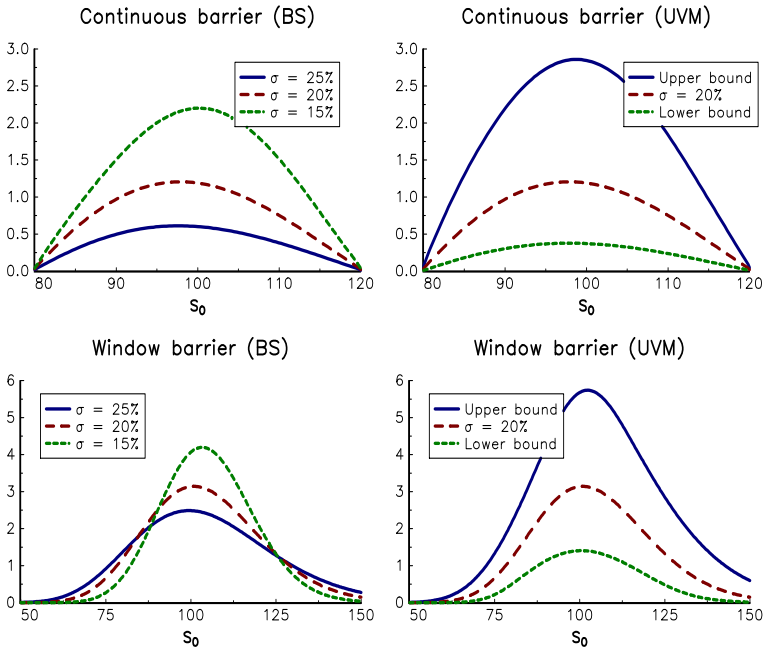


FIGURE 9.17: Comparing BS and UVM prices of the double KOC barrier option

protect them against downside risk. Let us see an example to understand the underlying mechanism of such derivative contracts.

We consider a cliquet option with a 3-year maturity on an equity index $S(t)$. The fixing dates corresponds to the end of each calendar year. We assume that the initial value S_0 of the index is equal to 100. The payoff of the cliquet option is:

$$f(S(T)) = N \cdot \left(\sum_{j=1}^{3} \max \left(0, \frac{S(T_j) - S(T_{j-1})}{S(T_{j-1})} \right) \right)$$

where $\{T_1, T_2, T_3\}$ are the fixing dates and N is the notional of the cliquet option. This cliquet option accumulates positive annual returns. In the following table, we have report four trajectories of $S(T_j)$:

$S(T_j)$	#1	#2	#3	#4
$S(0)$	100	100	100	100
$S(1)$	120	110	95	90
$S(2)$	85	125	95	50
$S(3)$	90	135	75	70
Coupon	25.9%	31.6%	0%	40%

More generally, the payoff of a ratchet is:

$$f(S(T)) = N \cdot \min \left(C_g, \max \left(F_g, \sum_{j=1}^{n} \max(F_\ell, \min(C_\ell, R_j - K_\ell)) - K_g \right) \right)$$

where C_g is the global cap, F_g is the global floor, K_g is the global strike, C_ℓ is the local cap, F_ℓ is the local floor and K_ℓ is the local strike. Here, R_j is the return between two fixing

dates:

$$R_j = \frac{S(T_j) - S(T_{j-1})}{S(T_{j-1})}$$

At the maturity, the buyer of the cliquet option receives the sum of periodic returns subject to local and global caps, floors and strikes. In the market, one of the most common payoffs is the following:

$$f(S(T)) = N \cdot \max\left(F_g, \sum_{j=1}^{n} \max\left(0, \min\left(C_\ell, \frac{S(T_j)}{S(T_{j-1})} - 1\right)\right)\right)$$

With this payoff, the option buyer is hedged against the fall of the asset price and has the guarantee to have a minimum return that is equal to the global floor F_g. On the contrary, the option buyer limits the upside risk by introducing the local cap C_ℓ. Therefore, the price of the option is bounded:

$$e^{-rT} \cdot F_g \leq f(S(T)) \leq e^{-rT} \cdot \max(F_g, nC_\ell)$$

The fundamental issue of cliquet option pricing is the choice of the volatility model to price the forward call option:

$$\mathbb{E}\left[\left(\frac{S(T_j)}{S(T_{j-1})} - 1\right)^+ \middle| \mathcal{F}_0\right]$$

At first sight, we might consider the following solutions:

- we may use the implied forward volatility between T_{j-1} and T_j, which is calculated as follows:

$$T_j \cdot \Sigma^2(T_j) = T_{j-1} \cdot \Sigma^2(T_{j-1}) + (T_j - T_{j-1}) \cdot \Sigma^2(T_{j-1}, T_j)$$

- we may also use the implied volatility of maturity $T_j - T_{j-1}$ at the date T_{j-1}; this implies to have a dynamic model of the implied volatility surface.

Since the payoff is locally non-convex, it is not possible to calculate a conservative price using the Black-Scholes model. In this case, the choice of a good implied volatility is inappropriate.

Wilmott (2002) illustrates the difficulty of pricing cliquet options by comparing Black-Scholes and uncertain volatility models. The BS price can be calculated using the Monte Carlo method[35]. Another solution is to derive the corresponding PDE. In this case, we have to introduce two additional variables: $S' = S(T_{j-1})$ is the value of $S(t)$ at the previous fixing date and Q is a variable to keep track of the payoff:

$$Q = \sum_{j=1}^{n} \max\left(0, \min\left(C_\ell, \frac{S(T_j)}{S(Tt_{j-1})} - 1\right)\right)$$

The value of the option depends then on four state variables:

$$V = V(t, S, S', Q)$$

[35] For that, we simulate the asset price at the fixing dates $\{0, T_0, \ldots, T_n, T\}$ using the risk-neutral probability measure $\mathbb{Q}$ and we calculate the mean of the discounted payoff.

We deduce that $V(t,S,S',Q)$ satisfies the following PDE between two fixing dates T_{j-1} and T_j:

$$\frac{1}{2}\sigma^2 S^2 \partial_S^2 V(\cdot) + b(t) S \partial_S V(\cdot) + \partial_t V(\cdot) - r(t) V(\cdot) = 0$$

whereas the final condition is:

$$V(T,S,S',Q) = N \cdot \max(F_g, Q)$$

As noted by Wilmott (2002), $V(t,S,S',Q)$ must also satisfy the jump condition at the fixing date T_j:

$$V(T_j,S,S',Q) = V\left(T_j^+, S, S, Q + \max\left(0, \min\left(C_\ell, \frac{S}{S'} - 1\right)\right)\right)$$

This jump condition initializes the new value of S' for the next period $[T_{j-1}, T_j]$ and update the payoff Q. By introducing the state variable $x = S/S'$, Wilmott reduces the dimension of the problem to three variables t, x and Q:

$$\begin{cases} \frac{1}{2}\sigma^2 x^2 \partial_x^2 V(t,x,Q) + b(t) x \partial_x V(t,x,Q) + \partial_t V(t,x,Q) - r(t) V(t,x,Q) = 0 \\ V(T_j,x,Q) = V\left(T_j^+, 1, Q + \max(0, \min(C_\ell, x-1))\right) \\ V(T,x,Q) = N \cdot \max(F_g, Q) \end{cases}$$

This PDE can easily be solved numerically and the price of the cliquet option is equal to $V(0,1,0)$. For the uncertain volatility model, we have exactly the same PDE, except that the quadratic term is replaced by $\frac{1}{2}\sigma^2(\mathbf{\Gamma}(t,x)) x^2 \partial_x^2 V(t,x,Q)$.

Example 88 *We consider a cliquet option with the following parameters:* $r = 5\%$, $b = 5\%$, $F_g = 10\%$, $C_\ell = 12\%$ *and* $N = 1$. *The maturity is equal to 5 years, and there are 5 annual fixing dates. The volatility* $\sigma(t)$ *lies in the range 20% to 30%.*

In Figure 9.18, we show the PDE solution $V(0,x,0)$ for constant volatility and volatility ranges. We notice that the BS price is not very sensitive to the volatility. With respect to the mid volatility $\sigma = 25\%$, the BS price increases by 1.35% if the volatility is 30% and decreases by 1.57% if the volatility is 20%. On the contrary, the UVM price range $(V^+ - V^-)$ represents 34% of the BS price. This result depends on the values of the global floor and the local cap. An illustration is provided in Figure 9.19, which gives the relationship[36] between the cliquet option price $V(0,1,0)$ and the local cap C_ℓ.

9.2.2 The shifted log-normal model

This model assumes that the asset price $S(t)$ is a linear transformation of a log-normal random variable $X(t)$:

$$S(t) = \alpha(t) + \beta(t) X(t)$$

where $\beta(t) \geq 0$. Then, the payoff of the European call option is:

$$\begin{aligned} f(S(T)) &= (S(T) - K)^+ \\ &= (\alpha(T) + \beta(T) X(T) - K)^+ \\ &= \beta(T)\left(X(T) - \frac{K - \alpha(T)}{\beta(T)}\right)^+ \end{aligned}$$

[36] The parameters are those given in Example 88.

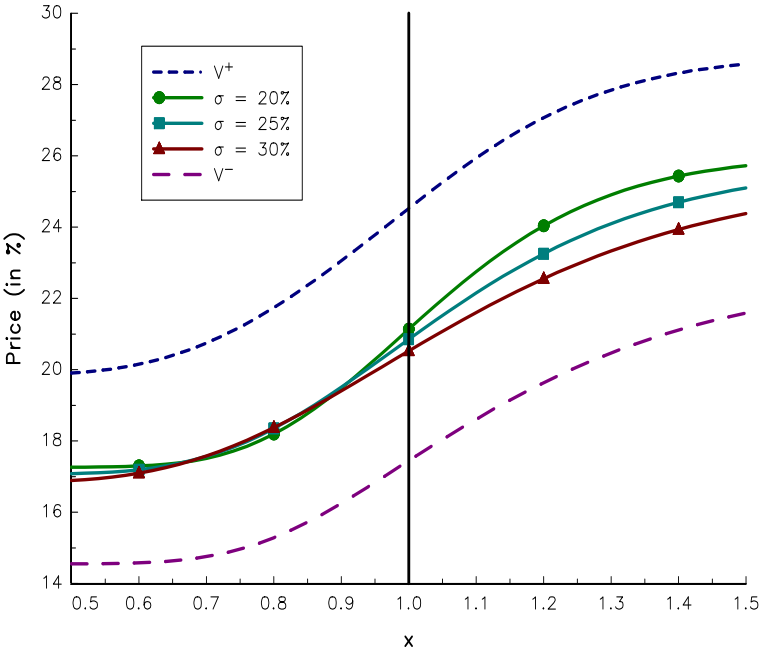


FIGURE 9.18: Comparing BS and UVM prices of the cliquet option

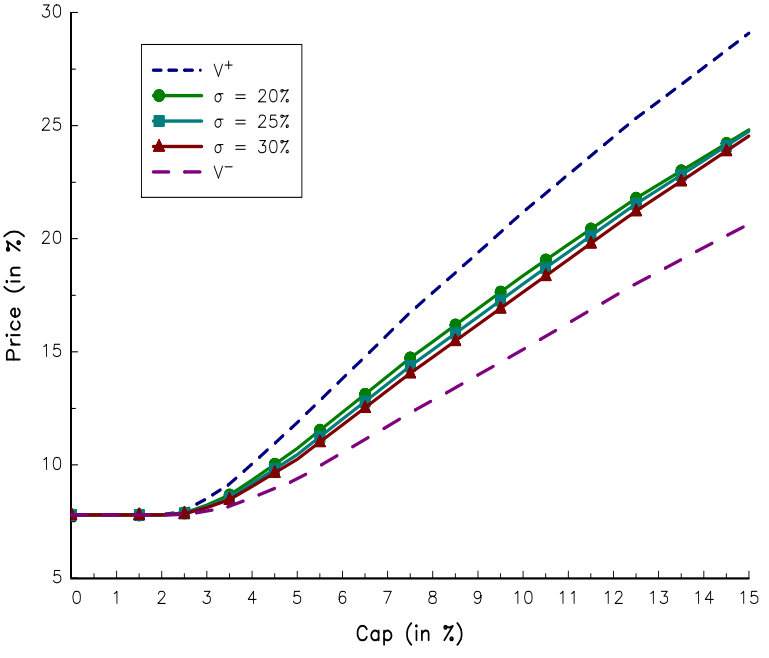


FIGURE 9.19: Influence of the local cap on the cliquet option price

This type of approach is interesting because the pricing of options can then be done using the Black-Scholes formula:

$$\mathcal{C}(0, S_0) = \beta(T) C_{\text{BS}}\left(X_0, \frac{K-\alpha(T)}{\beta(T)}, \sigma_X, T, b_X, r\right)$$

where b_X and σ_X are the drift and diffusion coefficients of $X(t)$ under the risk-neutral probability measure $\mathbb{Q}$. This modeling framework has been introduced by Rubinstein (1983) and popularized by Damiano Brigo and Fabio Mercurio in a series of working papers written between 2000 and 2003[37]. This model was originally used in order to generate a volatility skew, but it is now extensively used in interest rate derivatives because it extends the Black model when facing negative interest rates.

9.2.2.1 The fixed-strike parametrization

Let us suppose that:

$$S(t) = \alpha + \beta \exp\left(\left(b^{\mathbb{Q}}(t) - \frac{1}{2}\sigma^2\right)t + \sigma W^{\mathbb{Q}}(t)\right)$$

We have $S_0 = \alpha + \beta$ meaning that:

$$S(t) = \alpha + (S_0 - \alpha) \exp\left(\left(b^{\mathbb{Q}}(t) - \frac{1}{2}\sigma^2\right)t + \sigma W^{\mathbb{Q}}(t)\right) \tag{9.25}$$

Let b the cost-of-carry parameter of the asset. Under the risk-neutral probability measure, the martingale condition is:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-bt} S(t) \mid \mathcal{F}_0\right] = S_0$$

Since we have $\mathbb{E}^{\mathbb{Q}}[S(t)] = \alpha + (S_0 - \alpha) e^{b^{\mathbb{Q}}(t)t}$, we deduce that the no-arbitrage condition implies that:

$$\alpha + (S_0 - \alpha) e^{b^{\mathbb{Q}}(t)t} = S_0 e^{bt}$$

or:

$$b^{\mathbb{Q}}(t) = \frac{1}{t} \ln\left(\frac{S_0 e^{bt} - \alpha}{S_0 - \alpha}\right)$$

The payoff of the European call option is:

$$\begin{aligned} f(S(T)) &= (S(T) - K)^+ \\ &= ((S(T) - \alpha) - (K - \alpha))^+ \end{aligned}$$

We deduce that the price of the option is given by:

$$\mathcal{C}(0, S_0) = C_{\text{BS}}\left(S_0 - \alpha, K - \alpha, \sigma, T, b^{\mathbb{Q}}(T), r\right) \tag{9.26}$$

In Figure 9.20, we report the volatility skew generated by the SLN model when the current price S_0 of the asset is 100, the maturity T is one year, the cost-of-carry b is 5% and the interest rate 5 is 5%. We notice that the parameter σ of the SLN model is not of the same magnitude than the implied volatility of the BS model. This is due to the shift α. When α is positive (or negative), we have $\sigma > \Sigma(T, K)$ (or $\sigma < \Sigma(T, K)$).

[37] See Brigo and Mercurio (2002a) for a survey of their different works.

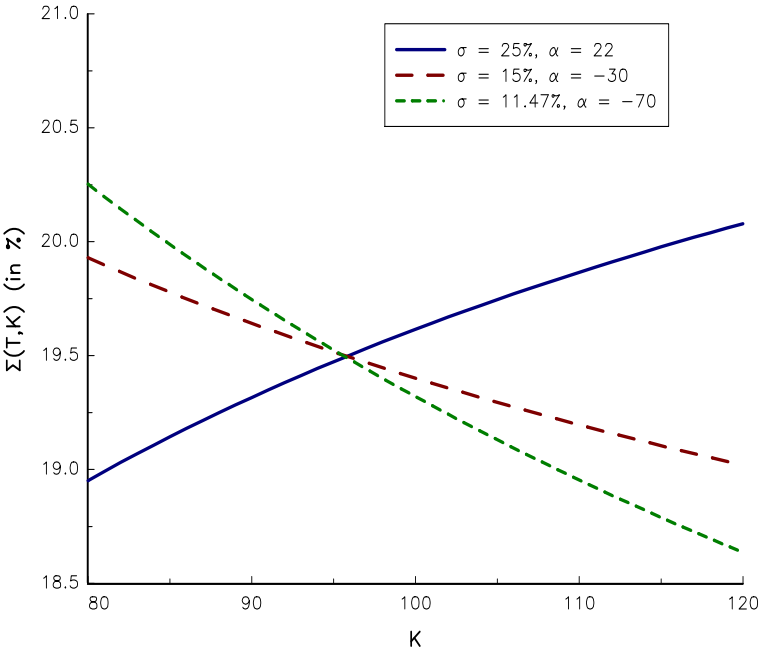


FIGURE 9.20: Volatility skew generated by the SLN model (fixed-strike parametrization)

9.2.2.2 The floating-strike parametrization

Let us now suppose that:

$$S(t) = \alpha e^{\varphi t} + \beta e^{\left(b-\frac{1}{2}\sigma^2\right)t+\sigma W^{\mathbb{Q}}(t)}$$

We have $S_0 = \alpha + \beta$ and $\mathbb{E}^{\mathbb{Q}}[S(t)] = \alpha e^{\varphi t} + \beta e^{bt}$. We deduce that the stochastic process $e^{-bt}S(t)$ is a $\mathcal{F}_t$-martingale if it is equal to:

$$S(t) = \alpha e^{bt} + (S_0 - \alpha)\, e^{\left(b-\frac{1}{2}\sigma^2\right)t+\sigma W^{\mathbb{Q}}(t)} \tag{9.27}$$

The payoff of the European call option becomes:

$$\begin{aligned} f(S(T)) &= (S(T) - K)^+ \\ &= \left(\left(S(T) - \alpha e^{bT}\right) - \left(K - \alpha e^{bT}\right)\right)^+ \end{aligned}$$

It follows that the option price is equal to:

$$\mathcal{C}(0, S_0) = C_{\text{BS}}\left(S_0 - \alpha, K - \alpha e^{bT}, \sigma, T, b, r\right) \tag{9.28}$$

Examples of Volatility skew are given in Figure 9.21 with the same parameters than those we have used in Figure 9.20.

Remark 103 *At first sight, the floating-strike parametrization seems to be different than the fixed-strike parametrization. In practice, the parameters* (α, σ) *are calibrated for each maturity* T*. This explains that the two parametrizations are very close.*

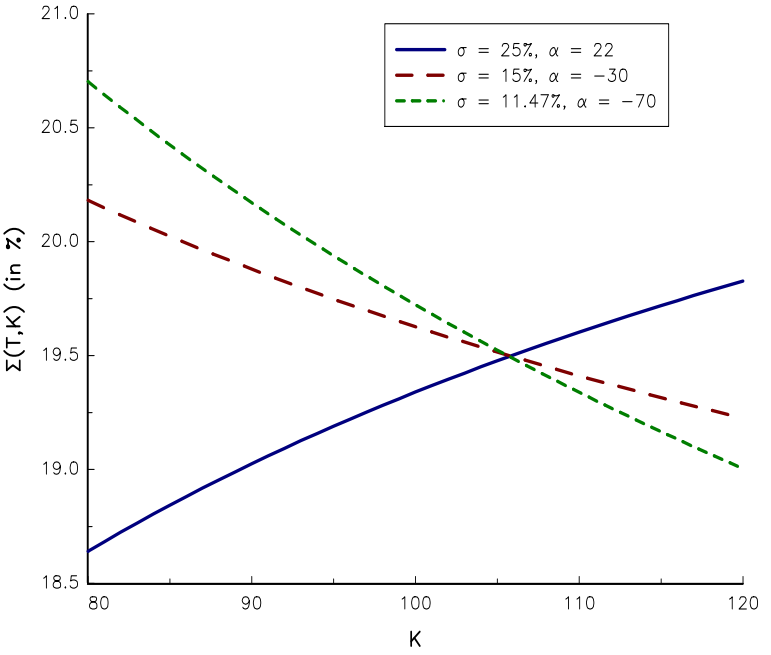


FIGURE 9.21: Volatility skew generated by the SLN model (floating-strike parametrization)

9.2.2.3 The forward parametrization

If we consider the forward price $F(t)$ instead of the spot price $S(t)$, the two models coincide because we have $b = 0$. In this case, the dynamics of the forward price is:

$$\mathrm{d}F(t) = \sigma(F(t) - \alpha)\,\mathrm{d}W^{\mathbb{Q}}(t) \tag{9.29}$$

and the price of the option is given by the Black formula[38]:

$$\boldsymbol{\mathcal{C}}(0, S_0) = C_{\text{Black}}(F_0 - \alpha, K - \alpha, \sigma, T, r) \tag{9.30}$$

In Equations (9.29) and (9.30), we impose that $\alpha < F_0$ and $\alpha < K$. This implies that $F(t) \in [\alpha, \infty)$. This model is appealing for fixed income derivatives, because the interest rate may be negative when α is negative. In this case, we have:

$$\begin{aligned}\mathrm{d}F(t) &= (\sigma F(t) - \alpha\sigma)\,\mathrm{d}W^{\mathbb{Q}}(t)\\ &= (\sigma_1 F(t) + \sigma_2)\,\mathrm{d}W^{\mathbb{Q}}(t)\end{aligned}$$

where $\sigma_1 = \sigma$ and $\sigma_2 = -\alpha\sigma > 0$. We obtain a stochastic differential equation whose diffusion coefficient is a mix of log-normal and Gaussian volatilities.

[38] We recall that the Black formula can be viewed as a special case of the Black-Scholes formula when the cost-of-carry parameter b is equal to zero:

$$C_{\text{Black}}(x, K, \sigma, T, r) = C_{\text{BS}}(x, K, \sigma, T, 0, r)$$

Lee and Wang (2012) prove the following results:

- monotonicity in strike:

$$\operatorname{sign}\left(\frac{\partial \Sigma(T,K)}{\partial K}\right) = \operatorname{sign}\alpha$$

- upper and lower bounds:

$$\begin{cases} \Sigma(T,K) < \sigma & \text{if } \alpha > 0 \\ \Sigma(T,K) > \sigma & \text{if } \alpha < 0 \end{cases}$$

- sharpness of bound:

$$\lim_{K\to\infty} \Sigma(T,K) = \sigma$$

- short-expiry behavior:

$$\lim_{T\to 0} \Sigma(T,K) = \begin{cases} \dfrac{\sigma \ln(F_0/K)}{\ln\left((F_0-\alpha)/(K-\alpha)\right)} & \text{if } K \neq F_0 \\ \sigma\left(1-\alpha F_0^{-1}\right) & \text{if } K = F_0 \end{cases}$$

The implied volatility formula does not depend on the maturity T and is only valid when T is equal to zero. However, it is a good approximation for other maturities as shown in Table 9.6. We use the previous parameters and three different maturities (one-month, one-year and five-year).

TABLE 9.6: Error of the SLN implied volatility formula (in bps)

K	$(\alpha = 22, \sigma = 25\%)$			$(\alpha = -70, \sigma = 12\%)$		
	1M	1Y	5Y	1M	1Y	5Y
80	1.0	11.1	57.0	−0.9	−12.9	−66.0
90	0.7	10.6	54.1	−1.0	−11.9	−61.4
100	0.9	10.2	51.6	−1.1	−11.3	−57.3
110	1.0	9.7	49.6	−0.8	−10.8	−53.8
120	0.7	9.3	47.7	−0.6	−10.3	−51.3

9.2.2.4 Mixture of SLN distributions

One limitation of the SLN model is that it only produces a volatility skew, and not a volatility smile. In order to obtain a U-shaped curve, Brigo and Mercurio (2002b) suggest that the (risk-neutral) probability density function $f(x)$ of the asset price density is given by the mixture of known basic densities:

$$f(x) = \sum_{j=1}^{m} p_j f_j(x)$$

where f_j is the j^{th} basic density, $p_j > 0$ and $\sum_{j=1}^{m} p_j = 1$. Let $G(S(T))$ be the payoff of an European option. We have:

$$\begin{aligned} \mathcal{C}(0,S_0) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT} G(S(T)) \middle| \mathcal{F}_0\right] \\ &= \int e^{-rT} G(S(T)) f(x)\, \mathrm{d}x \end{aligned}$$

We deduce that:

$$\begin{aligned}\boldsymbol{\mathcal{C}}(0,S_0) &= \int e^{-rT} G(S(T)) \sum_{j=1}^{m} p_j f_j(x)\,\mathrm{d}x \\ &= \sum_{j=1}^{m} p_j \int e^{-rT} G(S(T))\, f_j(x)\,\mathrm{d}x \\ &= \sum_{j=1}^{m} p_j \mathbb{E}^{\mathbb{Q}_j}\left[e^{-rT} G(S(T)) \middle| \mathcal{F}_0\right]\end{aligned}$$

where $\mathbb{Q}_j$ is the j^{th} probability measure. It is then straightforward to price an European option using formulas of basic models. If we consider a mixture of two shifted log-normal models, the price of the European call option is equal to:

$$\begin{aligned}\boldsymbol{\mathcal{C}}(0,S_0) &= p \cdot C_{\text{SLN}}(S_0,K,\sigma_1,T,b,r,\alpha_1) + \\ &\quad (1-p)\cdot C_{\text{SLN}}(S_0,K,\sigma_2,T,b,r,\alpha_2)\end{aligned}$$

where C_{SLN} is the formula of the SLN model[39]. The model has five parameters: σ_1, σ_2, α_1, α_2 and p.

Example 89 *We consider a calibration set of five options, whose strike and implied volatilities are equal to:*

K_j	80	90	100	110	120
$\Sigma(1,K_j)$	21%	19%	18.25%	18.5%	19%

The current value of the asset price is equal to 100*, the maturity of options is one year, the cost-of-carry parameter is set to* 0 *and the interest rate is* 5%*.*

The parameters are estimated by minimizing the weighted least squares:

$$\min \sum_{j=1}^{n} w_j \left(\hat{C}_j - C_{\text{SLN}}(S_0,K_j,\sigma_1,\sigma_2,T_j,b,r,\alpha_1,\alpha_2,p)\right)^2$$

where:

$$\hat{C}_j = C_{\text{BS}}(S_0,K_j,\Sigma(T_j,K_j),T_j,b,r)$$

and w_j is the weight of the j^{th} option. We consider three parameterizations: (#1) the weights w_j are uniform, and we impose that $\alpha_1 = \alpha_2$ and $p = 50\%$; (#2) the weights w_j are uniform, and p is set to 25%; (#3) the weights w_j are inversely proportional to option prices $\hat{C}_j$, and p is set to 50%. Results are given in Table 9.7 and Figure 9.22. We notice that α_1 and α_2 can take large values. Shifted log-normal models are generally presented as a low perturbation of the Black-Scholes model. In practice, they are very different.

9.2.2.5 Application to binary, corridor and barrier options

One of the difficulties when using the Black-Scholes model with exotic options is the choice of the implied volatility. In the case of an European call option, it is obvious to use the implied volatility $\Sigma(T,K)$ that corresponds to the strike and the maturity of the option. In the case of a double barrier option, we can use the implied volatility $\Sigma(T,K)$

[39] It corresponds to one of the three expressions (9.26), (9.28), and (9.30).

TABLE 9.7: Calibrated parameters of the mixed SLN model

Model	#1	#2	#3
σ_1	16.5%	8.2%	10.2%
σ_2	7.3%	17.2%	21.7%
α_1	−53.3	−289.7	−145.2
α_1	−53.3	19.6	47.4
p	50.0%	25.0%	50.0%

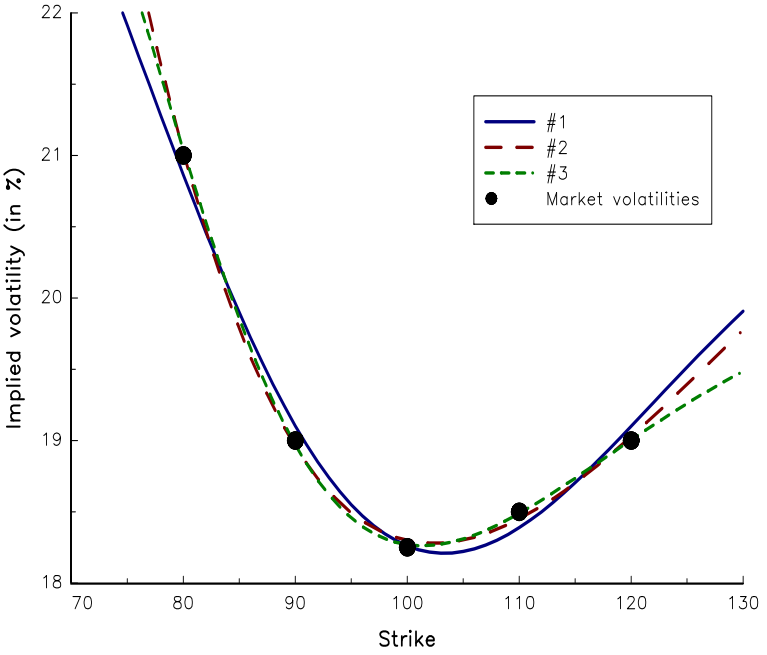


FIGURE 9.22: Implied volatility (in %) of calibrated mixed SLN models

that corresponds to the strike of the option, the implied volatility $\Sigma(T, L)$ that corresponds to the lower barrier of the option, the implied volatility $\Sigma(T, H)$ that corresponds to the higher barrier of the option, or another implied volatility. In fact, there is no satisfactory answer.

Let $S(t)$ be the asset price at time t. The payoff of the binary cash-or-nothing call option is:

$$f(S(T)) = \mathbb{1}\{S(T) > K\}$$

We deduce that:

$$\mathbf{BCC}(0, S_0) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(s)\,\mathrm{d}s} \cdot \mathbb{1}\{S(T) > K\} \middle| \mathcal{F}_0\right]$$

If we consider the Black-Scholes model, we obtain:

$$\mathbf{BCC}(0, S_0) = e^{-rT}\Phi(d_2)$$

We can replicate this option by using the classical dynamic delta hedging approach presented on page 495. Here, we consider another framework, which is called the static hedging method. The hedging portfolio consists in:

- a long position on the European call option with strike K;
- a short position on the European call option with strike $K+\varepsilon$.

If the notional of each option is set to ε, the value of the hedging portfolio at time t is equal to:

$$V(t) = \frac{1}{\varepsilon} \cdot \boldsymbol{\mathcal{C}}(t, S(t), K) - \frac{1}{\varepsilon} \cdot \boldsymbol{\mathcal{C}}(t, S(t), K+\varepsilon)$$

It follows that the value of the hedging strategy is equal to:

$$X(t) = \mathbf{BCC}(t, S(t)) - V(t)$$

We notice that:

$$\begin{aligned}
\lim_{\varepsilon \to 0} X(T) &= \mathbf{BCC}(T, S(T)) - \lim_{\varepsilon \to 0} V(T) \\
&= \mathbb{1}\{S(T) > K\} - \lim_{\varepsilon \to 0} \frac{(S(t)-K)^+ - (S(T)-K-\varepsilon)^+}{\varepsilon} \\
&= \mathbb{1}\{S(T) > K\} - \mathbb{1}\{S(T) > K\} \\
&= 0
\end{aligned}$$

The no-arbitrage condition implies that:

$$\begin{aligned}
\mathbf{BCC}(t, S(t)) &= \lim_{\varepsilon \to 0} \frac{\boldsymbol{\mathcal{C}}(t, S(t), K) - \boldsymbol{\mathcal{C}}(t, S(t), K+\varepsilon)}{\varepsilon} \\
&= -\lim_{\varepsilon \to 0} \frac{\boldsymbol{\mathcal{C}}(t, S(t), K+\varepsilon) - \boldsymbol{\mathcal{C}}(t, S(t), K)}{\varepsilon} \\
&= -\frac{\partial \boldsymbol{\mathcal{C}}(t, S(t), K)}{\partial K}
\end{aligned}$$

This result is valid only if the volatility is constant. If the volatility is not constant, the price $\mathbf{BCC}(t, S(t))$ becomes:

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \frac{\boldsymbol{\mathcal{C}}(t, S(t), K, \Sigma(T,K)) - \boldsymbol{\mathcal{C}}(t, S(t), K+\varepsilon, \Sigma(T,K+\varepsilon))}{\varepsilon} \\
= &-\frac{\partial \boldsymbol{\mathcal{C}}(t, S(t), K, \Sigma(T,K))}{\partial K} - \frac{\partial \boldsymbol{\mathcal{C}}(t, S(t), K, \Sigma(T,K))}{\partial \Sigma} \cdot \frac{\partial \Sigma(T,K)}{\partial K} \\
= &\ \mathbf{BCC}_{\mathrm{BS}}(t, S(t), \Sigma(T,K)) - \boldsymbol{\upsilon}_{\mathrm{BS}}(t, S(t), \Sigma(T,K))\, \omega(T,K)
\end{aligned}$$

where $\mathbf{BCC}_{\mathrm{BS}}(t, S(t), \Sigma(T,K))$ is the Black-Scholes price with implied volatility $\Sigma(T,K)$, $\boldsymbol{\upsilon}_{\mathrm{BS}}(t, S(t), \Sigma(T,K))$ is the Black-Scholes vega for the European call option and $\omega(T,K)$ is the skew of the volatility surface:

$$\omega(T,K) = \frac{\partial \Sigma(T,K)}{\partial K}$$

This framework, called the skew-method (SM) model, shows that taking into account the volatility smile cannot be reduced to choosing the right implied volatility, because we have:

$$\mathbf{BCC}_{\mathrm{SM}}(t, S(t)) \neq \mathbf{BCC}_{\mathrm{BS}}(t, S(t), \Sigma(T,K))$$

Example 90 *We price a binary call option when the underlying asset price is* 100*, the maturity of the option is* 6 *months, and the parameters* b *and* r *are equal to* 5%*. The skew* $\omega(T,K)$ *of the implied volatility can take the values* 0*,* -20 *and* $+20$ *bps. We consider two cases for the implied volatility: (1)* $\Sigma(T,K)$ *is equal to* 20%*, (2)* $\Sigma(T,K)$ *is a linear function with respect to* K*:*

$$\Sigma(T,K) = \Sigma(T,S_0) + \omega(T,K) \cdot (K - S_0)$$

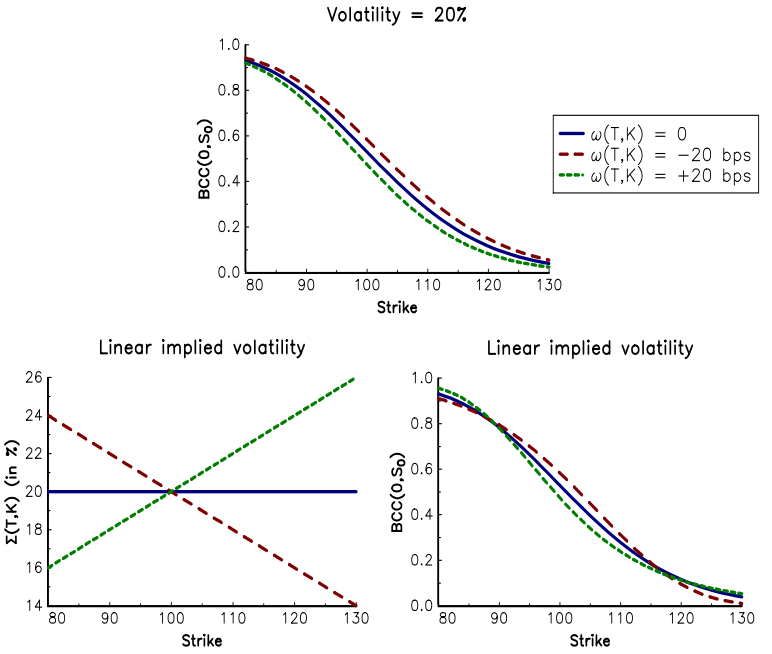


FIGURE 9.23: Impact of the implied volatility skew on the binary option price

Figure 9.23 represents the relationship between the binary call option price $\mathbf{BCC}(0, S_0)$ and the strike K. The first panel assumes that the implied volatility $\Sigma(T, K)$ is equal to 20%. We verify that:

$$\begin{cases} \omega(T,K) < 0 \Rightarrow \mathbf{BCC}_{\text{SM}}(0,S_0) < \mathbf{BCC}_{\text{BS}}(0,S_0,\Sigma(T,K)) \\ \omega(T,K) > 0 \Rightarrow \mathbf{BCC}_{\text{SM}}(0,S_0) > \mathbf{BCC}_{\text{BS}}(0,S_0,\Sigma(T,K)) \end{cases}$$

However, the results shown in the first panel may be misleading, because it is not possible to compare the price for two different strikes. Indeed, if $K_2 > K_1$ and $\omega(T,K) > 0$ for every strike K, this implies that $\Sigma(T,K_2) > \Sigma(T,K_1)$, $\mathbf{BCC}_{\text{BS}}(0,S_0,\Sigma(T,K_2)) > \mathbf{BCC}_{\text{BS}}(0,S_0,\Sigma(T,K_1))$, but $\boldsymbol{\upsilon}_{\text{BS}}(0,S_0,\Sigma(T,K_2)) > \boldsymbol{\upsilon}_{\text{BS}}(0,S_0,\Sigma(T,K_1))$. A higher implied volatility increases the binary option price thanks to the impact on the Black-Scholes price, but also reduces it thanks to the impact on the vega. Therefore, the second and third panels are more useful to understand the dynamics of the binary option price with respect to the strike. We observe that it is more complex because of the two contrary effects.

We now assume that the shifted log-normal model is the right model. We have:

$$\begin{aligned} \mathbb{1}\{S(T) > K\} &\Leftrightarrow \mathbb{1}\{\alpha(T) + \beta(T) X(T) > K\} \\ &\Leftrightarrow \mathbb{1}\left\{(S_0 - \alpha) e^{\left(b - \frac{1}{2}\sigma^2\right)T + \sigma W^{\mathbb{Q}}(T)} > K - \alpha e^{bT}\right\} \end{aligned}$$

We deduce that:

$$\mathbf{BCC}_{\text{SLN}}(0,S_0) = f_{\text{BS}}\left(S_0 - \alpha, K - \alpha e^{bT}, \sigma, T, b, r\right) \tag{9.31}$$

where f_{BS} is the Black-Scholes formula of the BCC option. Equation (9.31) is equivalent to shift the current price and the option strike.

TABLE 9.8: Price of the binary call option ($\alpha = -50, \sigma = 15\%$)

K	$\Sigma(T,K)$	$\omega(T,K)$	BS	SLN	SM
80	23.64	−5.47	0.8087	0.8184	0.8184
90	23.14	−4.57	0.6761	0.6895	0.6895
100	22.72	−3.87	0.5160	0.5306	0.5306
110	22.36	−3.34	0.3582	0.3715	0.3715
120	22.05	−2.92	0.2271	0.2374	0.2374

TABLE 9.9: Price of the binary call option ($\alpha = 50, \sigma = 40\%$)

K	$\Sigma(T,K)$	$\omega(T,K)$	BS	SLN	SM
80	16.71	17.25	0.8937	0.8780	0.8780
90	18.21	13.13	0.7390	0.7055	0.7055
100	19.39	10.51	0.5364	0.4971	0.4971
110	20.34	8.69	0.3546	0.3202	0.3202
120	21.14	7.35	0.2209	0.1953	0.1953

We consider the following parameters: $S_0 = 100$, $T = 1$, $b = 5\%$ and $r = 5\%$. The SLN parameters α and σ are equal to -50 and 15%. In Table 9.8, we price the binary call option with three models: the Black-Scholes model with the implied volatility $\Sigma(T,K)$, the SLN model and the SM approximation using the implied volatility $\Sigma(T,K)$ and the volatility skew $\omega(T,K)$. We remark that the Black-Scholes model produces bad option prices, whereas the SM prices are equal to those obtained with the SLN model. We obtain the same conclusion with an increasing smile as shown in Table 9.9.

The previous analysis can be extended to many other payoffs including corridor and barrier options. For instance, the holder of a corridor option receives a coupon at maturity, the magnitude of which depends on the behavior of a specified spot rate during the lifetime of the corridor. A special case is the range binary corridor option that pays a fixed coupon c if the asset stays within the range $[L,H]$:

$$f(S(T)) = c\sum_{j=1}^{n} \mathbb{1}\{S(T_j) \in [L,H]\}$$

where $\{T_1,\ldots,T_n\}$ are the fixing dates of the corridor option. Since we have:

$$\begin{aligned}\mathbb{1}\{S(T_j) \in [L,H]\} &\Leftrightarrow \mathbb{1}\{L \leq S(T_j) \leq H\}\\ &\Leftrightarrow \mathbb{1}\{S(T_j) \geq L\} - \mathbb{1}\{S(T_j) \geq H\}\end{aligned}$$

we deduce that the price $\mathbf{CC}(0,S_0)$ is related to a series of BCC cash flows:

$$\mathbf{CC}(0,S_0) = c\sum_{j=1}^{n}(\mathbf{BCC}(0,S_0,L) - \mathbf{BCC}(0,S_0,H))$$

where $\mathbf{BCC}(0,S_0,K)$ is the price of the cash-or-nothing binary call option, whose strike is K. We can then use SLN, mixed-SLN or SM models in order to take into account the volatility smile.

Remark 104 *In the case of barrier options, we can use the Black-Scholes formulas of Rubinstein and Reiner (1991) by shifting the parameters S_0, K, L and H:*

$$\begin{cases} S_0 \to S_0 - \alpha \\ K \to K - \alpha_0 e^{bT} \\ L \to L - \alpha_0 e^{bT} \\ H \to H - \alpha_0 e^{bT} \end{cases}$$

9.2.3 Local volatility model

The local volatility model has been proposed by Dupire (1994) using continuous-time modeling and, Derman and Kani (1994) in a binomial tree framework. It is one of the most famous smile models with Heston and SABR models. We assume that the risk-neutral dynamics of the asset price is given by the following SDE:

$$\mathrm{d}S(t) = bS(t)\,\mathrm{d}t + \sigma(t, S(t))\,S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

We can then retrieve the local volatility surface $\sigma(t, S)$ from the implied volatility surface $\Sigma(T, K)$, because the knowledge of all European option prices is sufficient to estimate the unique risk-neutral diffusion (Dupire, 1994).

9.2.3.1 Derivation of the forward equation

The Fokker-Planck equation Using Appendix A.3.6 on page 1072, the risk-neutral probability density function $q_t(T, S)$ of the asset price $S(T)$ satisfies the forward Chapman-Kolmogorov equation:

$$\frac{\partial\, q_t(T,S)}{\partial\, T} = -\frac{\partial\,[bSq_t(T,S)]}{\partial\, S} + \frac{1}{2}\frac{\partial^2\,\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial\, S^2}$$

The initial condition is:

$$q_t(t, S) = \mathbb{1}\{S = S_t\}$$

where S_t is the value of $S(t)$ that is known at time t.

The Breeden-Litzenberger formulas On page 508, we have seen that the risk-neutral probability measure is related to the prices of European options. In particular, we have found that:

$$\begin{aligned} \boldsymbol{\mathcal{C}}_t(T,K) &= e^{-r(T-t)}\int_K^\infty (S-K)\,q_t(T,S)\,\mathrm{d}S \\ \frac{\partial\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial\, K} &= -e^{-r(T-t)}\int_K^\infty q_t(T,S)\,\mathrm{d}S \\ \frac{\partial^2\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial\, K^2} &= e^{-r(T-t)}q_t(T,K) \end{aligned}$$

Main result We also have:

$$\begin{aligned} \frac{\partial\,\boldsymbol{\mathcal{C}}_t(T,K)}{\partial\, T} &= -r\boldsymbol{\mathcal{C}}_t(T,K) + e^{-r(T-t)}\int_K^\infty (S-K)\,\frac{\partial\, q_t(T,S)}{\partial\, T}\,\mathrm{d}S \\ &= -r\boldsymbol{\mathcal{C}}_t(T,K) + e^{-r(T-t)}\mathcal{I} \end{aligned}$$

Using the Fokker-Planck equation, we obtain:

$$\begin{aligned}
\mathcal{I} &= \int_K^\infty (S-K)\left(\frac{1}{2}\frac{\partial^2\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial S^2} - \frac{\partial\left[bSq_t(T,S)\right]}{\partial S}\right)\mathrm{d}S \\
&= \frac{1}{2}\int_K^\infty (S-K)\frac{\partial^2\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial S^2}\,\mathrm{d}S - \\
&\quad \int_K^\infty (S-K)\frac{\partial\left[bSq_t(T,S)\right]}{\partial S}\,\mathrm{d}S \\
&= \frac{1}{2}\mathcal{I}_1 - \mathcal{I}_2
\end{aligned}$$

Using an integration by parts, we have:

$$\begin{aligned}
\mathcal{I}_1 &= \int_K^\infty (S-K)\frac{\partial^2\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial S^2}\,\mathrm{d}S \\
&= \left[(S-K)\frac{\partial\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial S}\right]_K^\infty - \\
&\quad \int_K^\infty \frac{\partial\left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]}{\partial S}\,\mathrm{d}S \\
&= 0 - \left[\sigma^2(T,S)\,S^2 q_t(T,S)\right]_K^\infty \\
&= \sigma^2(T,K)\,K^2 q_t(T,K)
\end{aligned}$$

We notice that[40]:

$$\begin{aligned}
\mathcal{I}_2 &= \int_K^\infty (S-K)\frac{\partial\left[bSq_t(T,S)\right]}{\partial S}\,\mathrm{d}S \\
&= \left[(S-K)\,bSq_t(T,S)\right]_K^\infty - b\int_K^\infty Sq_t(T,S)\,\mathrm{d}S \\
&= -b\int_K^\infty Sq_t(T,S)\,\mathrm{d}S \\
&= -be^{r(T-t)}\left(\mathcal{C}_t(T,K) + K\frac{\partial\mathcal{C}_t(T,K)}{\partial K}\right)
\end{aligned}$$

The expression of $\mathcal{I}$ is then equal to:

$$\mathcal{I} = \frac{1}{2}\sigma^2(T,K)\,K^2 q_t(T,K) + be^{r(T-t)}\left(\mathcal{C}_t(T,K) - K\frac{\partial\mathcal{C}_t(T,K)}{\partial K}\right)$$

[40] Using Breeden-Litzenberger formulas, we have:

$$\begin{aligned}
e^{r(T-t)}\mathcal{C}_t(T,K) &= \int_K^\infty (S-K)\,q_t(T,S)\,\mathrm{d}S \\
&= \int_K^\infty Sq_t(T,S)\,\mathrm{d}S - K\int_K^\infty q_t(T,S)\,\mathrm{d}S \\
&= \int_K^\infty Sq_t(T,S)\,\mathrm{d}S - Ke^{r(T-t)}\frac{\partial\mathcal{C}_t(T,K)}{\partial K}
\end{aligned}$$

It follows that:

$$\begin{aligned}\frac{\partial \mathcal{C}_t(T,K)}{\partial T} &= -r\mathcal{C}_t(T,K) + \frac{1}{2}\sigma^2(T,K)K^2\frac{\partial^2 \mathcal{C}_t(T,K)}{\partial K^2} + \\ &\quad b\left(\mathcal{C}_t(T,K) - K\frac{\partial \mathcal{C}_t(T,K)}{\partial K}\right)\end{aligned}$$

We conclude that:

$$\begin{aligned}\frac{1}{2}\sigma^2(T,K)K^2\frac{\partial^2 \mathcal{C}_t(T,K)}{\partial K^2} - bK\frac{\partial \mathcal{C}_t(T,K)}{\partial K} - \\ \frac{\partial \mathcal{C}_t(T,K)}{\partial T} + (b-r)\mathcal{C}_t(T,K) = 0\end{aligned} \tag{9.32}$$

Differences between backward and forward PDE approaches Equation (9.32) is very important because it can be interpreted as the dual of the backward PDE (9.2):

$$\begin{cases} \frac{1}{2}\sigma^2(t,S)S^2\partial_S^2 V(t,S) + bS\partial_S V(t,S) + \partial_t V(t,S) - rV(t,S) = 0 \\ V(T,S(T)) = f(T,S(T),K) \end{cases}$$

where $V(t,S)$ is the price of the European option, whose terminal payoff is $f(T,S(T),K)$. In the case of Dupire model, the pricing formula becomes:

$$\begin{cases} \frac{1}{2}\sigma^2(T,K)K^2\partial_K^2 V(T,K) - bK\partial_K V(T,K) - \\ \qquad \partial_T V(T,K) + (b-r)V(T,K) = 0 \\ V(t,K) = f(t,S_t,K) \end{cases}$$

where $V(T,S)$ is the price of the European option, whose initial payoff is $f(t,S_t,K)$. In the backward formulation, the state variables are t and S, whereas the fixed variables are T and K. In the backward formulation, the state variables become T and K, whereas the fixed variables are now the current time[41] t and the current asset price S_t. This is not the only difference between the two approaches. Indeed, the backward PDE approach suggests that we can hedge the option using a dynamic portfolio of the underlying asset, whereas the forward PDE approach suggests that we can hedge the option using a static portfolio of call and put options.

We consider the pricing of an European call option with the following parameters: $S_0 = 100$, $K = 100$, $\sigma(t,S) = 20\%$, $T = 0.5$, $b = 2\%$ and $r = 5\%$. In the case of the backward PDE, we consider the usual boundary conditions:

$$\begin{cases} \mathcal{C}(t,S) = 0 \\ \partial_S \mathcal{C}(t,+\infty) = 1 \end{cases}$$

For the forward PDE, the boundary conditions are[42]:

$$\begin{cases} \partial_K \mathcal{C}(T,0) = -1 \\ \mathcal{C}(T,+\infty) = 0 \end{cases}$$

In Figure 9.24, we show the relative error (expressed in bps) of numerical solutions when considering the Crank-Nicholson scheme. In the case of the backward PDE, the state variable

[41] t can be equal to zero.

[42] We can also use the following specifications:

$$\begin{cases} \mathcal{C}(T,0) = e^{(b-r)T}S_0 \\ \partial_K \mathcal{C}(T,+\infty) = 0 \end{cases}$$

is the current asset price S_0, and we obtain all the option prices when the strike is equal to 100. In the case of the forward PDE, the state variable is the strike K, and we obtain all the option prices when the current asset price is equal to 100. We notice that the relative errors are equivalent when S_0 is equal to K. In fact, the efficiency of the numerical algorithms will depend on the relative position between S_0 and K.

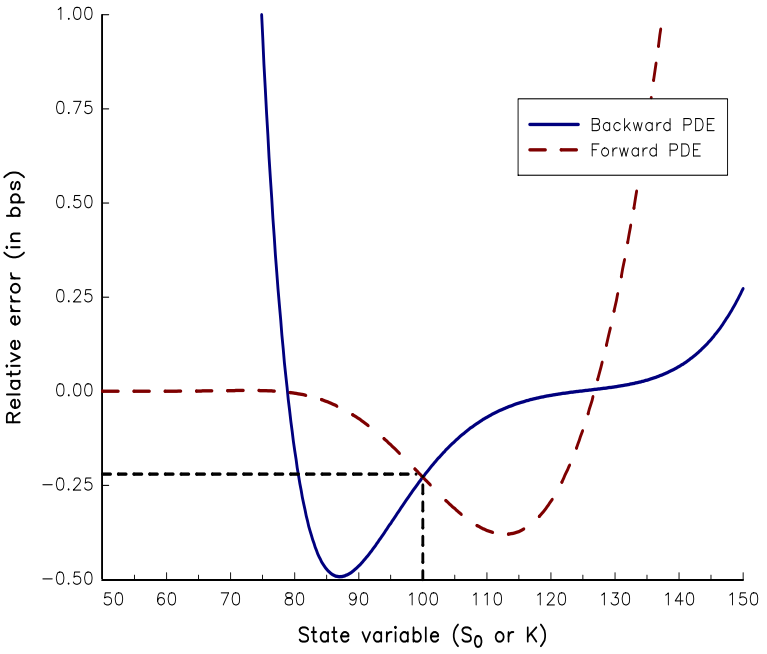


FIGURE 9.24: Relative error of backward and forward PDE numerical solutions

9.2.3.2 Duality between local volatility and implied volatility

We can inverse Equation (9.32) in order to relate the expression of the local volatility and the price of the call option:

$$\sigma^2(T,K) = 2\frac{bK\partial_K \boldsymbol{\mathcal{C}}(T,K) + \partial_T \boldsymbol{\mathcal{C}}(T,K) - (b-r)\boldsymbol{\mathcal{C}}(T,K)}{K^2\partial_K^2 \boldsymbol{\mathcal{C}}(T,K)}$$

In Exercise 9.4.8 on page 599, we show that $\sigma(T,K)$ can also be written with respect to the implied volatility $\Sigma(T,K)$:

$$\sigma(T,K) = \sqrt{\frac{A(T,K)}{B(T,K)}} \tag{9.33}$$

where:

$$\begin{aligned} A(T,K) &= \Sigma^2(T,K) + 2bKT\Sigma(T,K)\,\partial_K\Sigma(T,K) + \\ &\quad 2T\Sigma(T,K)\,\partial_T\Sigma(T,K) \end{aligned}$$

and:

$$\begin{aligned} B(T,K) &= 1 + 2K\sqrt{T}d_1\partial_K\Sigma(T,K) + K^2T\Sigma(T,K)\,\partial_K^2\Sigma(T,K) + \\ &\quad K^2Td_1d_2\left(\partial_K\Sigma(T,K)\right)^2 \end{aligned}$$

Equation (9.33) is the key finding of Dupire (1994). Indeed, knowing the implied volatility surface, we can retrieve the unique local volatility function that matches the set of all European call and put option prices.

Many results have been derived from Equation (9.33). For instance, if there is no skew[43], the local volatility function does not depend on the strike[44]:

$$\sigma^2(T) = \Sigma^2(T) + 2T\Sigma(T)\frac{\partial \Sigma(T)}{\partial T} \tag{9.34}$$

On the contrary, the local volatility always depends on the maturity T even if there is no time-variation in the implied volatility[45].

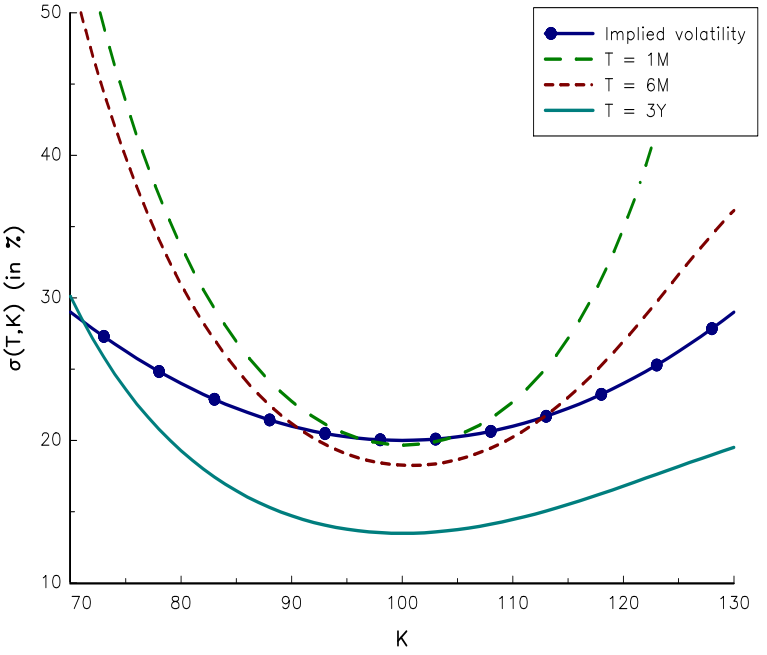


FIGURE 9.25: Calibrated local volatility $\sigma(T, S)$ (in %)

Example 91 *We assume that the implied volatility is equal to:*

$$\Sigma(T, K) = \Sigma_0 + \alpha(S_0 - K)^2$$

where $\Sigma_0 = 20\%$, $\alpha = 1$ *bp*, $S_0 = 100$ *and* $b = 5\%$.

Figure 9.25 shows the calibrated local volatility for different values of T. We verify the time-variation property of the local volatility. We notice that Equation (9.34) is equivalent to:

$$\sigma^2(T) = \frac{\partial T\Sigma^2(T)}{\partial T}$$

or:

$$\Sigma^2(T) = \frac{1}{T}\int_0^T \sigma^2(t)\,\mathrm{d}t$$

The implied variance is then the time series average of the local variance.

[43] We have $\Sigma(T, K) = \Sigma(T)$.
[44] This result is obtained by setting $\partial_K \Sigma(T, K)$ and $\partial_K^2 \Sigma(T, K)$ equal to 0 in Equation (9.33).
[45] We have $\Sigma(T, K) = \Sigma(K)$.

Another important result concerns the behavior of the implied volatility near expiry. Let x be the log-moneyness:

$$\begin{aligned} x &= \varphi(T,K) \\ &= \ln\frac{S_0}{K} + bT \end{aligned}$$

We introduce the functions $\tilde{\Sigma}$ and $\tilde{\sigma}$ such that $\Sigma(T,K) = \tilde{\Sigma}(T,\varphi(T,K))$ and $\sigma(T,K) = \tilde{\sigma}(T,\varphi(T,K))$. Berestycki *et al.* (2002) showed that the implied volatility is the harmonic mean of the local volatility[46]:

$$\frac{1}{\tilde{\Sigma}(0,x)} = \int_0^1 \frac{\mathrm{d}y}{\tilde{\sigma}(0,xy)}$$

It follows that:

$$\frac{\partial\tilde{\Sigma}(0,0)}{\partial x} = \frac{1}{2}\frac{\partial\tilde{\sigma}(0,0)}{\partial x}$$

The ATM slope of the implied volatility near expiry is equal to one half the slope of the local volatility.

9.2.3.3 Dupire model in practice

One of the problems is the availability of the call/put prices for all maturities and all strikes. In practice, we only know the option price for some maturities T_m and some strikes K_i. This is why we have to use a calibration method to obtain the continuous volatility surface $\Sigma(T,K)$.

Time interpolation We note $\upsilon(T,K)$ the total implied variance:

$$\upsilon(T,K) = T\Sigma^2(T,K)$$

The linear interpolation of the total implied variance gives:

$$\upsilon(T,K) = w\cdot\upsilon(T_m,K_m(T)) + (1-w)\cdot\upsilon(T_{m+1},K_{m+1}(T))$$

where $T\in[T_m,T_{m+1}]$ and:

$$w = \frac{T_{m+1}-T}{T_{m+1}-T_m}$$

We deduce that:

$$\begin{aligned} \Sigma^2(T,K) &= \frac{T_m(T_{m+1}-T)}{T(T_{m+1}-T_m)}\Sigma^2(T_m,K_m(T)) + \\ &\quad \frac{T_{m+1}(T-T_m)}{T(T_{m+1}-T_m)}\Sigma^2(T_{m+1},K_{m+1}(T)) \\ &= a_m(T)\,\Sigma^2(T_m,K_m(T)) + b_{m+1}(T)\,\Sigma^2(T_{m+1},K_{m+1}(T)) \end{aligned}$$

where:

$$a_m(T) = \frac{T_m(T_{m+1}-T)}{T(T_{m+1}-T_m)}$$

[46] See Exercise 9.4.8 on page 599 for the proof of this result.

and:

$$b_{m+1}(T) = \frac{T_{m+1}(T - T_m)}{T(T_{m+1} - T_m)} = 1 - a_m(T)$$

In the previous scheme, we interpolate the total variance for the strike K and the maturity T by considering the pairs $(T_m, K_m(T))$ and $(T_{m+1}, K_{m+1}(T))$. Generally, the strikes $K_m(T)$ and $K_{m+1}(T)$ are a translation of the strike K:

$$\begin{cases} K_m(T) = k_m \cdot (T) K \\ K_{m+1}(T) = k_{m+1} \cdot (T) K \end{cases}$$

with $k_m(T_m) = 1$ and $k_{m+1}(T_{m+1}) = 1$. The simplest rule is $k_m(T) = k_{m+1}(T) = 1$. Another method is to define $k_m(T_m) = e^{-b(T-T_m)} \leq 1$ and $k_{m+1}(T_{m+1}) = e^{b(T_{m+1}-T)} \geq 1$.

Example 92 *We assume that the implied volatility is equal to:*

$$\Sigma(T_m, K) = \Sigma_m + \alpha_m (K - 100)^2$$

where $\Sigma_m = 20\% + 0.005 \cdot (T_m - 1.0)$, $\alpha_m = 0.05 \cdot T_m$ *bps and* T_m *is equal to* 1, 2, 3, 4 *and* 5 *years. The cost-of-carry parameter* b *is set to* 5%.

We have represented the implied volatility $\Sigma(T_m, K)$ in the first panel in Figure 9.26. We can then compute the volatility surface. When T is lower than the first observed maturity or higher than the last observed maturity, we can extrapolate the implied volatility in several ways. The simplest method is to assume that the implied volatility is constant. In the third panel, we have reported the interpolated implied volatility with respect to the maturity T for three different strikes. We notice that it is curved between two interpolating knots due to the effect of the square root transformation.

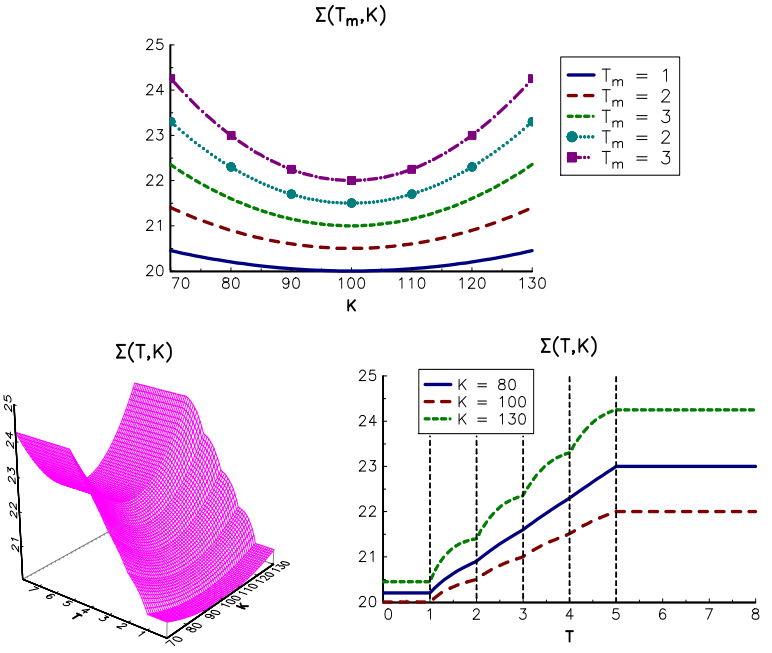


FIGURE 9.26: Time interpolation of the implied volatility

Non-parametric interpolation We note $\mathcal{S}_m(K)$ the non-parametric function that give the value of $\Sigma(T_m, K)$ for all values of strike K. The calculation of the local volatility surface implies to calculate the quantities $\partial_K \Sigma(T,K)$, $\partial_K^2 \Sigma(T,K)$ and $\partial_T \Sigma(T,K)$. We use the shortened notations: $\mathcal{S}_m = \mathcal{S}_m(K_m(T))$, $\mathcal{S}'_m = \mathcal{S}'_m(K_m(T))$, $\mathcal{S}''_m = \mathcal{S}''_m(K_m(T))$, $\mathcal{S}_{m+1} = \mathcal{S}_{m+1}(K_{m+1}(T))$, $\mathcal{S}'_{m+1} = \mathcal{S}'_{m+1}(K_{m+1}(T))$ and $\mathcal{S}''_{m+1} = \mathcal{S}''_{m+1}(K_{m+1}(T))$. We have:

$$\begin{aligned}
\Sigma(T,K)\,\partial_K \Sigma(T,K) &= \frac{1}{2}\partial_K \Sigma^2(T,K) \\
&= a_m(T)\,k_m(T)\,\mathcal{S}_m\mathcal{S}'_m + \\
&\quad b_{m+1}(T)\,k_{m+1}(T)\,\mathcal{S}_{m+1}\mathcal{S}'_{m+1}
\end{aligned}$$

For the second term, we obtain:

$$\begin{aligned}
\Sigma(T,K)\,\partial_K^2 \Sigma(T,K) &= \frac{1}{2}\partial_K^2 \Sigma^2(T,K) - (\partial_K \Sigma(T,K))^2 \\
&= a_m(T)\,k_m^2(T)\left(\mathcal{S}_m\mathcal{S}''_m + (\mathcal{S}'_m)^2\right) + \\
&\quad b_{m+1}(T)\,k_{m+1}^2(T)\left(\mathcal{S}_{m+1}\mathcal{S}''_{m+1} + (\mathcal{S}'_{m+1})^2\right) - \\
&\quad (\partial_K \Sigma(T,K))^2
\end{aligned}$$

Since we have:

$$a'_m(T) = \frac{-T_m T_{m+1}}{T^2(T_{m+1} - T_m)}$$

and:

$$b'_{m+1}(T) = \frac{T_m T_{m+1}}{T^2(T_{m+1} - T_m)}$$

we deduce that the last term is equal to[47]:

$$\begin{aligned}
\Sigma(T,K)\,\partial_T \Sigma(T,K) &= \frac{1}{2}\partial_T \Sigma^2(T,K) \\
&= \frac{1}{2}a'_m(T)\,\Sigma^2(T_m, K_m(T)) + \\
&\quad \frac{1}{2}b'_{m+1}(T)\,\Sigma^2(T_{m+1}, K_{m+1}(T)) + \\
&\quad a_m(T)\,\Sigma(T_m, K_m(T))\,\partial_T \Sigma(T_m, K_m(T)) + \\
&\quad b_{m+1}(T)\,\Sigma(T_{m+1}, K_{m+1}(T))\,\partial_T \Sigma(T_{m+1}, K_{m+1}(T)) \\
&= \frac{1}{2}(\mathcal{S}_{m+1} - \mathcal{S}_m)\frac{T_m T_{m+1}}{T^2(T_{m+1} - T_m)} + \\
&\quad a_m(T)\,\mathcal{S}_m\mathcal{S}'_m K \partial_T k_m(T) + \\
&\quad b_m(T)\,\mathcal{S}_{m+1}\mathcal{S}'_{m+1} K \partial_T k_{m+1}(T)
\end{aligned}$$

In the case where $k_m(T) = k_{m+1}(T) = 1$, the previous formula reduces to:

$$\Sigma(T,K)\,\partial_T \Sigma(T,K) = \frac{1}{2}(\mathcal{S}_{m+1} - \mathcal{S}_m)\frac{T_m T_{m+1}}{T^2(T_{m+1} - T_m)}$$

In practice, we don't observe the function $\mathcal{S}_m(K)$, but only few values of $\Sigma(T_m, K_i)$ for some maturities T_m and some strikes K_i. An example is given in Table 9.10. We assume that

[47] We use the fact that $\partial_T K_m(T) = K\partial_T k_m(T)$ and $\partial_T K_{m+1}(T) = K\partial_T k_{m+1}(T)$.

TABLE 9.10: Calibration set

$T_m = 1/12$									
K_i	87.0	92.0	96.0	98.0	100.0	103.0	106.0	110.0	116.0
$\Sigma(T_m, K_i)$	13.7	13.7	13.3	13.2	13.0	13.1	13.2	13.5	13.5
$T_m = 3/12$									
K_i	77.0	85.0	93.0	97.0	101.0	106.0	111.0	121.0	134.0
$\Sigma(T_m, K_i)$	14.9	14.9	14.1	14.0	13.5	13.8	14.2	15.1	15.1
$T_m = 6/12$									
K_i	66.0	78.0	89.0	96.0	102.0	111.0	119.0	136.0	161.0
$\Sigma(T_m, K_i)$	16.8	16.8	15.5	15.0	14.5	15.0	15.5	16.8	16.8
$T_m = 1$									
K_i	53.0	69.0	86.0	96.0	104.0	119.0	133.0	166.0	217.0
$\Sigma(T_m, K_i)$	19.0	19.0	17.0	16.0	15.5	16.5	17.5	18.5	18.5
$T_m = 2$									
K_i	37.0	56.0	80.0	96.0	103.0	137.0	163.0	229.0	347.0
$\Sigma(T_m, K_i)$	21.9	21.9	20.0	18.5	18.5	19.0	19.5	20.8	20.8

five maturities are quoted (1M, 3M, 6M, 1Y and 2Y). For each maturity, we observe the implied volatility (expressed in %) for 9 strikes. This is why we have to use an interpolation method. In Figure 9.27, we have represented the function $\mathcal{S}_m(K)$ obtained with the cubic spline method[48]. One of the issues is the interpolated implied volatility on the wings. Here, we have chosen to keep the cubic spline values, but an alternative approach is to assume that the smile is constant before the first strike and after the last strike. Let us assume that $S_0 = 100$, $b = 5\%$ and $r = 5\%$. Using the time approximation approach, we obtain the implied volatility surface given in Figure 9.28. The implied volatility is constant when $T \leq 1/12$ and $T \geq 2$. Finally, the local volatility surface is reported in Figure 9.29. We notice that it is not a smooth function. This is why we can use cubic spline approximation or other smoothing methods in place of cubic spline interpolation[49]. However, we not not retrieve exactly the quoted implied volatilities with this approach.

Remark 105 *In real life, the number of strikes may be different from one maturity to another, and may be smaller. For example, in the case of currency options*[50]*, we generally have* 5 *quoted options (ATM,* 10*-delta call,* 25*-delta call,* 10*-delta put and* 25*-delta put).*

Parametric calibration In the previous section, $\Sigma(T, K)$ and $\sigma(T, K)$ are calibrated using non-parametric approaches such as the cubic spline method. This produces a disorderly local volatility surface. In order to avoid this problem, we can use a parametric framework. For instance, we can calibrate $\Sigma(T, K)$ using the SABR model. Another popular approach is to consider the stochastic volatility inspired or SVI parametrization.

We recall that the total implied variance is equal to:

$$\upsilon(T, K) = T\Sigma^2(T, K)$$

We assume that $\upsilon(T, K) = \tilde{\upsilon}(T, x)$ and $\Sigma(T, K) = \tilde{\Sigma}(T, x)$ where x is the log-moneyness:

$$x = \varphi(T, K) = \ln \frac{K}{F(T)} = \ln \frac{K}{S_0 e^{bT}}$$

[48] See Appendix A.1.2.1 on page 1035.
[49] See Crépey (2003) and Fengler (2009).
[50] FX vanilla options are generally quoted in terms of volatility with respect to a fixed delta, and not in terms of premium with respect to a given strike.

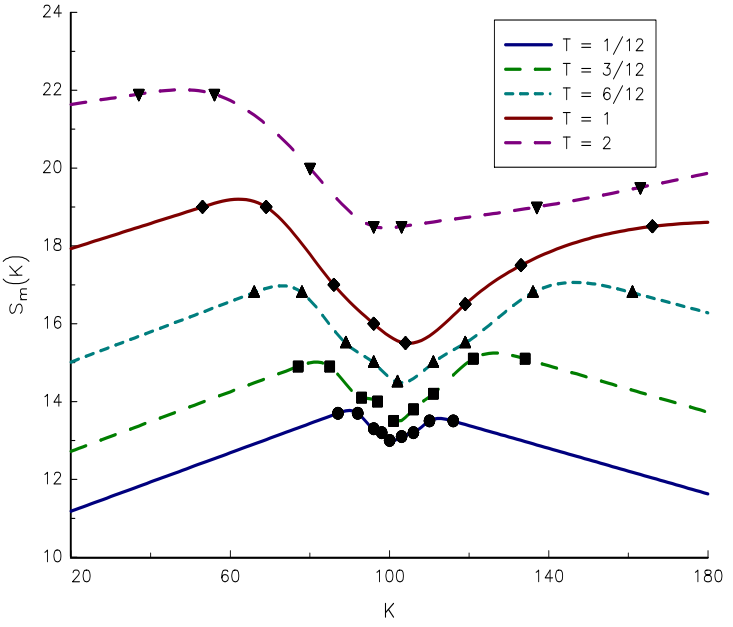


FIGURE 9.27: Cubic spline interpolation $\mathcal{S}_m(K)$ (in %)

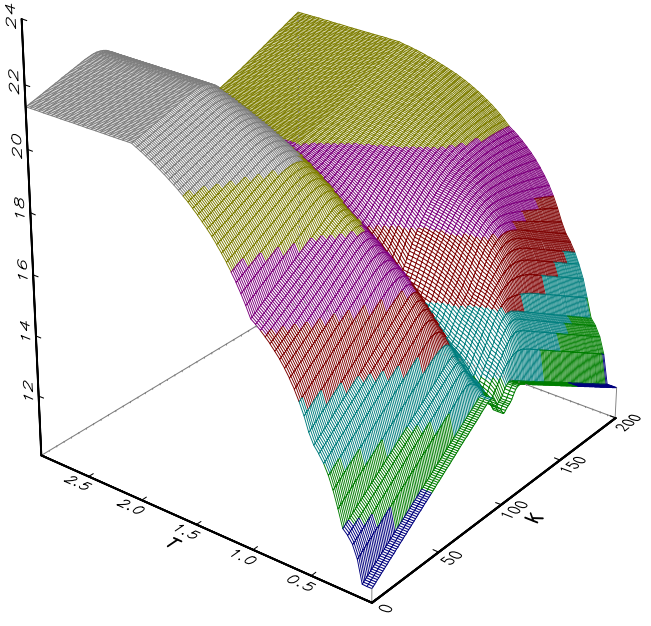


FIGURE 9.28: Implied volatility surface $\Sigma(T, K)$ (in %)

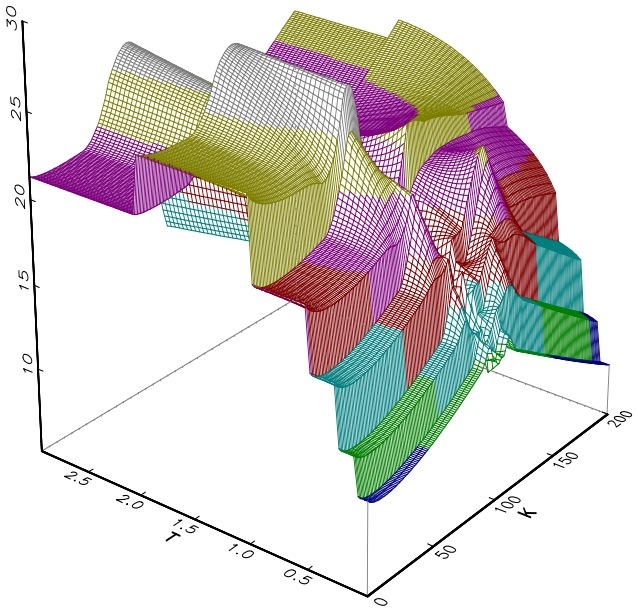


FIGURE 9.29: Local volatility surface $\sigma(T,K)$ (in %)

Let $\tilde{v}_T(x) = \tilde{v}(T,x)$ be the total implied variance for a given maturity slice. Gatheral (2004) introduces the following SVI parametrization:

$$\tilde{v}_T(x) = \alpha + \beta\left(\rho(x-m) + \sqrt{(x-m)^2+\sigma^2}\right)$$

where $\beta > 0$, $\sigma > 0$ and $\rho \in [-1,1]$. We have:

$$\tilde{v}_T(m) = \alpha + \beta\sigma$$

and:

$$\begin{cases} \lim_{x\to-\infty} \tilde{v}_T(x) = \alpha - \beta(1-\rho)(x-m) \\ \lim_{x\to\infty} \tilde{v}_T(x) = \alpha + \beta(1+\rho)(x-m) \end{cases}$$

Gatheral deduces that α controls the general level, β influences the slope of the wings, σ changes the curvature of the smile, ρ impacts the symmetry of the smile while m shifts the smile.

Example 93 *We assume that* $\alpha = 2\%$, $\beta = 0.3$, $\sigma = 10\%$, $\rho = -40\%$ *and* $m = 0$. *Figure 9.30 shows the impact of each parameter on the total variance* $\tilde{v}_T(x)$.

Gatheral and Jacquier (2014) show that a volatility surface is free of static arbitrage if and only if it is free of calendar spread arbitrage[51] and each time slice is free of butterfly arbitrage[52]. The first property implies that:

$$\partial_T \tilde{v}(T,x) \geq 0$$

[51]This means that the price of an European option is monotone with the maturity.
[52]This means that the probability density function is non-negative for any given maturity T.

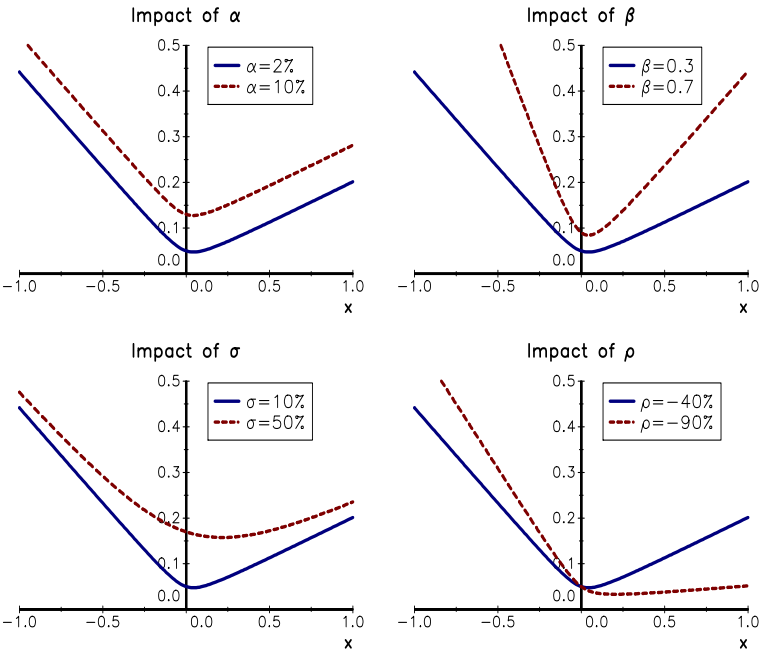


FIGURE 9.30: Impact of SVI parameters on the total variance $\tilde{v}_T(x)$

for all $x \in \mathbb{R}$. Thanks to Breeden and Litzenberger (1978), the second property is equivalent to verify that[53]:

$$\frac{\partial^2 \boldsymbol{C}(T,K)}{\partial K^2} \geq 0$$

These authors deduce then how the absence of static arbitrage impacts SVI parameters.

We consider the calibration set defined in Table 9.10 on page 554. We delete the two extreme strikes of each maturity[54]. In Figure 9.31, we show the SVI parametrization for each maturity. By considering the time interpolation presented previously, we can define the implied volatility surface $\Sigma(T,K)$ and then calculate the local $\sigma(T,K)$. These two volatility surfaces are reported in Figure 9.31.

Hedging coefficients Let $\Sigma(T,K,S_t)$ and $\sigma(T,K,S_t)$ be the implied and local volatility surfaces that depend on the current price S_t. We also write the value of the option $V(T,K,S_t)$ as a function of the maturity T, the strike K and the current price S_t. The delta of the option is then equal to:

$$\boldsymbol{\Delta}_t = \frac{\partial V(T,K,S_t)}{\partial S_t}$$

If we use the finite difference approximation, we obtain:

$$\boldsymbol{\Delta} \approx \frac{V(T,K,S_t+\varepsilon) - V(T,K,S_t-\varepsilon)}{2\varepsilon}$$

[53] See Section 9.1.1.4 on page 508.

[54] In fact, we have added these two points in the calibration set in order to stabilize the non-parametric calibration. However, this approach is not adequate because volatility smile is linear and not constant at extreme strikes (Lee, 2004).

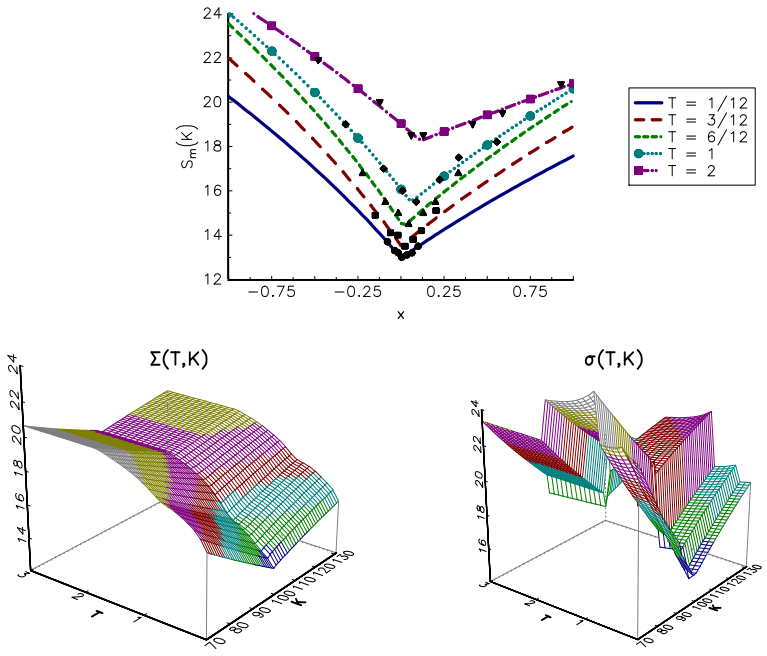


FIGURE 9.31: SVI parametrization, implied volatility $\Sigma(T,K)$ and local volatility $\sigma(T,K)$ (in %)

Computing the option price and its corresponding delta require then to calculate three local volatility surfaces[55] and solve the forward PDE three times[56]. This method can also be used to calculate the gamma of the option, because we have:

$$\mathbf{\Gamma} \approx \frac{V(T,K,S_t+\varepsilon) - 2V(T,K,S_t) + V(T,K,S_t-\varepsilon)}{\varepsilon^2}$$

The vega coefficient in a local volatility model is not well-defined. It can be measured with respect to the local volatility $\sigma(T,K,S_t)$ or the implied volatility $\Sigma(T,K,S_t)$. The most frequent approach is to measure the vega as the sensitivity of the price to a parallel shift of $\Sigma(T,K,S_t)$. We have:

$$\boldsymbol{\upsilon} = \frac{V'(T,K,S_t) - V(T,K,S_t)}{\varepsilon'}$$

where $V'(T,K,S_t)$ is the option price obtained when the implied volatility surface is $\Sigma(T,K,S_t)+\varepsilon'$.

One of the issues with the local volatility model is that greeks are not easy to compute and are not stable in the time and across strikes. This is a severe disadvantage, since the hedging of the option is not straightforward and generally less efficient than the hedging portfolio given by the Black-Scholes model:

> "*Market smiles and skews are usually managed by using local volatility models* a la *Dupire. We discover that the dynamics of the market smile predicted by local vol models is opposite of observed market behavior: when the price of the underlying decreases, local vol models predict that the smile shifts to higher*

[55] We have to calculate $\sigma(T,K,S_t-\varepsilon)$, $\sigma(T,K,S_t)$ and $\sigma(T,K,S_t+\varepsilon)$.
[56] We have to calculate $V(T,K,S_t-\varepsilon)$, $V(T,K,S_t)$ and $V(T,K,S_t+\varepsilon)$.

prices; when the price increases, these models predict that the smile shifts to lower prices. Due to this contradiction between model and market, delta and vega hedges derived from the model can be unstable and may perform worse than naive Black-Scholes' hedges" (Hagan *et al.*, 2002, page 84).

9.2.3.4 Application to exotic options

Another shortcoming of the local volatility model is the unrealistic probability distribution of the conditional random variable $S(t_2) \mid S(t_1)$. This is why this model is only used for European options, and not for path-dependent derivatives. In particular, it has been popular in the 1990s and 2000s for pricing European barrier options.

We consider the calibration set given in Table 9.10 on page 554. We assume that S_0, $b = 5\%$ and $5 = 5\%$. We price different payoffs given in Table 9.11, whose parameters are $K = 100$, $L = 90$ and $H = 115$. The maturity is set to one year. Prices are calculated with a Crank-Nicholson scheme with 2 000 discretization points[57] in space, 2 000 discretization points in time and traditional boundary conditions[58]. Results are given in column LV. We can compare them with Black-Scholes prices calculated with implied volatilities[59] $\Sigma_1 = 16\%$ and $\Sigma_2 = 15.5\%$. For each payoff and each value of implied volatility, we report two values of the option price: one obtained by solving the PDE and another one calculated with the analytical formulas of Rubinstein and Reiner (1991). We observe some differences between the two prices, because the PDE price depends on the choice of the discretization scheme and the boundary conditions. We notice that the prices DOC, UOC, KOC and BCC calculated with the local volatility model are not in the interval of BS prices.

TABLE 9.11: Barrier option pricing with the local volatility model

Option	Payoff	LV	BS-PDE		BS-RR	
			Σ_1	Σ_2	Σ_1	Σ_2
Call	$(S(T) - K)^+$	8.85	8.96	8.78	8.96	8.78
Put	$(K - S(T))^+$	3.97	4.08	3.90	4.08	3.90
DOC	$\mathbb{1}\{S(t) > L\} \cdot (S(T) - K)^+$	7.98	8.14	8.05	8.11	8.02
DOP	$\mathbb{1}\{S(t) > L\} \cdot (K - S(T))^+$	0.26	0.27	0.28	0.25	0.27
UOC	$\mathbb{1}\{S(t) < H\} \cdot (S(T) - K)^+$	0.99	0.88	0.94	0.83	0.89
UOP	$\mathbb{1}\{S(t) < H\} \cdot (K - S(T))^+$	3.81	3.90	3.75	3.89	3.74
KOC	$\mathbb{1}\{S(t) \in [L, H]\} \cdot (S(T) - K)^+$	0.65	0.56	0.64	0.52	0.59
KOP	$\mathbb{1}\{S(t) \in [L, H]\} \cdot (K - S(T))^+$	0.20	0.20	0.22	0.19	0.21
BCC	$\mathbb{1}\{S(T) \geq K\}$	0.58	0.56	0.57	0.56	0.57
BCP	$\mathbb{1}\{S(T) \leq K\}$	0.37	0.39	0.38	0.39	0.38

[57] We assume that $S(t) \in [0, 200]$.

[58] We use the following Dirichlet and Neumann conditions:

$V(t, S^-) = 0$	$V(t, S^+) = 0$	$\partial_S V(t, S^-) = -1$	$\partial_S V(t, S^+) = 0$
Call, BCC	Put, BCP	Put, BCP	Call, BCC
DOC, DOP, UOC	DOP, UOC, UOP	UOP	DOC
KOC, KOP	KOC, KOP		

where $S^- = 0$ and $S^+ = 200$.

[59] $\Sigma_1 = 16\%$ and $\Sigma_2 = 15.5\%$ correspond to the two implied volatilities of strikes 96 and 104 for the one-year maturity.

9.2.4 Stochastic volatility models

The most popular approach to model the volatility smile is to consider that the volatility is not constant, but stochastic. In this case, we obtain a model with two state variables, which are the spot price $S(t)$ and the volatility $\sigma(t)$. After deriving the general formula of the fundamental pricing equation, we present Heston and SABR models, which are the two most important parametrizations of this class of models.

9.2.4.1 General analysis

Pricing formula We assume that the joint dynamics of the spot price $S(t)$ and the stochastic volatility $\sigma(t)$ is:

$$\begin{cases} \mathrm{d}S(t) = \mu(t) S(t)\ \mathrm{d}t + \sigma(t) S(t)\ \mathrm{d}W_1(t) \\ \mathrm{d}\sigma(t) = \zeta(\sigma(t))\ \mathrm{d}t + \xi(\sigma(t))\ \mathrm{d}W_2(t) \end{cases}$$

where $\mathbb{E}[W_1(t) W_2(t)] = \rho t$. $S(t)$ is a geometric Brownian motion with time-varying parameters $\mu(t)$ and $\sigma(t)$, whereas $\sigma(t)$ follows a general diffusion that does not depend on $S(t)$. In the Black-Scholes model, the volatility has the status of parameter. In this new approach, the volatility is a second state variable. The SV model is defined by the functions $\zeta(y)$ and $\xi(y)$.

Using Itô's lemma, we can show that the fundamental pricing equation defined on page 492 becomes[60]:

$$\begin{aligned} \frac{1}{2}\sigma^2 S^2 \partial_S^2 V(t,S,\sigma) + \rho\sigma S\xi(\sigma)\partial_{S,\sigma}^2 V(t,S,\sigma) + \frac{1}{2}\xi^2(\sigma)\partial_\sigma^2 V(t,S,\sigma) \\ + (\mu - \lambda_S\sigma) S\partial_S V(t,S,\sigma) + (\zeta(\sigma) - \lambda_\sigma\xi(\sigma))\partial_\sigma V(t,S,\sigma) \\ + \partial_t V(t,S,\sigma) - rV(t,S,\sigma) = 0 \end{aligned}$$

where $V(t,S,\sigma)$ is the price of the contingent claim, $V(T,S(T)) = f(S(T))$ and $f(S(T))$ is the option payoff. As previously, the market price of the spot risk $W_1(t)$ is:

$$\lambda_S(t) = \frac{\mu(t) - b(t)}{\sigma(t)}$$

By introduction the function $\zeta'(y)$:

$$\zeta'(\sigma(t)) = \zeta(\sigma(t)) - \lambda_\sigma(t)\xi(\sigma(t))$$

we obtain the following PDE:

$$\begin{aligned} \frac{1}{2}\sigma^2 S^2 \partial_S^2 V(t,S,\sigma) + \rho\sigma S\xi(\sigma)\partial_{S,\sigma}^2 V(t,S,\sigma) + \frac{1}{2}\xi^2(\sigma)\partial_\sigma^2 V(t,S,\sigma) \\ + bS\partial_S V(t,S,\sigma) + \zeta'(\sigma)\partial_\sigma V(t,S,\sigma) + \partial_t V(t,S,\sigma) - rV(t,S,\sigma) = 0 \end{aligned} \tag{9.35}$$

Equation (9.35) is the equivalent of Equation (9.2) on page 492 when the volatility is stochastic.

Using the Girsanov theorem, we deduce that the risk-neutral dynamics is:

$$\begin{cases} \mathrm{d}S(t) = b(t) S(t)\ \mathrm{d}t + \sigma(t) S(t)\ \mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}\sigma(t) = \zeta'(\sigma(t))\ \mathrm{d}t + \xi(\sigma(t))\ \mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

[60] We omit the dependence in t in order to simplify the notation.

The martingale solution is then equal to:

$$V_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(t)\,\mathrm{d}t} f\left(S\left(T\right)\right)\middle|\,\mathcal{F}_0\right]$$

We retrieve the formula obtained in the one-dimensional case. However, the computation of the expected value is now more complex since $S\left(T\right)$ depends on the trajectory of the volatility $\sigma\left(t\right)$.

Hedging portfolio The computation of greek coefficients is more complex in SV models. This is why the definition of the hedging portfolio is not straightforward and depends on the assumption on the smile dynamics. In the case of the Black-Scholes model, delta and vega sensitivities are equal to:

$$\boldsymbol{\Delta}_{\mathrm{BS}} = \frac{\partial\, V_{\mathrm{BS}}\left(S_0, K, \Sigma, T\right)}{\partial\, S_0}$$

and:

$$\boldsymbol{\upsilon}_{\mathrm{BS}} = \frac{\partial\, V_{\mathrm{BS}}\left(S_0, K, \Sigma, T\right)}{\partial\, \Sigma}$$

In the case of the stochastic volatility model, we have:

$$\boldsymbol{\Delta}_{\mathrm{SV}} = \frac{\partial\, V_{\mathrm{SV}}\left(S_0, K, \sigma_0, T\right)}{\partial\, S_0}$$

If we assume that $V_{\mathrm{SV}}\left(S_0, K, \sigma_0, T\right) = V_{\mathrm{BS}}\left(S_0, K, \Sigma_{\mathrm{SV}}\left(T, S_0\right), T\right)$, we obtain:

$$\begin{aligned}
\boldsymbol{\Delta}_{\mathrm{SV}} &= \frac{\partial\, V_{\mathrm{BS}}\left(S_0, K, \Sigma_{\mathrm{SV}}, T\right)}{\partial\, S_0} + \frac{\partial\, V_{\mathrm{BS}}\left(S_0, K, \Sigma_{\mathrm{SV}}, T\right)}{\partial\, \Sigma_{\mathrm{SV}}} \cdot \frac{\partial\, \Sigma_{\mathrm{SV}}\left(T, S_0\right)}{\partial\, S_0} \\
&= \boldsymbol{\Delta}_{\mathrm{BS}} + \boldsymbol{\upsilon}_{\mathrm{BS}} \cdot \frac{\partial\, \Sigma_{\mathrm{SV}}\left(T, S_0\right)}{\partial\, S_0}
\end{aligned}$$

Therefore, the delta of the SV model depends on the BS vega. Generally, we have $\partial_{S_0}\Sigma_{\mathrm{SV}}\left(T, S_0\right) \geq 0$ implying that $\boldsymbol{\Delta}_{\mathrm{SV}} \geq \boldsymbol{\Delta}_{\mathrm{BS}}$.

The calculation of the vega coefficient is a second issue. Indeed, the natural hedging portfolio should consist in two long/short exposures since we have two risk factors $S\left(t\right)$ and $\sigma\left(t\right)$. Therefore, we can define the vega sensitivity as follows:

$$\boldsymbol{\upsilon}_{\mathrm{SV}} = \frac{\partial\, V_{\mathrm{SV}}\left(S_0, K, \sigma_0, T\right)}{\partial\, \sigma_0}$$

However, this definition has no interest since the stochastic volatility $\sigma\left(t\right)$ cannot be directly or even indirectly trade. This is why most of traders prefer to use a BS vega:

$$\boldsymbol{\upsilon}_{\mathrm{SV}} = \frac{\partial\, V_{\mathrm{BS}}\left(S_0, K, \Sigma_{\mathrm{SV}}\left(T, S_0\right), T\right)}{\partial\, \Sigma_{\mathrm{SV}}}$$

Here, we make the assumption that the vega is calculated with respect to the implied volatility $\Sigma_{\mathrm{SV}}\left(T, S_0\right)$ deduced from the stochastic volatility model. It can be viewed as a pure Black-Scholes vega, but most of times, it corresponds to a shift of the implied volatility surface. This approach requires a new calibration of the stochastic volatility parameters. In some sense, the vega can be viewed as the difference between the prices obtained with two stochastic volatility models.

9.2.4.2 Heston model

Heston (1993) assumes that the stochastic differential equation of the spot price is equal to:

$$\begin{cases} \mathrm{d}S(t) = \mu S(t)\,\mathrm{d}t + \sqrt{v(t)}S(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}v(t) = \kappa(\theta - v(t))\,\mathrm{d}t + \xi\sqrt{v(t)}\,\mathrm{d}W_2(t) \end{cases}$$

where $S(0) = S_0$, $v(0) = v_0$ and $W(t) = (W_1(t), W_2(t))$ is a two-dimensional Wiener process with $\mathbb{E}[W_1(t) W_2(t)] = \rho t$. We notice that the stochastic variance $v(t)$ follows a CIR process: θ is the long-run variance, κ is the mean-reverting parameter and ξ is the volatility of the variance (also called the vovol parameter).

Remark 106 *We have $\sigma(t) = \sqrt{v(t)}$ and:*

$$\mathrm{d}\sigma(t) = \left(\left(\frac{\kappa\theta}{2} - \frac{\xi^2}{8}\right)\frac{1}{\sigma(t)} - \frac{1}{2}\kappa\sigma(t)\right)\mathrm{d}t + \frac{1}{2}\xi\,\mathrm{d}W_2(t)$$

The stochastic volatility is then an Ornstein-Uhlenbeck process if we impose $\theta = \xi^2/(4\kappa)$.

As the second state variable of the Heston model is the stochastic variance $v(t)$, the price $V(t, S, v)$ of the option must satisfy the PDE[61]:

$$\begin{aligned} &\frac{1}{2}vS^2\partial_S^2 V + \rho\xi vS\partial_{S,v}^2 V + \frac{1}{2}\xi^2 v\partial_v^2 V \\ +bS\partial_S V + (\kappa(\theta - v(t)) - \lambda v)\,\partial_v V + \partial_t V - rV &= 0 \end{aligned}$$

It follows that the risk-neutral dynamics is:

$$\begin{cases} \mathrm{d}S(t) = bS(t)\,\mathrm{d}t + \sqrt{v(t)}S(t)\,\mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}v(t) = (\kappa(\theta - v(t)) - \lambda v(t))\,\mathrm{d}t + \xi\sqrt{v(t)}\,\mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

In the case of European call and put options, Heston (1993) gives a closed-form solution of the price:

$$\begin{aligned} \boldsymbol{\mathcal{C}}_0 &= S_0 e^{(b-r)T} P_1 - K e^{-rT} P_2 \\ \boldsymbol{\mathcal{P}}_0 &= S_0 e^{(b-r)T}(P_1 - 1) - K e^{-rT}(P_2 - 1) \end{aligned}$$

where the probabilities P_1 and P_2 satisfy:

$$\begin{aligned} P_j &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\phi\ln K}\varphi_j(S_0, v_0, T, \phi)}{i\phi}\right)\mathrm{d}\phi \\ \varphi_j(S_0, v_0, T, \phi) &= \exp\left(C_j(T, \phi) + D_j(T, \phi)\,v_0 + i\phi\ln S_0\right) \\ C_j(T, \phi) &= ib\phi T + \frac{a_j}{\xi^2}\left((b_j - i\rho\xi\phi + d_j)\,T - 2\ln\left(\frac{1 - g_j e^{d_j T}}{1 - g_j}\right)\right) \\ D_j(T, \phi) &= \frac{b_j - i\rho\xi\phi + d_j}{\xi^2}\left(\frac{1 - e^{d_j T}}{1 - g_j e^{d_j T}}\right) \\ g_j &= \frac{b_j - i\rho\xi\phi + d_j}{b_j - i\rho\xi\phi - d_j} \\ d_j &= \sqrt{(i\rho\xi\phi - b_j)^2 - \xi^2(2iu_j\phi - \phi^2)} \end{aligned}$$

where $a_1 = a_2 = \kappa\theta$, $b_1 = \kappa + \lambda - \rho\xi$, $b_2 = \kappa + \lambda$, $u_1 = 1/2$ and $u_2 = -1/2$.

[61] Heston (1993) makes the assumption that $\lambda_v(t) \propto \sqrt{v}$.

The existence of these semi-analytical formulas for European options is one of the main factors for explaining the popularity of the Heston model. However, the implementation of the formulas is not straightforward since it requires computing the integral of the inverse Fourier transform. In particular, Kahl and Jäckel (2005) show that the evaluation of logarithms with complex arguments may produce a numerical instability. Numerical softwares will generally do the following computation:

$$\ln\left(\frac{1-g_j e^{d_j T}}{1-g_j}\right) = \ln|r| + i\varphi$$

where:

$$r = \left|\frac{1-g_j e^{d_j T}}{1-g_j}\right|$$

and:

$$\varphi = \arg\left(\frac{1-g_j e^{d_j T}}{1-g_j}\right)$$

However, the fact that $\varphi \in [-\pi, \pi]$ will create a discontinuity when integrating the function. In order to circumvent this problem, we note:

$$g_j = r(g_j)\, e^{i\varphi(g_j)}$$

and:

$$d_j = a(d_j) + ib(d_j)$$

Kahl and Jäckel (2005) deduce that:

$$\begin{aligned} g_j - 1 &= r(g_j)\, e^{i\varphi(g_j)} - 1 \\ &= \tilde{r} e^{i(\tilde{\varphi}_j + 2\pi\tilde{m})} \end{aligned}$$

where $\tilde{m} = \left\lfloor (2\pi)^{-1} (\varphi(g_j) + \pi) \right\rfloor$, $\tilde{\varphi}_j = \arg(g_j - 1)$ and $\tilde{r} = |g_j - 1|$. They also found that:

$$\begin{aligned} g_j e^{d_j T} - 1 &= r(g_j)\, e^{i\varphi(g_j)} e^{a(d_j)T + ib(d_j)T} - 1 \\ &= r(g_j)\, e^{a(d_j)T} e^{i(\varphi(g_j) + b(d_j)T)} - 1 \\ &= \breve{r} e^{i(\breve{\varphi}_j + 2\pi\breve{m})} \end{aligned}$$

where $\breve{m} = \left\lfloor (2\pi)^{-1} (\varphi(g_j) + b(d_j)T + \pi) \right\rfloor$, $\breve{\varphi}_j = \arg\left(g_j e^{d_j T} - 1\right)$ and $\breve{r} = \left|g_j e^{d_j T} - 1\right|$. Finally, they obtain:

$$\ln\left(\frac{1-g_j e^{d_j T}}{1-g_j}\right) = \ln\frac{\breve{r}}{\tilde{r}} + i\left(\breve{\varphi}_j - \tilde{\varphi}_j + 2\pi\breve{m} - 2\pi\tilde{m}\right)$$

In Figure 9.32, we show the functions $f_1(u)$ and $f_2(u)$ defined by:

$$f_j(u) = \mathrm{Re}\left(\frac{e^{-iu\ln K}\varphi_j(S_0, v_0, T, u)}{iu}\right)$$

The parameters are $S_0 = 100$, $K = 100$, $T = 30$, $b = 0.00$, $v_0 = 0.2$, $\kappa = 1$, $\theta = 0.2$, $\xi = 0.5$ and $\lambda = 0$. For $f_1(u)$, we use $\rho = 30\%$ whereas ρ is set to -30% for the function $f_2(u)$. We see the discontinuity produced by numerical softwares. The Kahl-Jäckel method produces continuous functions without jumps. The problem can sometimes affect the two functions $f_1(u)$ and $f_2(u)$. This is the case in Figure 9.33 with the following parameters $S_0 = 100$, $K = 100$, $T = 30$, $b = 0.05$, $v_0 = 4\%$, $\kappa = 0.5$, $\theta = 4\%$, $\xi = 0.7$, $\rho = -0.80$ and $\lambda = 0$. Again, the Kahl-Jäckel method performs the good correction.

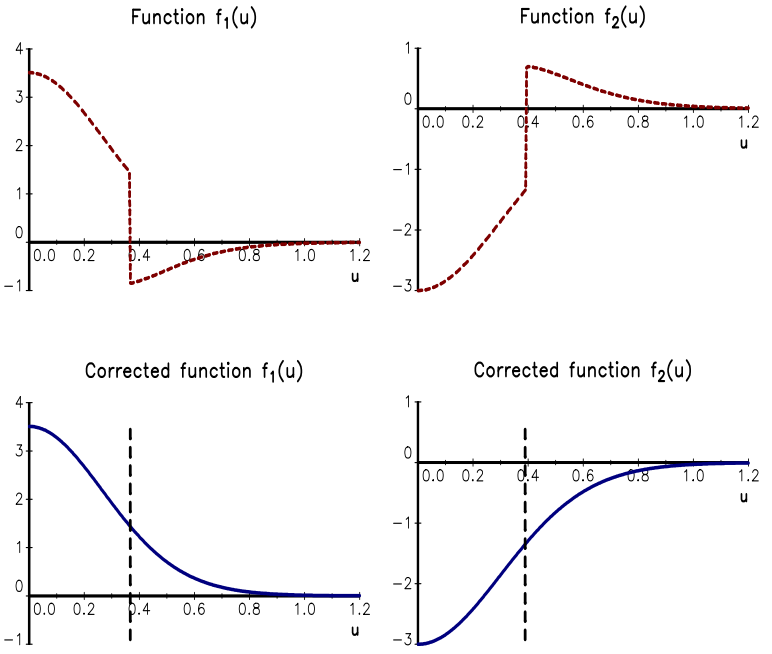


FIGURE 9.32: Functions $f_1(u)$ and $f_2(u)$ ($\kappa = 1$)

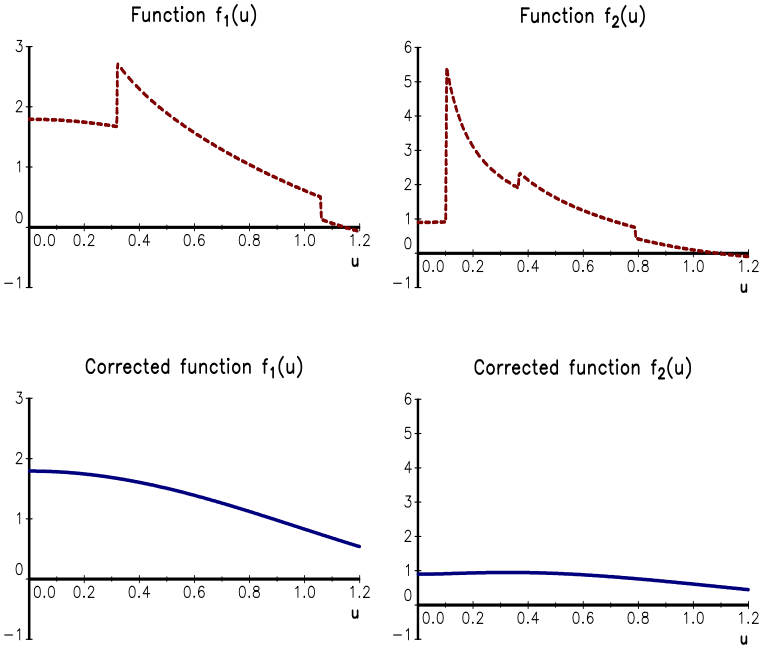


FIGURE 9.33: Functions $f_1(u)$ and $f_2(u)$ ($\kappa = 0.5$)

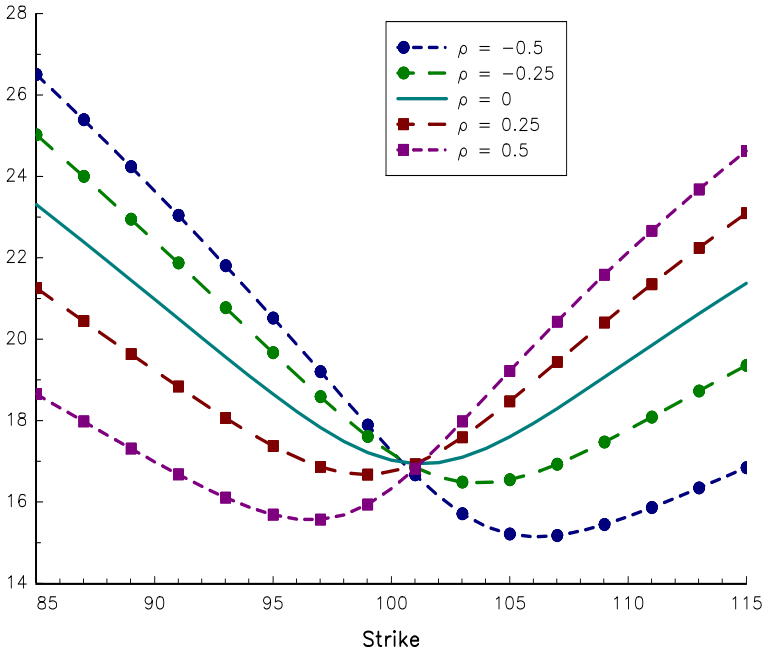


FIGURE 9.34: Implied volatility of the Heston model (in %)

Example 94 *The parameters are equal to* $S_0 = 100$, $b = r = 5\%$, $v_0 = \theta = 4\%$, $\kappa = 0.5$, $\xi = 0.9$ *and* $\lambda = 0$. *We consider the pricing of the European call option, whose maturity is three months.*

Figure 9.34 shows the implied volatility for different values of the strike K and the correlation ρ. We notice that the Heston model can produce different shapes of the volatility surface. In Figure 9.35, we have reported the skew of the implied volatility defined by:

$$\omega(T,K) = \frac{\partial \Sigma(T,K)}{\partial K}$$

Several authors have proposed approximations of the Heston implied volatility $\Sigma_t(T,K)$. We can cite Schönbucher (1999), Forde and Jacquier (2009), and Gatheral and Jacquier (2011). A more general approach has been proposed by Durrleman (2010), who assumes that the dynamics of S_t is Markovian with:

$$S(t) = S_0 \exp\left(\int_0^t \sigma(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t \sigma^2(s)\,\mathrm{d}s\right)$$

and:

$$\begin{cases} \mathrm{d}\sigma^2(t) = \mu(t)\,\mathrm{d}t - 2\sigma(t)\left(a(t)\,\mathrm{d}W(t) + \tilde{a}(t)\,\mathrm{d}\tilde{W}(t)\right) \\ \mathrm{d}\mu(t) = (\cdot)\,\mathrm{d}t + \omega(t)\,\mathrm{d}W(t) + (\cdot)\,\mathrm{d}\tilde{W}(t) \\ \mathrm{d}a(t) = m(t)\,\mathrm{d}t + u(t)\,\mathrm{d}W(t) + \tilde{u}(t)\,\mathrm{d}\tilde{W}(t) \\ \mathrm{d}\tilde{a}(t) = (\cdot)\,\mathrm{d}t + v(t)\,\mathrm{d}W(t) + (\cdot)\,\mathrm{d}\tilde{W}(t) \\ \mathrm{d}u(t) = (\cdot)\,\mathrm{d}t + x(t)\,\mathrm{d}W(t) + (\cdot)\,\mathrm{d}\tilde{W}(t) \end{cases}$$

where $(\cdot)$ is a generic symbol for a continuous adapted process. Durrleman (2010) shows that:

$$\Sigma_t^2(T,K) \simeq \sigma^2(t) + a(t)\,s(t) + \frac{b(t)\,\tau}{2} + \frac{c(t)\,s^2(t)}{2} + \frac{d(t)\,s(t)\,\tau}{2} + \frac{e(t)\,s^3(t)}{6}$$

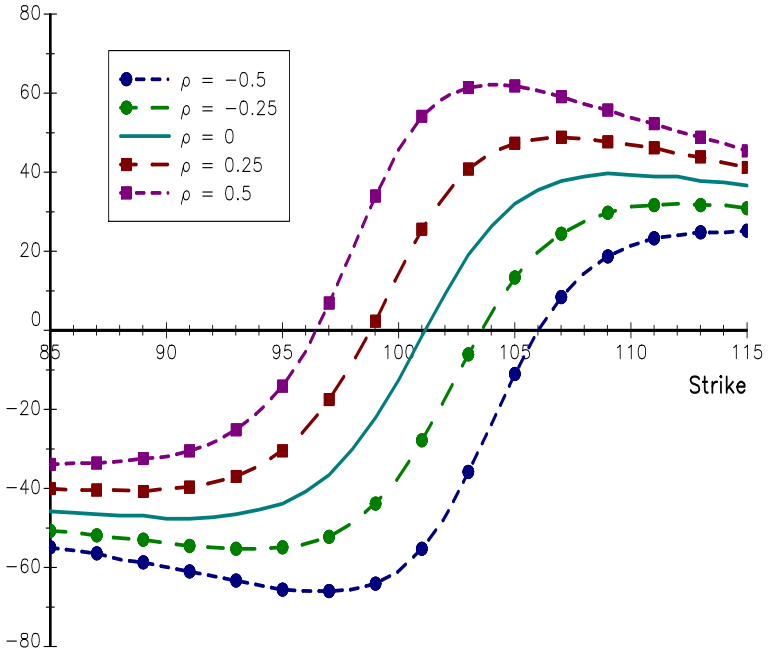


FIGURE 9.35: Skew of the Heston model (in bps)

where $\tau = T - t$ and $s(t) = \ln S(t) - \ln K$. The coefficients $b(t)$, $c(t)$, $d(t)$ and $e(t)$ are given by:

$$b(t) = \mu(t) - \frac{a^2(t)}{2} - \frac{2\tilde{a}^2(t)}{3} - a(t)\sigma^2(t) + \frac{2u(t)\sigma(t)}{3}$$

and:

$$\begin{aligned}
c(t) &= -\frac{2u(t)}{3\sigma(t)} - \frac{a^2(t)}{2\sigma^2(t)} + \frac{2\tilde{a}^2(t)}{3\sigma^2(t)} \\
d(t) &= \frac{2m(t)}{3} - \frac{\omega(t)}{3\sigma(t)} - \frac{x(t)}{2} - \frac{a(t)\mu(t)}{3\sigma^2(t)} + \frac{\tilde{a}(t)\tilde{u}(t)}{6\sigma(t)} + \frac{\tilde{a}(t)v(t)}{\sigma(t)} + \\
&\quad \frac{2a(t)\tilde{a}^2(t)}{3\sigma^2(t)} + \frac{2u(t)\sigma(t)}{3} - \frac{a^2(t)}{3} \\
e(t) &= \frac{x(t)}{2\sigma^2(t)} + \frac{2a(t)u(t)}{\sigma^3(t)} - \frac{3\tilde{a}(t)\tilde{u}(t)}{2\sigma^3(t)} - \frac{\tilde{a}(t)v(t)}{\sigma^3(t)} + \frac{3a^3(t)}{2\sigma^4(t)} - \\
&\quad \frac{4a(t)\tilde{a}^2(t)}{\sigma^4(t)}
\end{aligned}$$

In the case of the Heston model, we have:

$$\begin{cases}
\mathrm{d}S(t) = \sigma(t)S(t)\ \mathrm{d}W(t) \\
\mathrm{d}\sigma^2(t) = \kappa\left(\theta - \sigma^2(t)\right)\mathrm{d}t + \xi\sigma(t)\left(\rho\,\mathrm{d}W_t + \sqrt{1-\rho^2}\,\mathrm{d}\tilde{W}_t\right)
\end{cases}$$

It follows that $a(t) = -\frac{\xi\rho}{2}$, $\tilde{a}(t) = -\frac{\xi\sqrt{1-\rho^2}}{2}$ and $\omega(t) = m(t) = u(t) = \tilde{u}(t) = v(t) = x(t) = 0$. We deduce that:

$$\begin{aligned}
b(t) &= \kappa\left(\theta - \sigma^2(t)\right) + \frac{\xi\rho\sigma^2(t)}{2} - \frac{\xi^2}{6}\left(1 - \frac{\rho^2}{4}\right) \\
c(t) &= \frac{\xi^2}{6\sigma^2(t)}\left(1 - \frac{7\rho^2}{4}\right) \\
d(t) &= \frac{\kappa\xi\rho}{6}\left(\frac{\theta}{\sigma_t^2} - 1\right) - \frac{\xi^2\rho}{12}\left(\rho + \frac{\xi\left(1-\rho^2\right)}{\sigma^2(t)}\right) \\
e(t) &= \frac{\xi^3\rho}{2\sigma^4(t)}\left(1 - \frac{11\rho^2}{8}\right)
\end{aligned}$$

In Figure 9.36, we have generated the volatility surface using the Durrleman formula of the Heston model approximation. The parameters are $S(t) = 100$, $\sigma(t) = 20\%$, $\kappa = 0.5$, $\theta = 4\%$ and $\xi = 0.2$. We consider different values for the correlation parameter ρ and the maturity T. We notice that the Durrleman formula does not fit correctly the Heston smile when the absolute value $|\rho|$ of the correlation is high.

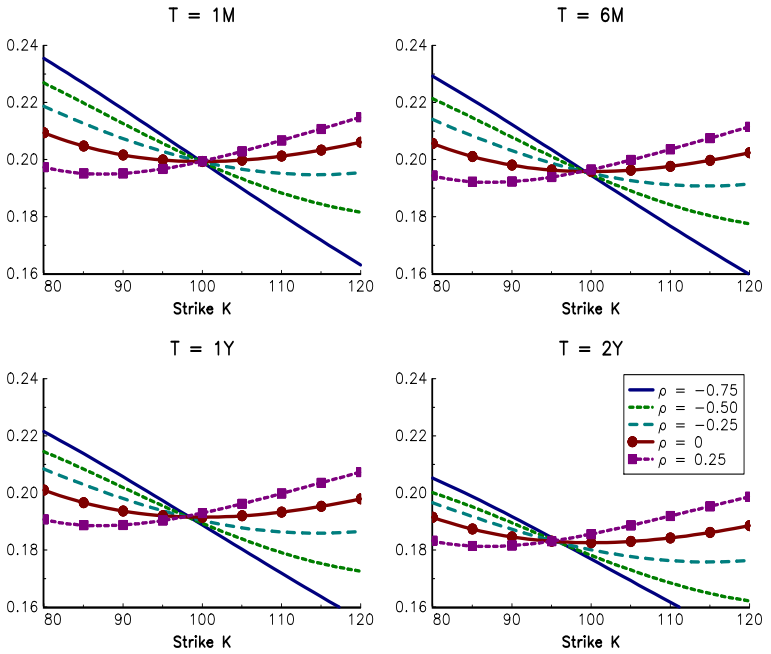


FIGURE 9.36: Implied volatility of the Durrleman formula (in %)

Example 95 *We assume that* $S(t) = 100$ *and* $T = 0.5$. *The volatility smile is given by the following values:*

K	90.00	95.00	100.00	105.00	110.00
$\Sigma_t(T,K)$ *(in %)*	20.25	19.92	19.67	19.49	19.38

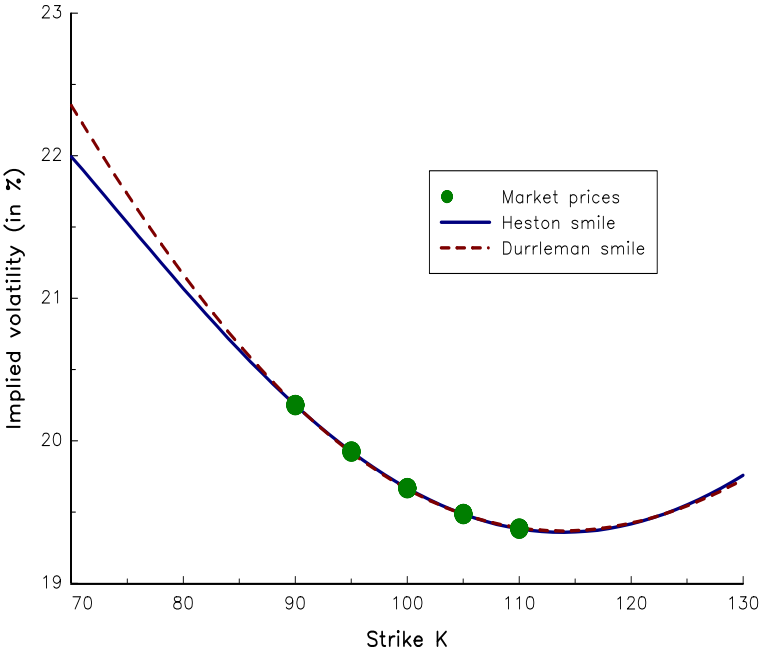


FIGURE 9.37: Calibration of the smile by the Heston model and the Durrleman formula

The calibration of the smile gives the following result[62]:

Model	$\sigma(t)$	κ	θ	ξ	ρ
Heston	0.201	0.980	0.040	0.192	−0.207
Durrleman	0.222	1.000	0.014	0.191	−0.193

The volatility surface of each calibrated model is represented in Figure 9.37. The results are very similar.

Remark 107 *The Heston model was very popular in the 2000s. Nevertheless, even if we have an analytical formula for the call and put prices, the absence of a true implied volatility formula was an obstacle of its development, and the use of the Heston model is today less frequent. The Heston model has then been replaced by the SABR model, because of the availability of an implied volatility formula.*

9.2.4.3 SABR model

Hagan *et al.* (2002) suggest using the SABR[63] model to take into account the smile effect. The dynamics of the forward rate $F(t)$ is given by:

$$\begin{cases} \mathrm{d}F(t) = \alpha(t) F(t)^{\beta} \mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}\alpha(t) = \nu\alpha(t) \mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

where $\mathbb{E}\left[W_1^{\mathbb{Q}}(t) W_2^{\mathbb{Q}}(t)\right] = \rho t$. Since $\beta \in [0,1]$, $\alpha(t)$ is not necessarily the instantaneous volatility of $F(t)$ except in the cases $\beta = 0$ (Gaussian volatility) and $\beta = 1$ (log-normal

[62] It consists of minimizing the sum of squared errors between observed implied volatilities and theoretical implied volatilities deduced from the option model.

[63] This is the acronym of stochastic $-\alpha-\beta-\rho$.

volatility). The model has also 4 parameters: α the current value of $\alpha(t)$, β the exponent of the forward rate, ν the log-normal volatility of $\alpha(t)$ and ρ the correlation between the two Brownian motions. One of the big interests of the SABR model is that we have an approximate formula of the implied Black volatility:

$$\begin{aligned}\Sigma_B(T,K) &= \frac{\alpha}{(F_0K)^{(1-\beta)/2}\left(1+\frac{(1-\beta)^2}{24}\ln^2\frac{F_0}{K}+\frac{(1-\beta)^4}{1920}\ln^4\frac{F_0}{K}\right)}\left(\frac{z}{\chi(z)}\right)\cdot \\ &\left(1+\left(\frac{(1-\beta)^2\alpha^2}{24(F_0K)^{1-\beta}}+\frac{\rho\alpha\nu\beta}{4(F_0K)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\nu^2\right)T\right)\end{aligned}$$

where $z=\nu\alpha^{-1}(F_0K)^{(1-\beta)/2}\ln\dfrac{F_0}{K}$ and $\chi(z)=\ln\left(\sqrt{1-2\rho z+z^2}+z-\rho\right)-\ln(1-\rho)$.

Let us see the interpretation of the parameters[64]. We have represented their impact in Figures[65] 9.38 and 9.39. The parameter β allows to define a stochastic log-normal model when β is equal to 1, or a stochastic normal model when β is equal to 0, or an hybrid model. The choice of β is generally exogenous. The main reason is that β is highly related to the dynamics of the ATM implied volatility. If β is equal to 1, we observe a simple translation of the smile when the forward rate moves (first panel in Figure 9.38). If β is equal to 0, the ATM implied volatility decreases when the forward rates increases (second panel in Figure 9.38). This explains the behavior of the backbone, which represents the dynamics of the ATM implied volatility when the forward rate varies (third panel in Figure 9.38).

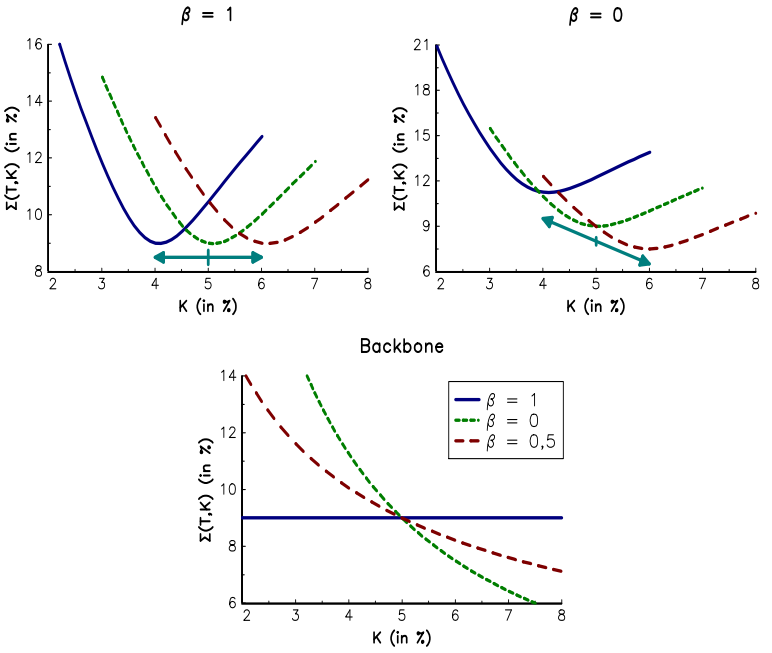


FIGURE 9.38: Impact of the parameter β

[64] In the following examples, we consider a one-year option, whose current forward rate F_0 is equal to 5%.
[65] The default values are $\alpha = 10\%$, $\beta = 1$, $\nu = 50\%$ and $\rho = 0$.

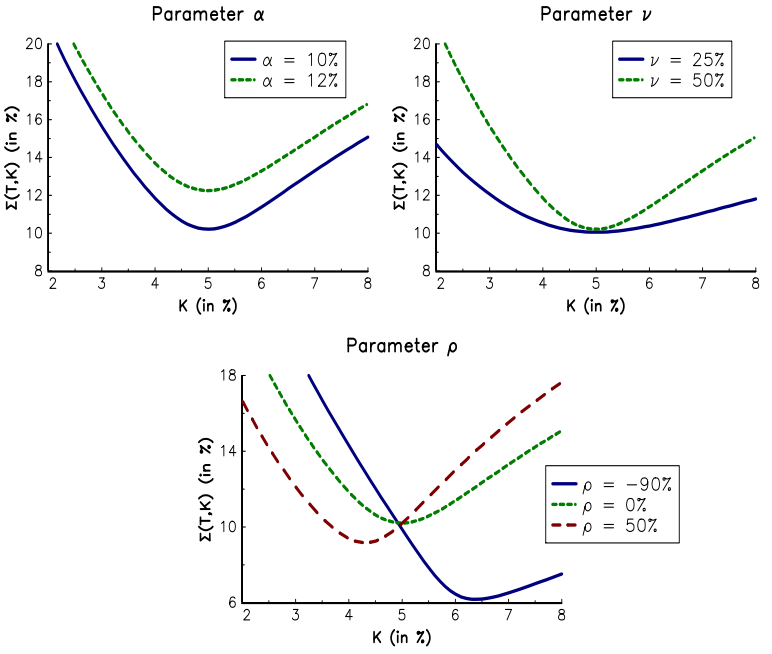


FIGURE 9.39: Impact of the parameters α, ν and ρ

The parameter α controls the level of implied volatilities (see Panel 1 in Figure 9.39). In particular, α is close to the value of the ATM volatility when β is equal to one[66]. ν is called the vovol (or vol-vol) parameter, because it measures the volatility of the volatility. ν impacts then the stochastic property of the volatility $\alpha(t)$. The limit case $\nu = 0$ corresponds to the constant volatility and we obtain the classical Black model[67]. An increase of ν tends to increase the slope of the implied volatility (see Panel 2 in Figure 9.39). The asymmetry of the smile is due to the parameter ρ. For instance, if ρ is negative, the skew is more important in the left side than in the right side (see Panel 3 in Figure 9.39).

Remark 108 *The parameters β and ρ impact the slope of the smile in a similar way. Then, they cannot be jointly identifiable. For example, let us consider the following smile when F_0 is equal to 5%: $\Sigma_B(1,3\%) = 13\%$, $\Sigma_B(1,4\%) = 10\%$, $\Sigma_B(1,5\%) = 9\%$ and $\Sigma_B(1,7\%) = 10\%$. If we calibrate this smile for different values of β, we obtain the following solutions:*

β	α	ν	ρ
0.0	0.0044	0.3203	0.2106
0.5	0.0197	0.3244	0.0248
1.0	0.0878	0.3388	−0.1552

We have represented the corresponding smiles in Figure 9.40 and we verify that the three sets of calibrated parameters give the same smile.

[66] In this case, we have:

$$\Sigma_B(T, F_0) = \alpha\left(1 + \left(\frac{\rho\alpha\nu}{4} + \frac{2-3\rho^2}{24}\nu^2\right)T\right)$$

It follows that $\Sigma_B(T, F_0)$ is exactly equal to α when ρ is equal to zero.

[67] When β is equal to one of course.

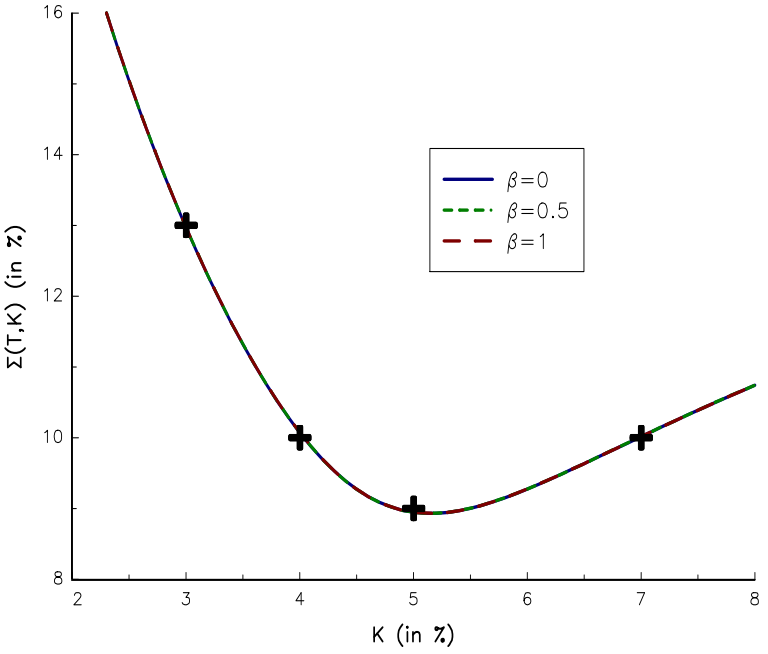


FIGURE 9.40: Implied volatility for different parameter sets (β, ρ)

We have seen that the choice of β is not important for calibrating the SABR model for a given maturity. We have already seen that the parameter β has a great impact on the dynamics on the backbone. Therefore, there are two approaches for estimating β:

1. β can be chosen from prior beliefs ($\beta = 0$ for the normal model, $\beta = 0.5$ for the CIR model and $\beta = 1$ for the log-normal model);
2. β can be statistically estimated by considering the dynamics of the forward rate.

TABLE 9.12: Calibration of the parameter β in the SABR model

Rate	Level		Difference		Empirical quantile of $\hat{\beta}_{t,t+h}$				
	$\hat{\beta}$	R_c^2	$\hat{\beta}$	R_c^2	10%	25%	50%	75%	90%
1y1y	−0.06	0.91	0.59	0.15	−2.01	−0.14	0.71	1.00	2.17
1y5y	−0.29	0.87	0.32	0.27	−1.80	−0.28	0.73	1.11	2.76
1y10y	−0.37	0.80	0.34	0.22	−2.04	−0.23	0.71	1.11	2.69
5y1y	0.42	0.29	0.35	0.22	−1.58	−0.31	0.71	1.00	2.38
5y5y	−0.01	0.73	0.23	0.28	−2.12	−0.36	0.61	1.00	2.52
5y10y	−0.10	0.69	0.27	0.23	−1.99	−0.30	0.70	1.05	2.58
10y1y	0.96	0.00	0.28	0.20	−1.88	−0.20	0.80	1.07	2.43
10y5y	−0.10	0.65	0.28	0.20	−2.02	−0.29	0.73	1.02	2.76
10y10y	−0.47	0.73	0.27	0.20	−1.71	−0.24	0.85	1.07	2.93

The second approach is based on the approximation of the ATM volatility:

$$\Sigma_t(T, F_t) \simeq \frac{\alpha}{F_t^{1-\beta}}$$

We have:

$$\ln \Sigma_t(T, F_t) = \ln \alpha + (\beta - 1) \ln F_t + u_t \tag{9.36}$$

We can then estimate β by considering the linear regression of the logarithm of the ATM volatility on the logarithm of the forward rate. However, these two variables are generally integrated of order one or $I(1)$. A better approach is then to consider the alternative linear regression[68]:

$$\ln \Sigma_{t+h}(T, F_{t+h}) - \ln \Sigma_t(T, F_t) = c + (\beta - 1)(\ln F_{t+h} - \ln F_t) + u_t \tag{9.37}$$

where c is a constant. In this case, the linear regression is performed using the difference and not the level of implied volatilities. Using the Libor EUR rates between 2000 and 2003, we obtain results given in Table 9.12. In the first column, we indicate the maturity and the tenor of the forward rate. The next two columns report the estimate $\hat{\beta}$ and the R-squared coefficient R_c^2 for the regression model (9.36). Then, we have the values of $\hat{\beta}$ and R_c^2 for the regression model[69] (9.37). We observe some strong differences between the two approaches (see also the probability density function of $\hat{\beta}$ in Figure 9.41). These results show that the regression model (9.36) produces bad results. However, it does not mean that the second regression model (9.36) is more robust. Indeed, we can calculate the exact value $\hat{\beta}_{t,t+h}$ that explains the dynamics of the ATM volatility from time t to time $t+h$:

$$\hat{\beta}_{t,t+h} = \frac{\ln\left(F_{t+h} \cdot \Sigma_{t+h}(T, F_{t+h})\right) - \ln\left(F_t \cdot \Sigma_t(T, F_t)\right)}{\ln F_{t+h} - \ln F_t}$$

In Table 9.12, we notice the wide dispersion of $\hat{\beta}_{t,t+h}$. On average, the parameter β is around 70%, but it can also take some large negative or positive values. This is why β is generally chosen from prior beliefs.

Once we have set the value of β, we estimate the parameters (α, ν, ρ) by fitting the observed implied volatilities. However, we have seen that α is highly related to the ATM volatility. Indeed, we have:

$$\Sigma_B(T, F_0) = \frac{\alpha}{F_0^{1-\beta}}\left(1 + \left(\frac{(1-\beta)^2\alpha^2}{24F_0^{2-2\beta}} + \frac{\rho\alpha\nu\beta}{4F_0^{1-\beta}} + \frac{2-3\rho^2}{24}\nu^2\right)T\right)$$

We deduce that:

$$\alpha^3\left(\frac{(1-\beta)^2 T}{24F_0^{2-2\beta}}\right) + \alpha^2\left(\frac{\rho\nu\beta T}{4F_0^{1-\beta}}\right) + \alpha\left(1 + \frac{2-3\rho^2}{24}\nu^2 T\right) - \Sigma_B(T, F_0)\, F_0^{1-\beta} = 0$$

Let $\alpha = g_\alpha(\Sigma_B(T, F_0), \nu, \rho)$ be the positive root of the cubic equation. Therefore, imposing that the smile passes through the ATM volatility $\Sigma_B(T, F_0)$ allows to reduce the calibration to two parameters (ν, ρ).

Example 96 *We consider the following smile:*

K *(in %)*	2.8	3.0	3.5	3.7	4.0	4.5	5.0	7.0
$\Sigma(T, K)$ *(in %)*	13.2	12.8	12.0	11.6	11.0	10.0	9.0	10.0

The maturity T is equal to one year and the forward rate F_0 is set to 5%.

[68] We have:

$$\frac{\Sigma_{t+h}(T, F_{t+h})}{\Sigma_t(T, F_t)} \simeq \left(\frac{F_{t+h}}{F_t}\right)^{\beta-1}$$

[69] In this case, we set h to one trading day.

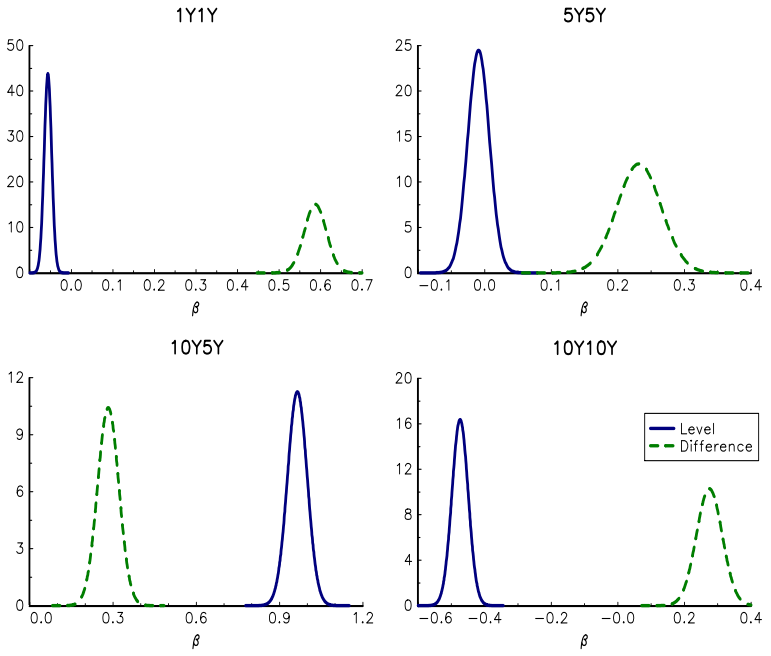


FIGURE 9.41: Probability density function of the estimate $\hat{\beta}$ (SABR model)

If we consider a stochastic log-normal model ($\beta = 1$), we obtain the following results:

Calibration	α (in %)	β	ν	ρ (in %)	RSS	Σ_{ATM} (in %)
#1	9.466	1.00	0.279	−23.70	0.630	9.51
#2	8.944	1.00	0.322	−22.90	1.222	9.00

RSS indicates the residual sum of squares (expressed in bps). In the first calibration, we estimate the three parameters α, ν and ρ. In this case, the residual sum of squares is equal to 0.63 bps, but the SABR ATM volatility is equal to 9.51%, which is far from the market ATM volatility. In the second calibration, we estimate the two parameters ν and ρ, whereas α is the solution of the cubic equation that fits the ATM volatility. We notice that the residual sum of squares has increased from 0.63 bps to 1.222 bps, but the SABR ATM volatility is exactly equal to the market ATM volatility. The two calibrated smiles are reported in Figure 9.42.

Remark 109 *One of the issues with implied volatility calibration is that we generally have more market prices for the put (or left) wing of the smile than its call (or right) wing. This implies that the put wing is better calibrated than the call wing, and we may observe a large difference between the calibrated ATM volatility and the market ATM volatility. Therefore, professionals prefer the second calibration.*

The sensitivities correspond to the following formulas[70]:

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial F_0} + \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma} \cdot \frac{\partial \Sigma_B(T,K)}{\partial F_0}$$

[70] If we consider the parametrization $\alpha = g_\alpha(\Sigma_{\text{ATM}}, \nu, \rho)$, we have:

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial F_0} + \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma} \cdot \left(\frac{\partial \Sigma_B(T,K)}{\partial F_0} + \frac{\partial \Sigma_B(T,K)}{\partial \alpha} \cdot \frac{\partial g_\alpha(\Sigma_{\text{ATM}}, \nu, \rho)}{\partial F_0} \right)$$

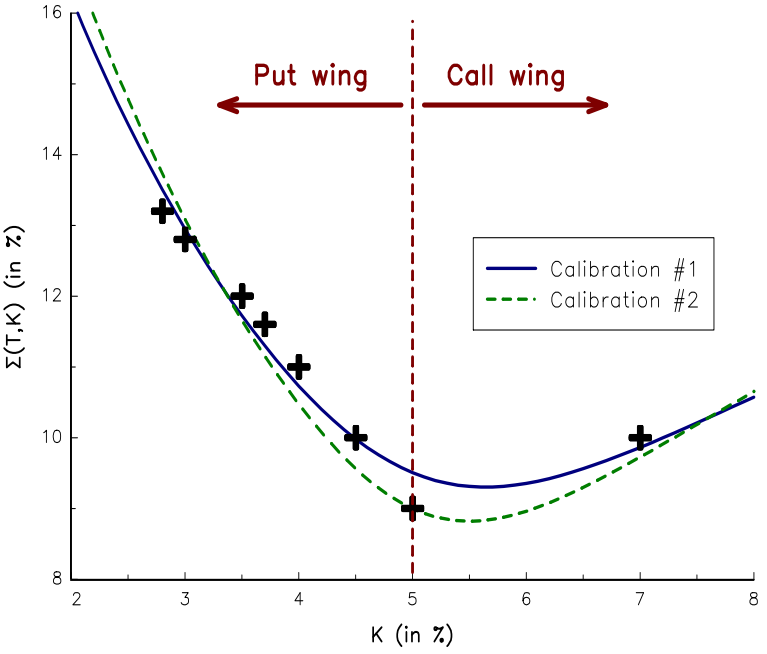


FIGURE 9.42: Calibration of the SABR model

and:

$$\boldsymbol{v} = \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma} \cdot \frac{\partial \Sigma_B(T,K)}{\partial \alpha}$$

To obtain these formulas, we apply the chain rule on the Black formula by assuming that the volatility Σ is not constant and depends on F_0 and α.

Remark 110 *We notice that the vega is defined with respect to the parameter α. This approach is little used in practice, because it is difficult to hedge this model parameter. This is why traders prefer to compute the vega with respect to the ATM volatility:*

$$\boldsymbol{v} = \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma} \cdot \frac{\partial \Sigma_B(T,K)}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \Sigma_{\text{ATM}}}$$

where $\Sigma_{\text{ATM}} = \Sigma_B(T, F_0)$.

Remark 111 *Bartlett (2006) proposes a refinement for computing the delta. Indeed, a shift in F_0 produces a shift in α, because the two processes $F(t)$ and $\alpha(t)$ are correlated. Since we have:*

$$\begin{aligned} \mathrm{d}\alpha(t) &= \nu\alpha(t)\,\mathrm{d}W_2^Q(t) \\ &= \nu\alpha(t)\left(\rho\,\mathrm{d}W_1^Q(t) + \sqrt{1-\rho^2}\,\mathrm{d}W(t)\right) \end{aligned}$$

and:

$$\mathrm{d}W_1^Q(t) = \frac{\mathrm{d}F(t)}{\alpha(t)\,F(t)^\beta}$$

we deduce that:

$$\mathrm{d}\alpha(t) = \frac{\nu\rho}{F(t)^\beta}\,\mathrm{d}F(t) + \nu\alpha(t)\sqrt{1-\rho^2}\,\mathrm{d}W(t)$$

The new delta is then:

$$\begin{aligned}
\boldsymbol{\Delta}^{\star} &= \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial F_0} + \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma}\left(\frac{\partial \Sigma_B(T,K)}{\partial F_0} + \frac{\partial \Sigma_B(T,K)}{\partial \alpha}\cdot\frac{\partial \alpha}{\partial F_0}\right) \\
&= \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial F_0} + \frac{\partial \boldsymbol{\mathcal{C}}_B}{\partial \Sigma}\left(\frac{\partial \Sigma_B(T,K)}{\partial F_0} + \frac{\nu\rho}{F(t)^{\beta}}\frac{\partial \Sigma_B(T,K)}{\partial \alpha}\right) \\
&= \boldsymbol{\Delta} + \frac{\nu\rho}{F(t)^{\beta}}\boldsymbol{\upsilon}
\end{aligned}$$

Therefore, this approach is particularly useful when we consider a delta hedging instead of a delta-vega hedging, since the new delta risk incorporates a part of the vega risk.

9.2.5 Factor models

Factor models are extensively used for modeling fixed income derivatives (Vasicek, CIR, HJM, etc.). They assume that interest rates are linked to some factors $X(t)$, which can be observable or not observable. For instance, the factor is directly the instantaneous interest rate $r(t)$ in Vasicek or CIR models. However, a one-factor model is generally limited and is not enough rich to fit the yield curve and the basic asset prices (caplets and swaptions). During a long time, academics have developed multi-factor models by considering explicit factors (level, slope, convexity, etc.). For instance, Brennan and Schwartz (1979) consider the short-term interest rate and the long-term interest rate, whereas Longstaff and Schwartz (1992) use the short-term interest rate and its volatility. Today, this type of approach is outdated and is replaced by a more pragmatic approach based on non-explicit factors.

9.2.5.1 Linear and quadratic Gaussian models

Let us assume that the instantaneous interest rate $r(t)$ is linked to the factors $X(t)$ under the risk-neutral probability $\mathbb{Q}$ as follows:

$$r(t) = \alpha(t) + \beta(t)^{\top} X(t) + X(t)^{\top} \Gamma(t) X(t)$$

where $\alpha(t)$ is a scalar, $\beta(t)$ is a $n \times 1$ vector and $\Gamma(t)$ is a $n \times n$ matrix. This parametrization encompasses different specific cases: one-factor model, affine model and quadratic model[71]. We also assume that the factors follow an Ornstein-Uhlenbeck process:

$$\mathrm{d}X(t) = (a(t) + B(t) X(t))\,\mathrm{d}t + \Sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

where $a(t)$ is a $n \times 1$ vector, $B(t)$ is a $n \times n$ matrix, $\Sigma(t)$ is a $n \times n$ matrix and $W^{\mathbb{Q}}(t)$ is a standard n-dimensional Brownian motion.

El Karoui *et al.* (1992a) show that there exists a family of $\hat{\alpha}(t,T)$, $\hat{\beta}(t,T)$ and $\hat{\Gamma}(t,T)$ such that the price of the zero-coupon bond $B(t,T)$ is given by:

$$B(t,T) = \exp\left(-\hat{\alpha}(t,T) - \hat{\beta}(t,T)^{\top} X(t) - X(t)^{\top} \hat{\Gamma}(t,T) X(t)\right)$$

[71] As shown by Filipović (2002), it is not necessary to use higher order because the only consistent polynomial term structure approaches are the affine and quadratic term structure models.

where $\hat{\alpha}(t,T)$, $\hat{\beta}(t,T)$ and $\hat{\Gamma}(t,T)$ solve a system of Riccati equations. If we assume that the matrix $\hat{\Gamma}(t,T)$ is symmetric, we obtain:[72]:

$$\begin{aligned}
\partial_t \hat{\alpha}(t,T) &= -\operatorname{tr}\left(\Sigma(t)\Sigma(t)^\top \hat{\Gamma}(t,T)\right) - \hat{\beta}(t,T)^\top a(t) + \\
&\quad \frac{1}{2}\hat{\beta}(t,T)^\top \Sigma(t)\Sigma(t)^\top \hat{\beta}(t,T) - \alpha(t) \\
\partial_t \hat{\beta}(t,T) &= -B(t)^\top \hat{\beta}(t,T) + 2\hat{\Gamma}(t,T)\Sigma(t)\Sigma(t)^\top \hat{\beta}(t,T) - \\
&\quad 2\hat{\Gamma}(t,T)a(t) - \beta(t) \\
\partial_t \hat{\Gamma}(t,T) &= 2\hat{\Gamma}(t,T)\Sigma(t)\Sigma(t)^\top \hat{\Gamma}(t,T) - \\
&\quad 2\hat{\Gamma}(t,T)B(t) - \Gamma(t)
\end{aligned}$$

with the boundary conditions $\hat{\alpha}(T,T) = \hat{\beta}(T,T) = \hat{\Gamma}(T,T) = \mathbf{0}$. We notice that the expression of the forward interest rate $F(t,T_1,T_2)$ is given by:

$$\begin{aligned}
F(t,T_1,T_2) &= -\frac{1}{T_2-T_1}\ln\frac{B(t,T_2)}{B(t,T_1)} \\
&= \frac{\hat{\alpha}(t,T_2) - \hat{\alpha}(t,T_1) + \left(\hat{\beta}(t,T_2) - \hat{\beta}(t,T_2)\right)^\top X(t)}{T_2-T_1} + \\
&\quad \frac{X(t)^\top \left(\hat{\Gamma}(t,T_2) - \hat{\Gamma}(t,T_1)\right) X(t)}{T_2-T_1}
\end{aligned}$$

We deduce that the instantaneous forward rate is equal to:

$$f(t,T) = \alpha(t,T) + \beta(t,T)^\top X(t) + X(t)^\top \Gamma(t,T) X(t)$$

where $\alpha(t,T) = \partial_T \hat{\alpha}(t,T)$, $\beta(t,T) = \partial_T \hat{\beta}(t,T)$ and $\Gamma(t,T) = \partial_T \hat{\Gamma}(t,T)$. It follows that $\alpha(t) = \alpha(t,t) = \partial_t \hat{\alpha}(t,t)$, $\beta(t) = \beta(t,t) = \partial_t \hat{\beta}(t,t)$ and $\Gamma(t) = \Gamma(t,t) = \partial_t \hat{\Gamma}(t,t)$.

Let $V(t,X)$ be the price of the option, whose payoff is $f(x)$. It satisfies the following PDE:

$$\begin{aligned}
&\frac{1}{2}\operatorname{trace}\left(\Sigma(t)\partial_X^2 V(t,X)\Sigma(t)^\top\right) + (a(t) + B(t)X)\partial_X V(t,X) + \\
&\qquad \partial_t V(t,X) - \left(\alpha(t) + \beta(t)^\top X + X^\top \Gamma(t) X\right) V(t,X) = 0
\end{aligned} \tag{9.38}$$

Once we have specified the functions $\alpha(t)$, $\beta(t)$, $\Gamma(t)$, $a(t)$, $B(t)$ and $\Sigma(t)$, we can then price the option by solving numerically the previous multidimensional PDE with the terminal condition $V(T,X) = f(X)$. Most of the time, the payoff is not specified with respect to the state variables X, but depends on the interest rate $r(t)$. In this case, we use the following transformation:

$$f(r) = f\left(\alpha(T) + \beta(T)^\top X + X^\top \Gamma(T) X\right)$$

Remark 112 *We can also calculate the price of the option by Monte Carlo methods. This approach is generally more efficient when the number of factors is larger than* 2.

[72] See Exercise 9.4.10 on page 601 and Ahn *et al.* (2002) for the derivation of the Riccati equations.

9.2.5.2 Dynamics of risk factors under the forward probability measure

We have:

$$\frac{\mathrm{d}B(t,T)}{B(t,T)} = r(t)\,\mathrm{d}t - \left(2\hat{\Gamma}(t,T)\,X(t) + \hat{\beta}(t,T)\right)^{\top}\Sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

We deduce that:

$$W^{\mathbb{Q}^{\star}(T)}(t) = W^{\mathbb{Q}}(t) + \int_0^t \Sigma(s)^{\top}\left(2\hat{\Gamma}(s,T)\,X(s) + \hat{\beta}(s,T)\right)\mathrm{d}s$$

defines a Brownian motion under $\mathbb{Q}^{\star}(T)$. It follows that:

$$\mathrm{d}X(t) = \left(\tilde{a}(t) + \tilde{B}(t)\,X(t)\right)\mathrm{d}t + \Sigma(t)\,\mathrm{d}W^{\mathbb{Q}^{\star}(T)}(t)$$

where:

$$\tilde{a}(t) = a(t) - \Sigma(t)\,\Sigma(t)^{\top}\,\hat{\beta}(t,T)$$

and:

$$\tilde{B}(t) = B(t) - 2\Sigma(t)\,\Sigma(t)^{\top}\,\hat{\Gamma}(t,T)$$

We conclude that $X(t)$ is Gaussian under any forward probability measure $\mathbb{Q}^{\star}(T)$:

$$X(t) \sim \mathcal{N}(m(0,t), V(0,t))$$

El Karoui *et al.* (1992a) show that the conditional mean and variance satisfies the following forward differential equations:

$$\begin{cases} \partial_T m(t,T) &= a(T) + B(T)\,m(t,T) - 2V(t,T)\,\Gamma(T)\,m(t,T) - \\ & \quad V(t,T)\,\beta(T) \\ \partial_T V(t,T) &= V(t,T)\,B(T)^{\top} + B(T)\,V(t,T) - 2V(t,T)\,\Gamma(T)\,V(t,T) + \\ & \quad \Sigma(T)\,\Sigma(T)^{\top} \end{cases}$$

If t is equal to zero, the initial conditions are $m(0,0) = X(0) = \mathbf{0}$ and $V(0,0) = \mathbf{0}$. If $t \neq 0$, we proceed in two steps: first, we calculate numerically the solutions $m(0,t)$ and $V(0,t)$, and second, we initialize the system with $m(t,t) = m(0,t)$ and $V(t,t) = V(0,t)$.

Remark 113 *In fact, the previous forward differential equations are not obtained under the traditional forward probability measure $\mathbb{Q}^{\star}(T)$, but under the probability measure $\mathbb{Q}^{\star}(t,T)$ defined by the following Radon-Nykodin derivative:*

$$\frac{\mathrm{d}\mathbb{Q}^{\star}(t,T)}{\mathrm{d}\mathbb{P}} = e^{-\int_0^T r(s)\,\mathrm{d}s} e^{\int_t^T f(t,s)\,\mathrm{d}s}$$

The reason is that we would like to price at time t any caplet with maturity T. Therefore, this is the maturity T and not the filtration $\mathcal{F}_t$ that moves.

9.2.5.3 Pricing caplets and swaptions

We reiterate that the formula of the Libor rate $L(t,T_{i-1},T_i)$ at time t between the dates T_{i-1} and T_i is:

$$L(t,T_{i-1},T_i) = \frac{1}{T_i - T_{i-1}}\left(\frac{B(t,T_{i-1})}{B(t,T_i)} - 1\right)$$

It follows that the price of the caplet is given by:

$$\text{Caplet} = B(0,t)\,\mathbb{E}^{\mathbb{Q}^{\star}(t)}\left[\left(B(t,T_{i-1}) - (1 + (T_i - T_{i-1})\,K)\,B(t,T_i)\right)^{+}\right]$$

where $\mathbb{Q}^{\star}(t)$ is the forward probability measure. We can then calculate the price using two approaches:

1. we can solve the partial differential equation;
2. we can calculate the mathematical expectation using numerical integration.

In the first approach, we consider the PDE (9.38) with the following payoff:

$$f(X) = \max(0, g(X))$$

where:

$$\begin{aligned} g(X) &= \exp\left(-\hat{\alpha}(t, T_{i-1}) - \hat{\beta}(t, T_{i-1})^\top X - X^\top \hat{\Gamma}(t, T_{i-1}) X\right) - \\ &\quad (1 + \delta_{i-1} K) \exp\left(-\hat{\alpha}(t, T_i) - \hat{\beta}(t, T_i)^\top X - X^\top \hat{\Gamma}(t, T_i) X\right) \end{aligned}$$

In the second approach, we have $X(t) \sim \mathcal{N}(m(0,t), V(0,t))$ under the forward probability $\mathbb{Q}^\star(t)$. We deduce that:

$$\text{Caplet}(t, T_{i-1}, T_i) = B(0,t) \int f(x) \phi_n(x; m(0,t), V(0,t)) \, \mathrm{d}x$$

This integral can be computed numerically using Gauss-Legendre quadrature methods.

For the swaption, the payoff is:

$$\begin{aligned} f(X) &= (\text{Sw}(T_0) - K)^+ \sum_{i=1}^{n} (T_i - T_{i-1}) B(T_0, T_i) \\ &= \left(B(T_0, T_0) - B(T_0, T_n) - K \sum_{i=1}^{n} (T_i - T_{i-1}) B(T_0, T_i)\right)^+ \\ &= \max(0, g(X)) \end{aligned}$$

where:

$$\begin{aligned} g(X) &= \exp\left(-\hat{\alpha}(T_0, T_0) - \hat{\beta}(T_0, T_0)^\top X - X^\top \hat{\Gamma}(T_0, T_0) X\right) - \\ &\quad \exp\left(-\hat{\alpha}(T_0, T_n) - \hat{\beta}(T_0, T_n)^\top X - X^\top \hat{\Gamma}(T_0, T_n) X\right) - \\ &\quad K \sum_{i=1}^{n} \delta_{i-1} \exp\left(-\hat{\alpha}(T_0, T_i) - \hat{\beta}(T_0, T_i)^\top X - X^\top \hat{\Gamma}(T_0, T_i) X\right) \end{aligned}$$

As previously, we can price the swaption by solving the PDE with the payoff $f(X)$ or by calculating the following integral:

$$\text{Swaption} = B(0, T_0) \int f(x) \phi_n(x; m(0, T_0), V(0, T_0)) \, \mathrm{d}x$$

9.2.5.4 Calibration and practice of factor models

The calibration of the model consists in fitting the functions $\alpha(t)$, $\beta(t)$, $\Gamma(t)$, $a(t)$, $B(t)$ and $\Sigma(t)$. Generally, professionals assume that $a(t) = 0$ and $B(t) = \mathbf{0}$. Indeed, if we consider the following transformation:

$$\tilde{X}(t) = e^{-\int_0^t B(s)\,\mathrm{d}s} X(t) - \int_0^t a(s) e^{-\int_0^s B(u)\,\mathrm{d}u} \, \mathrm{d}s$$

we obtain:

$$\begin{aligned}\mathrm{d}\tilde{X}(t) &= e^{-\int_0^t B(s)\,\mathrm{d}s}\Sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)\\ &= \tilde{\Sigma}(t)\,\mathrm{d}W(t)\end{aligned}$$

Without loss of generality, we can then set $\mathrm{d}X(t) = \Sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$, and the Riccati equations are simplified as follows:

$$\begin{cases}\partial_t\hat{\alpha}(t,T) = -\operatorname{tr}\left(\Sigma(t)\Sigma(t)^\top\hat{\Gamma}(t,T)\right) + \frac{1}{2}\hat{\beta}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\beta}(t,T) - \alpha(t)\\ \partial_t\hat{\beta}(t,T) = 2\hat{\Gamma}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\beta}(t,T) - \beta(t)\\ \partial_t\hat{\Gamma}(t,T) = 2\hat{\Gamma}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\Gamma}(t,T) - \Gamma(t)\end{cases}$$

If we consider an affine model, we retrieve the formula of Duffie and Huang (1996):

$$B(t,T) = \exp\left(-\hat{\alpha}(t,T) - \hat{\beta}(t,T)^\top X(t)\right)$$

where[73]:

$$\begin{cases}\partial_t\hat{\alpha}(t,T) = \frac{1}{2}\hat{\beta}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\beta}(t,T) - \alpha(t)\\ \partial_t\hat{\beta}(t,T) = -\beta(t)\end{cases}$$

First, we must fit the initial yield curve, which is noted $B(0,T)$. If we assume that $X(0) = \mathbf{0}$, we obtain:

$$\hat{\alpha}(t,T) = -\ln\frac{B(0,T)}{B(0,t)}$$

We notice that the computation of $\hat{\alpha}(t,T)$ allows to define $\alpha(t)$:

$$\alpha(t) = -\operatorname{tr}\left(\Sigma(t)\Sigma(t)^\top\hat{\Gamma}(t,T)\right) + \frac{1}{2}\hat{\beta}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\beta}(t,T) - \partial_t\hat{\alpha}(t,T)$$

because $\partial_t\hat{\alpha}(t,T)$ can be calculated using finite differences. Therefore, the problem dimension is reduced and the calibration depends on $\beta(t)$, $\Gamma(t)$ and $\Sigma(t)$. In order to calibrate these functions, we need to fit other products like caplets and swaptions. We have shown that these products can be priced using numerical integration. Therefore, the calibration of $\beta(t)$, $\Gamma(t)$ and $\Sigma(t)$ can be done without solving the PDE, which is time-consuming.

Let us now see what type of volatility smile is generated by quadratic and linear Gaussian factor models. We assume that the functions $\beta(t)$, $\Gamma(t)$ and $\Sigma(t)$ are piecewise constant functions, whose knots are $t_1^\star = 0.5$ and $t_2^\star = 0.5$. For instance, the function $\beta(t)$ is given by:

$$\beta(t) = \begin{cases}\beta_1 & \text{if } t \in [0, 0.5[\\ \beta_2 & \text{if } t \in [0.5, 1[\\ \beta_3 & \text{if } t \in [1, \infty)\end{cases}$$

where β_1, β_2 and β_3 are three scalars. Therefore, $\beta(t)$ is defined by the vector $(\beta_1, \beta_2, \beta_3)$. In a similar way, $\Gamma(t)$ and $\Sigma(t)$ are defined by the vectors $(\Gamma_1, \Gamma_2, \Gamma_3)$ and $(\Sigma_1, \Sigma_2, \Sigma_3)$. We

[73] In the general case $a(t) \neq 0$ and $B(t) \neq \mathbf{0}$, we have:

$$\begin{cases}\partial_t\hat{\alpha}(t,T) = -\hat{\beta}(t,T)^\top a(t) + \frac{1}{2}\hat{\beta}(t,T)^\top\Sigma(t)\Sigma(t)^\top\hat{\beta}(t,T) - \alpha(t)\\ \partial_t\hat{\beta}(t,T) = -B(t)^\top\hat{\beta}(t,T) - \beta(t)\end{cases}$$

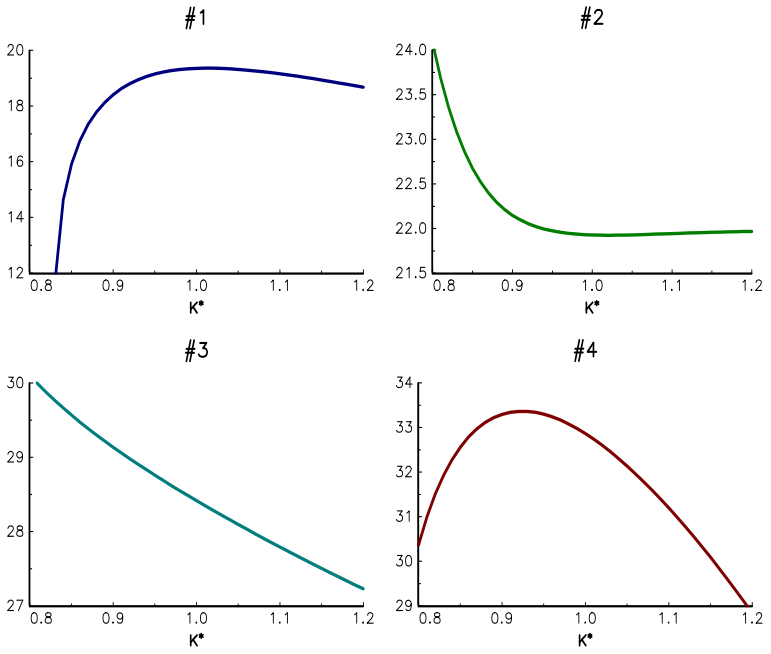


FIGURE 9.43: Volatility smiles generated by the quadratic Gaussian model

consider 4 parameter sets[74]:

Set	$(\beta_1, \beta_2, \beta_3)$	$(\Gamma_1, \Gamma_2, \Gamma_3)$	$(\Sigma_1, \Sigma_2, \Sigma_3)$
#1	$(0.3, 0.4, 0.5)$	$(-20, -10, 10)$	$(3, 3.2, 3.5)$
#2	$(0.3, 0.4, 0.5)$	$(20, 15, 10)$	$(3, 3.2, 3.5)$
#3	$(0.3, 0.4, 0.5)$	$(5, 5, 5)$	$(4, 3.5, 3)$
#4	$(0.3, 0.4, 0.5)$	$(-10, -10, -10)$	$(6, 5, 4)$

We also assume that the yield curve is flat and is equal to 5%. We consider the pricing of a caplet with $T_0 = T_1 - 2/365$, $T_1 = 0.5$ and $T_2 = 1.5$ for different strikes $K_i = K_i^\star \cdot \mathrm{Sw}(T_0)$ where $K_i^\star \in [0.8, 1.2]$. In Figure 9.43, we have reported the implied Black volatilities (in %) generated by the quadratic Gaussian model with the four parameter sets. We notice that the quadratic Gaussian model can generate different forms of volatility smiles. Since it is a little more flexible than the linear Gaussian model, we can obtained U-shaped and even reverse U-shaped volatility smiles.

9.3 Other model risk topics

In this section, we consider other risks than the volatility risk. In particular, we study the impact of dividends on option premia, the pricing of basket options and the liquidity risk.

[74] The volatilities $(\Sigma_1, \Sigma_2, \Sigma_3)$ are normalized by the factor $\sqrt{260} \times 10^{-4}$.

9.3.1 Dividend risk

9.3.1.1 Understanding the impact of dividends on option prices

Let us consider that the underlying asset pays a continuous dividend yield d during the life of the option. We have seen that the risk-neutral dynamics become:

$$\mathrm{d}S(t) = (r-d)\,S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t)$$

We deduce that the Black-Scholes formula is equal to:

$$\mathcal{C}_0 = S_0 e^{-dT}\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

where:

$$\begin{aligned} d_1 &= \frac{1}{\sigma\sqrt{T}}\left(\ln\frac{S_0}{K} + (r-d)\,T\right) + \frac{1}{2}\sigma\sqrt{T} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned}$$

We can also show that $\lim_{d\to\infty}\mathcal{C}_0 = 0$. In Figure 9.44, we report the price of the option when $K = 100$, $\sigma = 20\%$, $r = 5\%$ and $T = 0.5$. We consider different level of the dividend yield d. We notice that the call price is a decreasing function of the continuous dividend. If we consider put options instead of call options, the function becomes increasing.

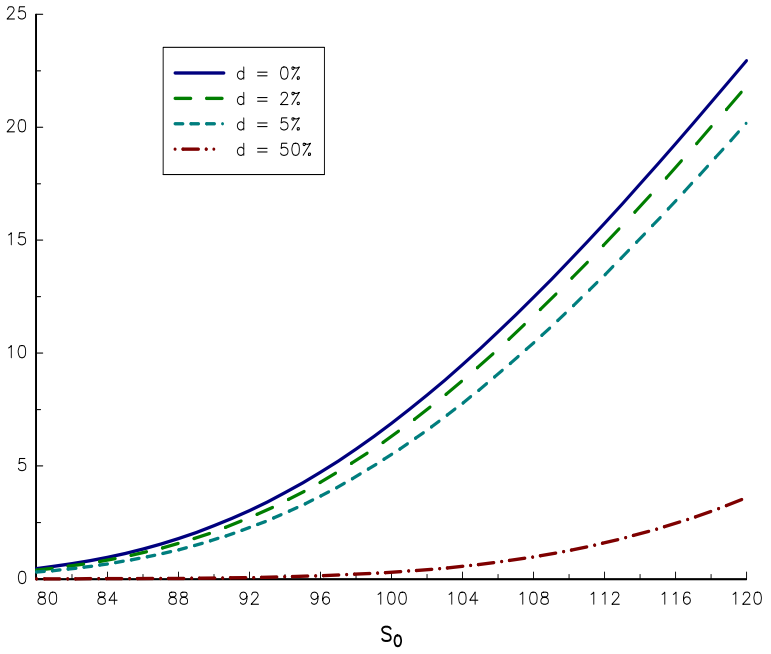


FIGURE 9.44: Impact of dividends on the call option price

We generally explain the impact of dividends because stock prices generally fall by the amount of the dividend on the ex-dividend date. Let $S(t)$ denote the value of the underlying asset at time t and D the discrete dividend paid at time t_D. We have:

$$S(t_D) = S\left(t_D^-\right) - D$$

The impact on the payoff is not the unique effect. Indeed, we recall that the option price is the cost of the replication portfolio. When the trader hedges the call option, he has a long exposure on the asset since the delta is positive. This implies that he receives the dividend of the asset. Therefore, the hedging cost of the call option is reduced. In the case of a put option, the trader has a short exposure and has to pay the dividend. As a result, the hedging cost of the put option is increased.

9.3.1.2 Models of discrete dividends

We denote by $S(t)$ the market price and $Y(t)$ an additional process that is assumed to be a geometric Brownian motion:

$$\mathrm{d}Y(t) = rY(t)\,\mathrm{d}t + \sigma Y(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

Following Frishling (2002), there are three main approaches to take into account discrete dividends. In the first approach, $Y(t)$ is the capital price process excluding the dividends and the market price $S(t)$ is equal to the sum of the capital price and the discounted value of future dividends:

$$S(t) = Y(t) + \sum_{t_k \in [t,T]} D(t_k)\, e^{-r(t_k - t)}$$

To price European options, we then replace the price S_0 by the adjusted price $Y_0 = S_0 - \sum_{t_k \leq T} D(t_k)\, e^{-rt_k}$. In the second approach, we define $D(t)$ as the sum of capitalized dividends paid until time t:

$$D(t) = \sum \mathbb{1}\{t_k < t\} \cdot D(t_k)\, e^{r(t - t_k)}$$

The market price $S(t)$ is equal to the difference between the cum-dividend price $Y(t)$ and the capitalized dividends (Haug *et al.*, 2003):

$$S(t) = Y(t) - D(t)$$

We deduce that:

$$\begin{aligned}(S(T) - K)^+ &= (Y(T) - D(T) - K)^+ \\ &= (Y(T) - (K + D(T)))^+ \\ &= (Y(T) - K')^+\end{aligned}$$

In the case of European options, we replace the strike K by the adjusted strike $K' = K + \sum_{t_k \leq T} D(t_k)\, e^{r(T - t_k)}$. The last approach considers the market price process as a discontinuous process:

$$\begin{cases} \mathrm{d}S(t) = rS(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) & \text{if } t_{k-1} < t < t_k \\ S(t) = S\left(t_k^-\right) - D(t_k) & \text{if } t = t_k \end{cases}$$

Therefore, we calculate the option price using finite differences or Monte Carlo simulations.

Remark 114 *The three models can be used to price exotic options, and not only European options. Generally, we do not have closed-form formulas and we calculate the price with numerical methods. For that, we have to define the risk-neural dynamics of $S(t)$. For instance, we have for the second model*[75]*:*

$$\mathrm{d}S(t) = \left(rS(t) - \sum \mathbb{1}\{t_k = t\} \cdot D(t_k)\, e^{r(t - t_k)}\right)\mathrm{d}t + \sigma\left(S(t) + D(t)\right)\mathrm{d}W^{\mathbb{Q}}(t)$$

[75] We notice that:

$$\mathrm{d}D(t) = \left(rD(t) + \sum \mathbb{1}\{t_k = t\} \cdot D(t_k)\, e^{r(t - t_k)}\right)\mathrm{d}t$$

Example 97 *We assume that* $S_0 = 100$, $K = 100$, $\sigma = 30\%$, $T = 1$, $r = 5\%$ *and* $b = 5\%$. *A dividend* $D(t_1)$ *will be paid at time* $t_1 = 0.5$.

Table 9.13 compares option prices when we use the three previous models. When $D(t_1)$ is equal to zero, the three models give the same price: the call option is equal to 14.23 whereas the put option is equal to 9.35. When the asset pays a dividend, the three models give different option prices. For instance, if the dividend is equal to 3, the call option is equal to 12.46 for Model #1, 12.81 for Model #2 and 12.69 for Model #3. We notice that the three models produce very different option prices[76]. Therefore, the choice of the dividend model has a big impact on the pricing of derivatives.

TABLE 9.13: Impact of the dividend on the option price

	Call			Put		
$D(t_1)$	(#1)	(#2)	(#3)	(#1)	(#2)	(#3)
0	14.23	14.23	14.23	9.35	9.35	9.35
3	12.46	12.81	12.69	10.51	10.86	10.64
5	11.34	11.92	11.69	11.34	11.92	11.59
10	8.78	9.93	9.42	13.66	14.80	14.20

Remark 115 *The previous models assume that dividends are not random at the inception date of the option. In practice, only the first dividend can be known if it has been announced before the inception date. This implies that dividends are generally unknown. Some authors have proposed option models with stochastic dividends, but they are not used by professionals. Most of the time, they use a very basic model. For instance, the Gordon growth model assumes that dividends increase at a constant rate* g*:*

$$D(t_k) = (1+g)^{(t_k - t_1)} D(t_1)$$

The parameter g *can be calibrated in order to match the forward prices.*

9.3.2 Correlation risk

Until now, we have studied the pricing and hedging of options that are based on one underlying asset. Banks have also developed derivatives with several underlying assets. In this case, the option price is sensitive to the covariance risk, which may be split between volatility risk and correlation risk. Here, we face two issues: the determination of implied correlations, and the hedging of the correlation risk.

9.3.2.1 The two-asset case

Pricing of basket options We consider the example of a basket option on two assets. Let $S_i(t)$ be the price process of asset i at time t. According to the Black-Scholes model, we have:

$$\begin{cases} \mathrm{d}S_1(t) = b_1 S_1(t)\,\mathrm{d}t + \sigma_1 S_1(t)\,\mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}S_2(t) = b_2 S_2(t)\,\mathrm{d}t + \sigma_2 S_2(t)\,\mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

where b_i and σ_i are the cost-of-carry and the volatility of asset i. Under the risk-neutral probability measure $\mathbb{Q}$, $W_1^{\mathbb{Q}}(t)$ and $W_2^{\mathbb{Q}}(t)$ are two correlated Brownian motions:

$$\mathbb{E}\left[W_1^{\mathbb{Q}}(t)\, W_2^{\mathbb{Q}}(t)\right] = \rho\, t$$

[76] We also notice that the price given by the third model is between the two prices calculated with the first and second models.

The option price associated to the payoff $(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+$ is the solution of the two-dimensional PDE:

$$\begin{aligned}\frac{1}{2}\sigma_1^2 S_1^2 \partial_{S_1}^2 \mathcal{C} + \frac{1}{2}\sigma_2^2 S_2^2 \partial_{S_2}^2 \mathcal{C} + \rho\sigma_1\sigma_2 S_1 S_2 \partial_{S_1,S_2}^2 \mathcal{C} + \\ b_1 S_1 \partial_{S_1} \mathcal{C} + b_2 S_2 \partial_{S_2} \mathcal{C} + \partial_t \mathcal{C} - r\mathcal{C} = 0\end{aligned}$$

with the terminal condition:

$$\mathcal{C}(T, S_1, S_2) = (\alpha_1 S_1 + \alpha_2 S_2 - K)^+$$

Using the Feynman-Kac representation theorem, we have:

$$\mathcal{C}_0 = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r\,\mathrm{d}t}(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+\right]$$

The value $\mathcal{C}_0$ can be calculated using numerical integration techniques such as Gauss-Legendre or Gauss-Hermite quadrature methods. In some cases, the two-dimensional problem can be reduced to one-dimensional integration. For instance, if $\alpha_1 < 0$, $\alpha_2 > 0$ and $K > 0$, we obtain[77]:

$$\mathcal{C}_0 = \int_{\mathbb{R}} \mathrm{BS}(S^\star(x), K^\star(x), \sigma^\star, T, b^\star, r)\,\phi(x)\,\mathrm{d}x$$

where $S^\star(x) = \alpha_2 S_2(0) e^{\rho\sigma_2\sqrt{T}x}$, $K^\star(x) = K - \alpha_1 S_1(0) e^{\left(b_1 - \frac{1}{2}\sigma_1^2\right)T + \sigma_1\sqrt{T}x}$, $\sigma^\star = \sigma_2\sqrt{1-\rho^2}$ and $b^\star = b_2 - \frac{1}{2}\rho^2\sigma_2^2$.

Example 98 *We assume that* $S_1(0) = S_2(0) = 100$, $\sigma_1 = \sigma_2 = 20\%$, $b_1 = 10\%$, $b_2 = 0$ *and* $r = 5\%$. *We calculate the price of a basket option, whose maturity* T *is equal to one year. For the other characteristics* (α_1, α_2, K), *we consider different set of parameters:* $(1, -1, 1)$, $(1, -1, 5)$, $(0.5, 0.5, 100)$, $(0.5, 0.5, 110)$ *and* $(0.1, 0.1, -5)$.

TABLE 9.14: Impact of the correlation on the basket option price

	α_1	1.0	1.0	0.5	0.5	0.1
	α_2	−1.0	−1.0	0.5	0.5	0.1
	K	1	5	100	110	−5
	−0.90	20.41	18.23	5.39	0.66	24.78
	−0.75	19.81	17.62	6.06	1.35	24.78
	−0.50	18.76	16.55	6.97	2.31	24.78
	−0.25	17.61	15.37	7.73	3.12	24.78
ρ	0.00	16.35	14.08	8.39	3.83	24.78
	0.25	14.94	12.61	8.99	4.46	24.78
	0.50	13.30	10.88	9.54	5.05	24.78
	0.75	11.29	8.66	10.05	5.59	24.78
	0.90	9.78	6.81	10.34	5.90	24.78

Using Gauss-Legendre quadratures, we obtain the prices of the basket option given in Table 9.14. We notice that the price can be an increasing, decreasing or independent function of the correlation parameter ρ.

[77] See Exercise 9.4.11 on page 602.

Remark 116 *We can extend the previous framework to other payoff functions. The PDE is the same, only the terminal condition changes:*

$$\mathcal{C}(T, S_1, S_2) = f(S_1(T), S_2(T))$$

where $f(S_1(T), S_2(T))$ *is the payoff function.*

Cega sensitivity The correlation risk studies the impact of the parameter ρ on the option price $\mathcal{C}_0$. For instance, Rapuch and Roncalli (2004) show that the price of the spread option, whose payoff is $(S_1(T) - S_2(T) - K)^+$, is a decreasing function of the correlation parameter ρ. They also extend this result to an arbitrary European payoff $f(S_1(T), S_2(T))$. In particular, they demonstrate that, if the cross-derivative $\partial^2_{1,2} f$ is a negative (resp. positive) measure, then the option price is decreasing (resp. increasing) with respect to ρ. For instance, the payoff function of the call option on the maximum of two assets is defined as $f(S_1, S_2) = (\max(S_1, S_2) - K)^+$. Since $\partial^2_{1,2} f(S_1, S_2) = -\mathbb{1}\{S_1 = S_2, S_1 > K\}$ is a negative measure, the option price decreases with respect to ρ. In the case of a Best-of call/call option, the payoff function is $f(S_1, S_2) = \max\left((S_1 - K_1)^+, (S_2 - K_2)^+\right)$ and we have:

$$\partial^2_{1,2} f(S_1, S_2) = -\mathbb{1}\{S_2 - K_2 - S_1 + K_1 = 0, S_1 > K_1, S_2 > K_2\}$$

We have the same behavior than the Max option. For the Min option, we remark that $\min(S_1, S_2) = S_1 + S_2 - \max(S_1, S_2)$. So, the option price is an increasing function of ρ. Other results could be found in Table 9.15.

TABLE 9.15: Relationship between the basket option price and the correlation parameter ρ

Option type	Payoff	Increasing	Decreasing
Spread	$(S_2 - S_1 - K)^+$		✓
Basket	$(\alpha_1 S_1 + \alpha_2 S_2 - K)^+$	$\alpha_1\alpha_2 > 0$	$\alpha_1\alpha_2 < 0$
Max	$(\max(S_1, S_2) - K)^+$		✓
Min	$(\min(S_1, S_2) - K)^+$	✓	
Best-of call/call	$\max\left((S_1 - K_1)^+, (S_2 - K_2)^+\right)$		✓
Best-of put/put	$\max\left((K_1 - S_1)^+, (K_2 - S_2)^+\right)$		✓
Worst-of call/call	$\min\left((S_1 - K_1)^+, (S_2 - K_2)^+\right)$	✓	
Worst-of put/put	$\min\left((K_1 - S_1)^+, (K_2 - S_2)^+\right)$	✓	

The sensitivity of the option price with respect to the correlation parameter ρ is called the cega:

$$\boldsymbol{c} = \frac{\partial \mathcal{C}_0}{\partial \rho}$$

Generally, it is difficult to fix a particular value of ρ, because a correlation is not a stable parameter. Moreover, the value of ρ used for pricing the option must reflect the risk-neutral distribution. Then, it is not obvious that the '*risk-neutral correlation*' is equal to the '*historical correlation*'. Most of the time, we only have an idea about the correlation range $\rho \in [\rho^-, \rho^+]$. The previous analysis leads us to define the lower and upper bounds of the option price when the cega is either positive or negative. We have:

$$\mathcal{C}_0 \in \begin{cases} [\mathcal{C}_0(\rho^-), \mathcal{C}_0(\rho^+)] & \text{if } \boldsymbol{c} \geq 0 \\ [\mathcal{C}_0(\rho^+), \mathcal{C}_0(\rho^-)] & \text{if } \boldsymbol{c} \leq 0 \end{cases}$$

We can define the conservative price by taking the maximum between $\mathcal{C}_0(\rho^-)$ and $\mathcal{C}_0(\rho^+)$.

Remark 117 *In the case where* $\rho^- = -1$ *and* $\rho^+ = 1$*, the bounds satisfy the one-dimensional PDE:*

$$\begin{cases} \frac{1}{2}\sigma_1^2 S^2 \partial_S^2 \boldsymbol{\mathcal{C}}(t,S) + b_1 S \partial_S \boldsymbol{\mathcal{C}}(t,S) + \partial_t \boldsymbol{\mathcal{C}}(t,S) - r\boldsymbol{\mathcal{C}}(t,S) = 0 \\ \boldsymbol{\mathcal{C}}(T,S) = f(S, g(S)) \end{cases}$$

where:

$$g(S) = S_2(0)\left(\frac{S}{S_1(0)}\right)^{\pm\sigma_2/\sigma_1} \exp\left(\left(b_2 - \frac{1}{2}\sigma_2^2 \pm \left(\frac{1}{2}\sigma_1\sigma_2 - \frac{\sigma_2}{\sigma_1}b_1\right)\right)T\right)$$

The implied correlation Like the implied volatility, the implied correlation is the value we put into the Black-Scholes formula to get the true market price. At first sight, the concept of implied correlation seems to be straightforward. For instance, let us consider composite options, whose payoff is defined by $(S_1(T) - kS_2(T))^+$. It is a special case of the general payoff $(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+$ where $\alpha_1 = 1$, $\alpha_2 = k$ and $K = 0$. The parameters are those given in Example 98. The values $(k, \boldsymbol{\mathcal{C}}_0)$ taken by the relative strike k and the market price $\boldsymbol{\mathcal{C}}_0$ are respectively equal to $(0.10, 95.61)$, $(0.20, 86.10)$, $(0.30, 76.59)$, $(0.40, 67.08)$, $(0.50, 57.57)$, $(0.60, 48.06)$, $(0.70, 38.62)$, $(0.80, 29.46)$, $(0.90, 21.12)$, $(1.00, 14.32)$, $(1.10, 9.45)$ and $(1.20, 6.30)$. Using these 12 market prices, we deduce the correlation smile with respect to k in Figure 9.45. We now consider the option, whose payoff is $\left(\frac{1}{2}S_1(T) + \frac{1}{2}S_2(T) - 100\right)^+$. Which correlation should be used? There is no obvious answer. Indeed, we notice that a correlation smile is always associated to a given payoff. This is why it is generally not possible to use a correlation smile deduced from one payoff function to price the option with another payoff function. Contrary to volatility, the concept of implied correlation makes sense, but not the concept of correlation smile.

Riding on the smiles Until now, we have assumed that the volatilities of the two assets are given. In practice, the two volatilities are unknown and must be deduced from the volatility smiles $\Sigma_1(K_1, T)$ and $\Sigma_2(K_2, T)$ of the two assets. The difficulty is then to find the corresponding strikes K_1 and K_2. In the case of the general payoff $(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+$, we have:

$$\begin{cases} (\alpha_1 = 1, \alpha_2 = 0, K \geq 0) \Rightarrow K_1 = K \\ (\alpha_1 = -1, \alpha_2 = 0, K \leq 0) \Rightarrow K_1 = -K \end{cases}$$

and:

$$\begin{cases} (\alpha_1 = 0, \alpha_2 = 1, K \geq 0) \Rightarrow K_2 = K \\ (\alpha_1 = 0, \alpha_2 = -1, K \leq 0) \Rightarrow K_2 = -K \end{cases}$$

The payoff of the spread option can be written as follows:

$$\begin{aligned} (S_1(T) - S_2(T) - K)^+ &= ((S_1(T) - K_1) + (K_2 - S_2(T)))^+ \\ &\leq \underbrace{(S_1(T) - K_1)^+}_{\text{Call}} + \underbrace{(K_2 - S_2(T))^+}_{\text{Put}} \end{aligned}$$

where $K_1 = K_2 + K$. Therefore, the price of the spread option can be bounded above by a call price on S_1 plus a put price on S_2. However, the implicit strikes can take different values. Let us assume that $S_1(0) = S_2(0) = 100$ and $K = 4$. Below, we give five pairs

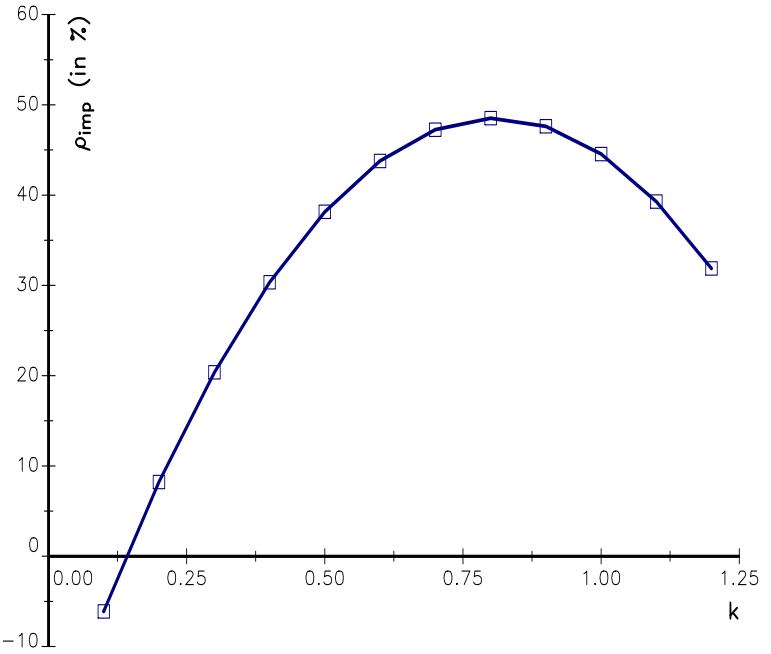


FIGURE 9.45: Correlation smile

(K_1, K_2) and the associated implied volatilities $(\Sigma_1(K_1, T), \Sigma_2(K_2, T))$:

Pair	#1	#2	#3	#4	#5
K_1	104	103	102	101	100
K_2	100	99	98	97	96
$\Sigma_1(K_1, T)$	16%	17%	18%	19%	20%
$\Sigma_2(K_2, T)$	20%	22%	24%	26%	28%
$\mathcal{C}_0$	10.77	11.37	11.99	12.61	13.24

We also compute the price of the spread option[78] and report it in the last row of the above table. We notice that the price varies from 10.77 to 13.24, even if we use the same correlation parameter. We face here an issue, because this simple example shows that two-dimensional option pricing is not just an extension of one-dimensional option pricing, and the concept of implied volatility becomes blurred.

9.3.2.2 The multi-asset case

How to define a conservative price? In the multivariate case, the PDE becomes:

$$\frac{1}{2}\sum_{i=1}^{n}\sigma_i^2 S_i^2 \partial^2_{S_i}\mathcal{C} + \sum_{i<j}^{n}\rho_{i,j}\sigma_i\sigma_j S_i S_j \partial^2_{S_i,S_j}\mathcal{C} + \sum_{i=1}^{n} b_i S_i \partial_i \mathcal{C} + \partial_t \mathcal{C} - r\mathcal{C} = 0$$

with the terminal value:

$$\mathcal{C}(T, S_1, \ldots, S_n) = f(S_1(T), \ldots, S_n(T))$$

[78] The parameters are $b_1 = 10\%$, $b_2 = 0\%$, $r = 5\%$, $\rho = 50\%$ and $T = 1$.

Here, $\rho_{i,j}$ is the correlation between the Brownian motions of S_i and S_j. Most of the time, the trader uses the same value ρ for all asset correlations $\rho_{i,j}$.

Rapuch and Roncalli (2004) show that the price is increasing (resp. decreasing) with respect to ρ if $\sum_{i<j}^{n} \sigma_i \sigma_j \partial^2_{S_i,S_j} f$ is a positive (resp. negative) measure. Let us consider the payoff function $f(S_1, S_2, S_3) = (S_1 + S_2 - S_3 - K)^+$, we have:

$$\sum_{i<j}^{n} \sigma_i \sigma_j \partial^2_{S_i,S_j} f = (\sigma_1\sigma_2 - \sigma_1\sigma_3 - \sigma_2\sigma_3) \cdot \mathbb{1}\{S_1 + S_2 - S_3 - K = 0\}$$

Hence, if $\sigma_1\sigma_2 - \sigma_1\sigma_3 - \sigma_2\sigma_3 > 0$, the price increases with respect to ρ, and if $\sigma_1\sigma_2 - \sigma_1\sigma_3 - \sigma_2\sigma_3 < 0$, the price decreases with respect to ρ. As a result, it is more difficult to define conservative prices for multi-asset options.

Issues with constant correlation matrices We consider a basket of n stocks. The basket volatility is given by:

$$\sigma_B = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i>j}^{n} \rho_{i,j} w_i w_j \sigma_i \sigma_j}$$

where w_i is the weight of asset i in the basket, σ_i the volatility of asset i and $\rho_{i,j}$ the correlation between asset i and asset j. The implied correlation ρ_{imp} of the basket is defined as the root of the following equation:

$$\sigma_B^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2 - 2\rho_{\text{imp}} \sum_{i>j}^{n} w_i w_j \sigma_i \sigma_j = 0$$

Skintzi and Refenes (2003) deduce that:

$$\rho_{\text{imp}} = \frac{\sigma_B^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{2 \sum_{i>j}^{n} w_i w_j \sigma_i \sigma_j}$$

Another expression of the implied correlation is[79]:

$$\rho_{\text{imp}} = \frac{\sigma_B^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2}$$

The concept of implied correlation has been very popular before the Global Financial Crisis. It was at the heart of a strategy known as volatility dispersion trading, which consists in selling variance swaps on an index and buying variance swaps on index components.

The previous analysis assumes a constant correlation matrix $\mathbb{C}_n(\rho)$ for modeling the dependance between asset returns. Over time, it has become the standard for pricing basket

[79] Indeed, we have:

$$\sigma_{\max} = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i>j} w_i w_j \sigma_i \sigma_j} = \sum_{i=1}^{n} w_i \sigma_i$$

implying that:

$$2 \sum_{i>j}^{n} w_i w_j \sigma_i \sigma_j = \left(\sum_{i=1}^{n} w_i \sigma_i\right)^2 - \sum_{i=1}^{n} w_i^2 \sigma_i^2$$

options with several assets. However, this approach implies a specific factor model. It is equivalent to assume that the underlying assets depend on a common risk factor with the same sensitivity. With such assumption, it is extremely difficult to estimate the conservative price of basket options with barriers, best-of/worst-of options, etc. To illustrate this problem, we consider the following payoff:

$$\left(S_1(T) - S_2(T) + S_3(T) - S_4(T) - K\right)_+ \cdot \mathbb{1}\left\{S_5(T) > L\right\}$$

We calculate the option price of maturity 3 months using the Black-Scholes model. We assume that $S_i(0) = 100$ and $\Sigma_i = 20\%$ for the five underlying assets, the strike K is equal to 5, the barrier L is equal to 105, and the interest rate r is set to 5%. In Figure 9.46, we report the option price when the correlation matrix is $\mathbb{C}_5(\rho)$. Since the option price decreases with respect to ρ, it can be bounded above by 2.20. If we simulate correlation matrices with uniform singular values, we notice that the maximum price of 2.20 is not a conservative price. For instance, if we consider the correlation matrix below, we obtain an option price of 3.99:

$$\mathbb{C} = \begin{pmatrix} 1.0000 & 0.2397 & 0.7435 & -0.1207 & 0.0563 \\ 0.2397 & 1.0000 & -0.0476 & -0.0260 & -0.1958 \\ 0.7435 & -0.0476 & 1.0000 & 0.2597 & 0.1153 \\ -0.1207 & -0.0260 & 0.2597 & 1.0000 & -0.7568 \\ 0.0563 & -0.1958 & 0.1153 & -0.7568 & 1.0000 \end{pmatrix}$$

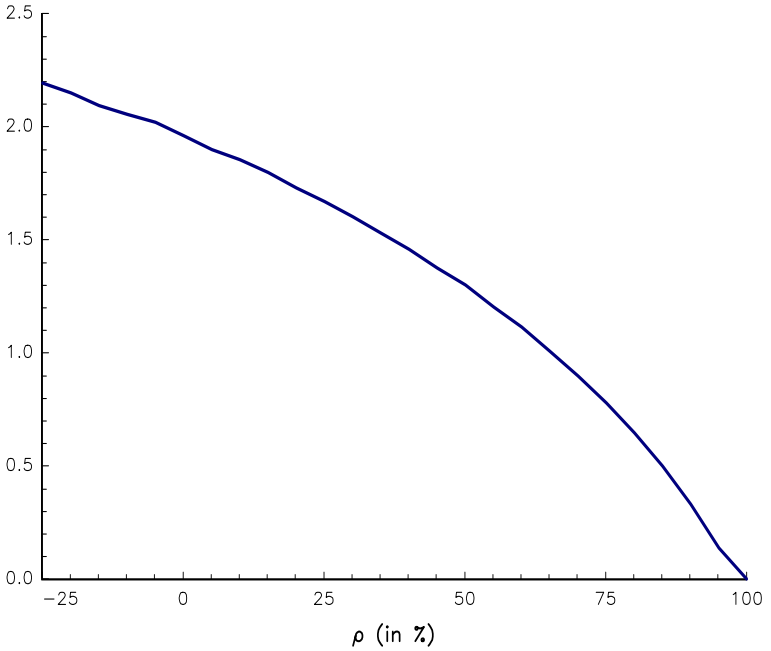


FIGURE 9.46: Price of the basket option with respect to the constant correlation

9.3.2.3 The copula method

Using Sklar's theorem, it comes that the multivariate risk-neutral distribution has the following canonical representation:

$$\mathbb{Q}\left(S_1(t), \ldots, S_n(t)\right) = \mathbf{C}^{\mathbb{Q}}\left(\mathbb{Q}_1\left(S_1(t)\right), \ldots, \mathbb{Q}_n\left(S_n(t)\right)\right)$$

$\mathbf{C}^{\mathbb{Q}}$ is called the risk-neutral copula (Cherubini and Luciano, 2002). The copula approach has been extensively used in order to derive the bounds of basket options. For instance, Rapuch and Roncalli (2004) extend the results presented in Section 9.3.2.1 on page 583 to the copula approach. In particular, they show that if the payoff function f is supermodular[80], then the option price increases with respect to the concordance order. More explicitly, we have:

$$\mathbf{C}_1 \prec \mathbf{C}_2 \Rightarrow \boldsymbol{\mathcal{C}}_0\left(S_1, S_2; \mathbf{C}_1\right) \leq \boldsymbol{\mathcal{C}}_0\left(S_1, S_2; \mathbf{C}_2\right)$$

Therefore, the previous results hold if we replace the Black-Scholes model with the Normal copula model. Thus, the spread option is a decreasing function of the Normal copula parameter ρ even if we use a local or stochastic volatility model in place of the Black-Scholes model. In a similar way, one can find lower and upper bounds of multi-asset option prices by considering lower and upper Fréchet copulas. As shown by Tankov (2011), these bounds can be improved significantly when partial information is available such as the prices of digital basket options.

In practice, the Normal copula model is extensively used for pricing multi-asset European-style option for two reasons:

1. The first one is that multi-asset option prices must be '*compatible*' with single-asset option prices. This means that it would be inadequate to price single-asset options with a complex model, e.g. the SABR model, and in the same time to price multi-asset options with the multivariate Black-Scholes model. Indeed, this decoupling approach creates arbitrage opportunities at the level of the bank itself.

2. The Normal copula model is a natural extension of the multivariate Black-Scholes model since the dependence function is the same.

Nevertheless, we face an issue because the pricing of the payoff $f\left(S_1\left(T\right), \ldots, S_n\left(T\right)\right)$ requires knowing the joint distribution of the random vector $\left(S_1\left(T\right), \ldots, S_n\left(T\right)\right)$, whose an analytical expression does not generally exist[81]. This is why multi-asset options are priced using the Monte Carlo method. However, the analytical distribution of the marginals are generally unknown. Therefore, we have to implement the method of empirical quantile functions described on page 806:

1. for each random variable $S_i\left(T\right)$, simulate m_1 random variates $S_{i,m}^{\star}$ and estimate the empirical distribution $\hat{\mathbf{F}}_i$;

2. simulate a random vector $\left(u_{1,j}, \ldots, u_{n,j}\right)$ from the copula function $\mathbf{C}\left(u_1, \ldots, u_n\right)$;

3. simulate the random vector $\left(S_{1,j}, \ldots, S_{n,j}\right)$ by inverting the empirical distributions $\hat{\mathbf{F}}_i$:

$$S_{i,j} \leftarrow \hat{\mathbf{F}}_i^{-1}\left(u_{i,j}\right)$$

   or equivalently:

$$S_{i,j} \leftarrow \inf\left\{x \left| \frac{1}{m_1} \sum_{m=1}^{m_1} \mathbb{1}\left\{x \leq S_{i,m}^{\star}\right\} \geq u_i \right.\right\}$$

[80] The function f is supermodular if and only if:

$$\Delta^{(2)} f := f\left(x_1 + \varepsilon_1, x_2 + \varepsilon_2\right) - f\left(x_1 + \varepsilon_1, x_2\right) - f\left(x_1, x_2 + \varepsilon_2\right) + f\left(x_1, x_2\right) \geq 0$$

for all $(x_1, x_2) \in \mathbb{R}^2$ and $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}_+^2$.

[81] An exception concerns the SABR model for which we have found an expression of the probability distribution thanks to the Breeden-Litzenberger representation.

4. repeat steps 2 and 3 m_2 times;
5. the MC estimate of the option price is equal to:

$$\hat{\mathcal{C}}_0 = e^{-rT}\left(\frac{1}{m_2}\sum_{j=1}^{m_2} f\left(S_{1,j},\ldots,S_{n,j}\right)\right)$$

It follows that the first step is used for estimating the distribution of $S_i(T)$. For this step, we use m_1 simulations of the single-asset option model. However, this step generates independent random variables. Therefore, the steps 2 and 3 are used in order to create the right dependence between $(S_1(T),\ldots,S_n(T))$.

Example 99 *We consider the two-asset option with the following payoff:*

$$f\left(F_1(T),F_2(T)\right) = 100\cdot\left(\max\left(\frac{F_1(T)}{F_1(0)}-1,\frac{F_2(T)}{F_2(0)}-1\right)-K\right)^+$$

where $F_1(t)$ and $F_2(t)$ are two forward rates. We assume that $F_1(0)=5\%$ and $F_2(0)=6\%$. The maturity of the option is equal to one year, whereas the strike of the option is set to 2%. Using the SABR model, we have calibrated the volatility smiles and we have obtained the following estimates:

	α	β	ν	ρ
F_1	8.944%	1.00	0.322	−22.901%
F_2	12.404%	1.00	0.280	16.974%

In Figure 9.47, we have reported the price of the two-asset option with respect to the dependence parameter ρ. For the Black-Scholes model, we use the ATM implied volatilities[82] and the parameter ρ represents the implied correlation. For the SABR model, we use the Normal copula model, and ρ is the copula parameter. We notice that the Black-Scholes model overestimates the option price compared to the SABR model. We also verified that the option price is a decreasing function with respect to ρ.

9.3.3 Liquidity risk

Liquidity risk can be incorporated in the theory of option pricing, but it requires solving a stochastic optimal control problem (Çetin *et al.*, 2004, 2006; Jarrow and Protter, 2007; Çetin *et al.*, 2010). In practice, these approaches are not used by professionals, but some theoretical results help to understand the impact of liquidity risk on option pricing. However, there is no satisfactory solution, and '*cooking recipes*' differ from one bank to another one, one trading desk to another one, one trader to another one. But the issue here is not to solve this problem, but to understand the model risk from a risk management perspective.

It is obvious that liquidity risk impacts trading costs, in particular the price of the replication strategy because of bid-ask spreads. Here, we don't want to focus on '*normal*' liquidity risk, but on '*trading*' liquidity risk. Option theory assumes that we can replicate the option, meaning that we can sell or buy the underlying asset at any time. For liquid assets, this assumption is almost verified even if we can face high bid-ask spread. For less liquid assets, this assumption is not verified. Let us consider one of the most famous examples, which concerns call options on Sharpe ratio. Starting from 2004, some banks proposed

[82] They are equal to 9% for F_1 and 12.5% for F_2.

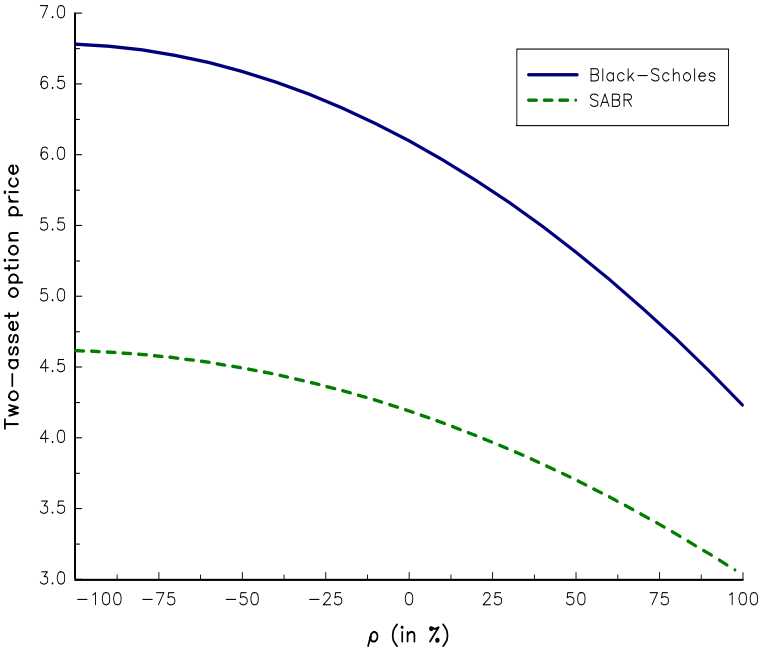


FIGURE 9.47: Comparison of the option price obtained with Black-Scholes and copula-SABR models

to investors a payoff of the form $(\mathrm{SR}(0;T)-K)^+$ where $\mathrm{SR}(0;T)$ is the Sharpe ratio of the underlying asset during the option period. This payoff is relatively easy to replicate. However, most of call options on Sharpe ratio have been written on mutual funds and hedge funds. The difficulty comes from the liquidity of these underlying assets. For instance, the trader does not know exactly the price of the asset when he executes his order because of the notice period[83]. This can be a big issue when the fund offers weekly or monthly liquidity. The second problem comes from the fact that the fund manager can impose lock-up period and gates. For instance, a gate limits the amount of withdrawals. During the 2008/2009 hedge fund crisis, many traders faced gate provisions and were unable to adjust their delta. This crisis marketed the end of call options on Sharpe ratio.

The previous example is an extreme case of the impact of liquidity on option trading. However, this type of problems is not unusual even with liquid markets, because liquidity is time-varying and may impact delta hedging at the worst possible time. Let us consider the replication of a call option. If the price of the underlying asset decreases sharply, the delta is reduced and the option trader has to sell asset shares. Because of their trend-following aspect, option traders generally buy assets when the market goes up and sell assets when the market goes down. However, we know that liquidity is asymmetric between these two market regimes. Therefore, it is more difficult to adjust the delta exposure when the market goes down, because of the lack of liquidity. This means that some payoffs are more sensitive to others.

[83] A subscription/redemption notice period requires that the investor informs the fund manager a certain period in advance before buying/selling fund shares.

9.4 Exercises

9.4.1 Option pricing and martingale measure

We consider the Black-Scholes model. The price process $S(t)$ follows a Geometric Brownian motion:

$$\mathrm{d}S(t) = \mu S(t)\ \mathrm{d}t + \sigma S(t)\ \mathrm{d}W(t)$$

and the risk-free asset $B(t)$ satisfies:

$$\mathrm{d}B(t) = rB(t)\ \mathrm{d}t$$

We consider a portfolio $(\phi(t), \psi(t))$ invested in the stock S and the risk-free bond B. We note $V(t)$ the value of this portfolio.

1. Show that:
$$\mathrm{d}V(t) = rV(t)\ \mathrm{d}t + \phi(t)\left(\mathrm{d}S(t) - rS(t)\ \mathrm{d}t\right)$$
2. We note $\tilde{V}(t) = e^{-rt}V(t)$ and $\tilde{S}(t) = e^{-rt}S(t)$. Show that:
$$\mathrm{d}\tilde{V}(t) = \phi(t)\ \mathrm{d}\tilde{S}(t)$$
3. Show that $\tilde{V}(t)$ is a martingale under the risk measure $\mathbb{Q}$. Deduce that:
$$V(t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[V(T) \middle| \mathcal{F}_t\right]$$
4. Define the corresponding martingale measure.
5. Calculate the price of the binary option $\mathbb{1}\{S(T) \geq K\}$.

9.4.2 The Vasicek model

Vasicek (1977) assumes that the instantaneous interest rate follows an Ornstein-Uhlenbeck process:

$$\begin{cases} \mathrm{d}r(t) = a(b - r(t))\ \mathrm{d}t + \sigma\ \mathrm{d}W(t) \\ r(t_0) = r_0 \end{cases}$$

and the risk price of the Wiener process is constant:

$$\lambda(t) = \lambda$$

We consider the pricing of a zero-coupon bond, whose maturity is equal to T.

1. Write the partial differential equation of the zero-coupon bond $B(t, r)$ when the interest rate $r(t)$ is equal to r.
2. Using the solution of the Ornstein-Uhlenbeck process given on page 1075, show that the random variable Z defined by:
$$Z = \int_0^T r(t)\ \mathrm{d}t$$
is Gaussian.
3. Calculate the first two moments.
4. Deduce the price of the zero-coupon bond.

9.4.3 The Black model

In the model of Black (1976), we assume that the price $F(t)$ of a forward or futures contract evolves as follows:

$$\mathrm{d}F(t) = \sigma F(t)\ \mathrm{d}W(t)$$

1. Write the PDE equation associated to the call option payoff:

$$\mathcal{C}(T) = \max(F(T) - K, 0)$$

when the interest rate is equal to r.

2. Using the Feynman-Kac representation theorem, deduce the current price of the call option.

3. We assume that the stock price $S(t)$ follows a geometric Brownian motion:

$$\mathrm{d}S(t) = \mu S(t)\ \mathrm{d}t + \sigma S(t)\ \mathrm{d}W(t)$$

Show that the Black formula can be used to price an European option, whose underlying asset is the futures contract of the stock.

4. What does the Black formula become if we assume that the interest rate $r(t)$ is stochastic and is independent of the forward price $F(t)$?

5. What is the problem if we consider that the interest rate $r(t)$ and the forward price $F(t)$ are not independent?

6. We reiterate that the price of the zero-coupon bond is given by:

$$B(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s)\,\mathrm{d}s}\middle|\,\mathcal{F}_t\right]$$

The instantaneous forward rate $f(t,T)$ is defined as follows:

$$f(t,T) = -\frac{\partial \ln B(t,T)}{\partial T}$$

We consider that the numéraire is the bond price $B(t,T)$ and we note $\mathbb{Q}^\star$ the associated forward probability measure.

(a) Show that:

$$\frac{\partial B(t,T)}{\partial T} = -B(t,T)\cdot\mathbb{E}^{\mathbb{Q}^\star}\left[f(T,T)|\,\mathcal{F}_t\right]$$

(b) Deduce that $f(t,T)$ is an $\mathcal{F}_t$-martingale under the forward probability measure $\mathbb{Q}^\star$.

(c) Find the price of the call option, whose payoff is equal to:

$$\mathcal{C}(T) = \max(f(T,T) - K, 0)$$

9.4.4 Change of numéraire and Girsanov theorem

Part one

Let $X(t)$ and $Y(t)$ be two $\mathcal{F}_t$-adapted processes.

1. Calculate the stochastic differentials $\mathrm{d}(X(t)Y(t))$ and $\mathrm{d}(1/Y(t))$.

2. We note $Z(t)$ the ratio of $X(t)$ and $Y(t)$. Show that:

$$\frac{\mathrm{d}Z(t)}{Z(t)} = \frac{\mathrm{d}X(t)}{X(t)} - \frac{\mathrm{d}Y(t)}{Y(t)} + \frac{\langle \mathrm{d}Y(t), \mathrm{d}Y(t)\rangle}{Y^2(t)} - \frac{\langle \mathrm{d}X(t), \mathrm{d}Y(t)\rangle}{X(t)Y(t)}$$

Part two

Let $S(t)$ be the price of an asset. Under the probability measure $\mathbb{Q}$, $S(t)$ has the following dynamics:

$$\mathrm{d}S(t) = \mu_S(t) S(t)\ \mathrm{d}t + \sigma_S(t) S(t)\ \mathrm{d}W^{\mathbb{Q}}(t)$$

The corresponding numéraire is denoted by $M(t)$ and we have:

$$\mathrm{d}M(t) = \mu_M(t) M(t)\ \mathrm{d}t + \sigma_M(t) M(t)\ \mathrm{d}W^{\mathbb{Q}}(t)$$

We now consider another numéraire $N(t)$ whose dynamics is given by:

$$\mathrm{d}N(t) = \mu_N(t) N(t)\ \mathrm{d}t + \sigma_N(t) N(t)\ \mathrm{d}W^{\mathbb{Q}}(t)$$

and we note $\mathbb{Q}^\star$ the probability measure associated to $N(t)$. We assume that:

$$\mathrm{d}S(t) = \mu_S^\star(t) S(t)\ \mathrm{d}t + \sigma_S(t) S(t)\ \mathrm{d}W^{\mathbb{Q}^\star}(t)$$

1. Why can we assume that the diffusion coefficient of $S(t)$ is the same under the two probability measures $\mathbb{Q}$ and $\mathbb{Q}^\star$?

2. Find the process $g(t)$ such that:

$$\mathrm{d}W^{\mathbb{Q}^\star}(t) = \mathrm{d}W^{\mathbb{Q}}(t) - g(t)\ \mathrm{d}t$$

   Let $Z(t)$ be the Radon-Nikodym derivative defined by:

$$Z(t) = \frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}}$$

   Show that:

$$\frac{\mathrm{d}Z(t)}{Z(t)} = g(t)\ \mathrm{d}W^{\mathbb{Q}}(t)$$

3. We recall that another expression of $Z(t)$ is:

$$Z(t) = \frac{N(t)/N(0)}{M(t)/M(0)}$$

   Deduce that:

$$g(t) = \sigma_N(t) - \sigma_M(t)$$

   Find the expression of $\mu_N(t)$.

4. Show that changing the numéraire is equivalent to change the drift:

$$\mu_S^\star(t) = \mu_S(t) + \sigma_S(t)(\sigma_N(t) - \sigma_M(t))$$

5. Deduce that:

$$\mu_S^\star(t)\ \mathrm{d}t - \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}N(t)}{N(t)} \right\rangle = \mu_S(t)\ \mathrm{d}t - \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \frac{\mathrm{d}M(t)}{M(t)} \right\rangle$$

   and:

$$\mu_S^\star(t)\ \mathrm{d}t = \mu_S(t)\ \mathrm{d}t + \left\langle \frac{\mathrm{d}S(t)}{S(t)}, \mathrm{d}\ln\frac{N(t)}{M(t)} \right\rangle$$

Part three

Under the risk-neutral probability measure $\mathbb{Q}$, we assume that the asset price and the numéraire are given by the following stochastic differential equations:

$$\mathrm{d}S(t) = r(t) S(t)\ \mathrm{d}t + \sigma_S(t) S(t)\ \mathrm{d}W_S^{\mathbb{Q}}(t)$$

and:

$$\mathrm{d}N(t) = r(t) N(t)\ \mathrm{d}t + \sigma_N(t) N(t)\ \mathrm{d}W_N^{\mathbb{Q}}(t)$$

where $N(0) = 1$, $W_S^{\mathbb{Q}}(t)$ and $W_N^{\mathbb{Q}}(t)$ are two Wiener processes and $\mathbb{E}\left[W_S^{\mathbb{Q}}(t) W_N^{\mathbb{Q}}(t)\right] = \rho t$. We note $\tilde{S}(t) = S(t)/N(t)$ the asset price expressed in the numéraire $N(t)$.

1. Find the stochastic differential equation of $\tilde{S}(t)$:

$$\tilde{S}(t) = \frac{S(t)}{N(t)}$$

2. Let $Q^\star$ be the martingale measure associated to the numéraire $N(t)$.

   (a) We assume that $\sigma_N(t) = 0$. Show that the discounted asset price is an $\mathcal{F}_t$-martingale under the risk-neutral probability measure.

   (b) We consider the case $W_S^{\mathbb{Q}}(t) = W_N^{\mathbb{Q}}(t)$. Using Girsanov theorem, show that:

$$\mathrm{d}\tilde{S}(t) = \tilde{\sigma}(t) \tilde{S}(t)\ \mathrm{d}W^{\mathbb{Q}^\star}(t)$$

   where $W^{\mathbb{Q}^\star}$ is a Brownian motion under the probability measure $Q^\star$ and $\tilde{\sigma}(t)$ is a function to be defined.

   (c) What does this result become in the general case?

9.4.5 The HJM model and the forward probability measure

We assume that the instantaneous forward rate $f(t, T_1)$ is given by the following stochastic differential equation:

$$\mathrm{d}f(t, T_1) = \alpha(t, T_1)\ \mathrm{d}t + \sigma(t, T_1)\ \mathrm{d}W^{\mathbb{Q}}(t)$$

where $\mathbb{Q}$ is the risk-neutral probability measure.

1. We consider the forward probability measure $\mathbb{Q}^\star(T_2)$ where $T_2 \geq T_1$. Define the corresponding numéraire $N(t)$ and show that the Radon-Nikodym derivative is equal to:

$$\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} = e^{-\int_0^{T_2} (r(t) - f(0,t))\,\mathrm{d}t}$$

2. We recall that the dynamics of the instantaneous spot rate $r(t)$ is:

$$r(t) = r(0) + \int_0^t \left(\sigma(s,t) \int_s^t \sigma(s,u)\ \mathrm{d}u\right) \mathrm{d}s + \int_0^t \sigma(s,t)\ \mathrm{d}W^{\mathbb{Q}}(s)$$

   Show that:

$$\frac{\mathrm{d}\mathbb{Q}^\star}{\mathrm{d}\mathbb{Q}} = e^{\int_0^{T_2} a(t,T_2)\,\mathrm{d}t + \int_0^{T_2} b(t,T_2)\,\mathrm{d}W^{\mathbb{Q}}(t)}$$

where:

$$a(t,T_2) = -\int_t^{T_2} \left(\sigma(t,v) \int_t^v \sigma(t,u)\, \mathrm{d}u \right) \mathrm{d}v$$

and:

$$b(t,T_2) = -\int_t^{T_2} \sigma(t,v)\, \mathrm{d}v$$

3. Using the drift restriction in the HJM model, show that:

$$W^{\mathbb{Q}^\star(T_2)}(t) = W^{\mathbb{Q}}(t) - \int_0^t b(s,T_2)\, \mathrm{d}s$$

is a Brownian motion under the forward probability measure $\mathbb{Q}^\star(T_2)$.

4. Find the dynamics of $f(t,T_1)$ under the forward probability measure $\mathbb{Q}^\star(T_2)$.

5. Show that $f(t,T_1)$ is a martingale under the forward probability measure $\mathbb{Q}^\star(T_1)$.

6. We recall that the price of the zero-coupon bond satisfies the SDE:

$$\mathrm{d}B(t,T) = r(t)\, B(t,T)\, \mathrm{d}t + b(t,T)\, B(t,T)\, \mathrm{d}W^{\mathbb{Q}}(t)$$

   (a) Show that:

$$\frac{B(t,T_2)}{B(t,T_1)} = \frac{B(s,T_2)}{B(s,T_1)} e^{X(s,t)}$$

   where $X(s,t)$ is a random variable to define.

   (b) Deduce that $B(t,T_2)/B(t,T_1)$ is a martingale under $\mathbb{Q}^\star(T_1)$.

9.4.6 Equivalent martingale measure in the Libor market model

Let $L_i(t) = L(t,T_i,T_{i+1})$ be the forward Libor rate when resetting and maturity dates are respectively equal to T_i and T_{i+1}. Under the forward probability measure $\mathbb{Q}^\star(T_{i+1})$, the dynamics of $L_i(t)$ is given by the following SDE:

$$\mathrm{d}L_i(t) = \gamma_i(t)\, L_i(t)\, \mathrm{d}W_i^{\mathbb{Q}^\star(T_{i+1})}(t)$$

1. Using the definition of the Libor rate, find the relationship between $B(t,T_{j+1})/B(t,T_j)$ and $L_j(t)$. Let $T_{k+1} > T_{i+1}$. Deduce an expression of the ratio:

$$\frac{B(t,T_{k+1})}{B(t,T_{i+1})}$$

in terms of Libor rates $L_j(t)$ $(j = i+1,\ldots,k)$.

2. We change the probability measure from $\mathbb{Q}^\star(T_{i+1})$ to $\mathbb{Q}^\star(T_{k+1})$. Define the numéraires $M(t)$ and $N(t)$ associated to $\mathbb{Q}^\star(T_{i+1})$ to $\mathbb{Q}^\star(T_{k+1})$. Deduce an expression of $Z(t)$:

$$Z(t) = \frac{\mathrm{d}\mathbb{Q}^\star(T_{k+1})}{\mathrm{d}\mathbb{Q}^\star(T_{i+1})}$$

in terms of Libor rates $L_j(t)$ $(j = i+1,\ldots,k)$.

3. Calculate $\mathrm{d}\ln Z(t)$.

4. Calculate the drift ζ defined by:
$$\zeta = \left\langle \frac{\mathrm{d}L_i(t)}{L_i(t)}, \mathrm{d}\ln Z(t) \right\rangle$$

5. Show that the dynamics of $L_i(t)$ under the forward probability measure $\mathbb{Q}^\star(T_{k+1})$ is given by:
$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \mu_{i,k}(t)\,\mathrm{d}t + \gamma_i(t)\,\mathrm{d}W_k^{\mathbb{Q}^\star(T_{k+1})}(t)$$
where $\mu_{i,k}(t)$ is a drift to determine.

6. What does the previous results become if $T_{k+1} < T_{i+1}$?

9.4.7 Displaced diffusion option pricing

Brigo and Mercurio (2002a) consider the diffusion process $X(t)$ given by:
$$\begin{cases} \mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W^{\mathbb{Q}}(t) \\ X(0) = X_0 \end{cases}$$
They assume that the asset price $S(t)$ is an affine transformation of $X(t)$:
$$S(t) = \alpha(t) + \beta(t) \cdot X(t)$$
where $\beta(t) > 0$.

1. By applying Itô's lemma to $S(t)$, find the condition on $\alpha(t)$ and $\beta(t)$ in order to satisfy the martingale condition:
$$\mathbb{E}^{\mathbb{Q}}\left[e^{-bt} \cdot S(t) \mid \mathcal{F}_0\right] = S_0$$
where b is the cost-of-carry parameter.

2. We consider the CEV process:
$$\mathrm{d}X(t) = \mu(t)\,X(t)\,\mathrm{d}t + \sigma(t)\,X(t)^\gamma\,\mathrm{d}W^{\mathbb{Q}}(t)$$
where $\gamma \in [0,1]$. Show that the solutions of $\alpha(t)$ and $\beta(t)$ are:
$$\begin{cases} \alpha(t) = \alpha_0 \cdot \exp(bt) \\ \beta(t) = \beta_0 \cdot \exp\left(\int_0^t (b - \mu(s))\,\mathrm{d}s\right) \end{cases}$$

3. Deduce the SDE of $S(t)$.

4. We consider the case $\gamma = 1$. Give the SDE of $X(t)$. Calculate the solutions of $X(t)$ and $S(t)$.

5. Give the price of the European call option, whose payoff is equal to $(S(T) - K)^+$.

6. We now assume that $\sigma(t) = \sigma$.
   (a) Using the formula of Lee and Wang (2012), give an approximation of the implied volatility $\Sigma(T, K)$.
   (b) Calculate the volatility skew:
   $$\omega(T, K) = \frac{\partial\,\Sigma(T, K)}{\partial\,K}$$
   (c) Give the price of the binary call option in the case of the BS model.
   (d) Deduce the BCC price when we consider the SLN model.
   (e) Give an approximation of the BCC price based on the implied volatility skew.

9.4.8 Dupire local volatility model

We assume that:

$$\mathrm{d}S(t) = bS(t)\,\mathrm{d}t + \sigma(t, S(t))\,S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

1. Give the forward equation for pricing the call option $\mathcal{C}(T,K)$. Deduce the expression of the local variance $\sigma^2(T,K)$.
2. Using the Black-Scholes formula, find the relationship between the local volatility $\sigma(T,K)$ and the implied volatility $\Sigma(T,K)$.
3. We consider the discounted payoff function:

   $$\tilde{f}(T, S(T)) = e^{-r(T-t)}(S(T) - K)^+$$

   Using Itô's lemma, calculate the derivative of the call option with respect to the maturity:

   $$\partial_T \mathcal{C}(T,K) = \frac{\mathbb{E}\left[\mathrm{d}\tilde{f}(T, S(T)) \middle| \mathcal{F}_t\right]}{\mathrm{d}T}$$

4. Calculate $\partial_K \mathcal{C}(T,K)$ and $\partial_K^2 \mathcal{C}(T,K)$ using the discounted payoff function. Retrieve the forward equation[84] of Dupire (1994).
5. We introduce the log-moneyness x:

   $$\begin{aligned} x &= \varphi(T,K) \\ &= \ln\frac{S_0}{K} + bT \end{aligned}$$

   and the functions $\tilde{\sigma}(T,x)$ and $\tilde{\Sigma}(T,x)$, which are defined by the relationships:

   $$\Sigma(T,K) = \tilde{\Sigma}(T, \varphi(T,K))$$

   and:

   $$\sigma(T,K) = \tilde{\sigma}(T, \varphi(T,K))$$

   (a) Calculate d_1, d_2 and $d_1 d_2$.

   (b) Write the derivatives $\partial_K \Sigma(T,K)$, $\partial_T \Sigma(T,K)$ and $\partial_K^2 \Sigma(T,K)$ using the variables T and x.

   (c) Deduce the relationship between $\tilde{\sigma}(T,x)$ and $\tilde{\Sigma}(T,x)$.

   (d) Show that:

   $$\partial_x \tilde{\Sigma}(0,0) = \frac{1}{2}\partial_x \tilde{\sigma}(0,0)$$

9.4.9 The stochastic normal model

Let $F(t)$ be the forward rate. We assume that the dynamics of $F(t)$ is given by the SABR model:

$$\begin{cases} \mathrm{d}F(t) = \alpha(t)\,F(t)^{\beta}\,\mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}\alpha(t) = \nu\alpha(t)\,\mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

where $\mathbb{E}\left[W_1^{\mathbb{Q}}(t)\,W_2^{\mathbb{Q}}(t)\right] = \rho t$. In what follows, we consider the special case $\beta = 0$.

[84] This approach has also been proposed by Derman *et al.* (1996).

1. How to transform the Black volatility $\Sigma_B(T,K)$ into the implied normal volatility $\Sigma_N(T,K)$?

2. Give the expression of the implied normal volatility[85] $\Sigma_N(T,K)$ for the general case $\beta \in [0,1]$.

3. Deduce the formula of $\Sigma_N(T,K)$ when $\beta = 0$.

4. What is the ATM normal volatility?

5. Calculate $\partial_K \Sigma_N(T,K)$.

6. Recall the price of the call option for the normal model, whose volatility is σ_N.

7. We now assume that σ_N is equal to the SABR normal volatility $\Sigma_N(T,K)$. Deduce the cumulative distribution function of $F(T)$.

8. By considering the following approximation[86]:

$$\sqrt{F_0 K}\ln\frac{F_0}{K} \simeq F_0 - K$$

calculate the probability density function of $F(T)$.

9. Show that:

$$F(t) = F_0 + \frac{\alpha}{\nu}\int_0^{\nu^2 t} \exp\left(-\frac{1}{2}s + W_2(s)\right) \mathrm{d}W_1(s)$$

where $W_1(t)$ and $W_2(t)$ have the same properties as $W_1^{\mathbb{Q}}(t)$ and $W_2^{\mathbb{Q}}(t)$.

10. We note:

$$X(t) = \int_0^t \exp\left(-\frac{1}{2}s + W_2(s)\right) \mathrm{d}W_1(s)$$

and:

$$M^a(t) = \exp\left(-\frac{1}{2}at + aW_2(t)\right)$$

Let us introduce the function $\Psi^{n,a}(t)$:

$$\Psi^{n,a}(t) = \mathbb{E}\left[X^n(t)\, M^a(t)\right]$$

where $n \in \mathbb{N}$ and $a \in \mathbb{R}_+$. Verify that $\Psi^{n,a}(t)$ satisfies the ordinary differential equation:

$$\frac{\mathrm{d}\Psi^{n,a}(t)}{\mathrm{d}t} = \frac{a(a-1)}{2}\Psi^{n,a}(t) + n\rho a\Psi^{n-1,a+1}(t) + \frac{n(n-1)}{2}\Psi^{n-2,a+2}(t)$$

where $\Psi^{n,a}(0) = 0$. What is the link between $\Psi^{n,a}(t)$ and the statistical moments of $F(t)$?

11. Calculate $\Psi^{0,a}(t)$, $\Psi^{1,a}(t)$, $\Psi^{2,a}(t)$, $\Psi^{3,0}(t)$ and $\Psi^{4,0}(t)$. Deduce the first four central moments of $F(t)$.

12. Calculate an approximation of the volatility, skewness and kurtosis of $F(t)$ when $t \simeq 0$.

[85] Hagan *et al.* (2002) calculate this expression in Appendix A.4 on page 102.
[86] Hagan *et al.* (2002), Equations (A67b) and (A68a), page 102.

13. We assume that $F_0 = 10\%$ and $T = 1$, and we consider the following smile:

K	7%	10%	13%
$\Sigma_B(T,K)$	30%	20%	30%

    (a) Calculate the equivalent normal volatility $\Sigma_N(T,K)$.

    (b) Calibrate the parameters of the stochastic normal model.

    (c) Draw the cumulative distribution function of $F(T)$. What is the problem?

    (d) Draw the probability density function of $F(T)$ when we consider the approximation $\sqrt{F_0 K}\ln\dfrac{F_0}{K} \simeq F_0 - K$.

    (e) Calculate the skewness and the kurtosis of $F(T)$. Comment on these results.

9.4.10 The quadratic Gaussian model

We consider the quadratic Gaussian model:

$$r(t) = \alpha(t) + \beta(t)^\top X(t) + X(t)^\top \Gamma(t) X(t)$$

where the state variables $X(t)$ follow an Ornstein-Uhlenbeck process:

$$\mathrm{d}X(t) = (a(t) + B(t) X(t))\,\mathrm{d}t + \Sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

1. Find the PDE associated to the zero-coupon bond $B(t,T)$.

2. We assume that the solution of $B(t,T)$ has the following form:

$$B(t,T) = \exp\left(-\hat{\alpha}(t,T) - \hat{\beta}(t,T)^\top X(t) - X(t)^\top \hat{\Gamma}(t,T) X(t)\right)$$

    where $\hat{\Gamma}(t,T)$ is a symmetric matrix. Show that $\hat{\alpha}(t,T)$, $\hat{\beta}(t,T)$ and $\hat{\Gamma}(t,T)$ satisfy a system of ODEs.

3. Find a condition that $\hat{\Gamma}(t,T)$ is a symmetric matrix. Why do we need this hypothesis?

4. Let $\mathbb{Q}^\star(T)$ be the forward probability measure. Recall the dynamics of $X(t)$ under $\mathbb{Q}^\star(T)$. Using the explicit solution, demonstrate that $X(t)$ is Gaussian:

$$X(t) \sim \mathcal{N}(m(0,t), V(0,t))$$

    Find the dynamics of $m(0,t)$ and $V(0,t)$. Compare these results with those obtained by El Karoui *et al.* (1992a).

5. Define the Libor rate $L(t, T_{i-1}, T_i)$.

6. Demonstrate that the pricing formula of the caplet is equal to:

$$\text{Caplet} = B(0,t) \cdot \mathbb{E}^{\mathbb{Q}^\star(t)}[\max(0, g(X))]$$

    where $\mathbb{Q}^\star(t)$ is the forward probability measure and $g(x)$ is a function to define.

7. Show that:

$$\text{Caplet} = B(0,t) \int_{\mathcal{E}} h(x)\,\mathrm{d}x$$

    where $h(x) = g(x)\,\phi(x; m(0,t), V(0,t))$ and $\mathcal{E}$ is a set to define.

8. We consider the following function:

$$\mathcal{J}(a,b,c,m,V,x_1,x_2) = \int_{x_1}^{x_2} \frac{e^{-ax^2-bx-c}}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(x-m)^2}\,\mathrm{d}x$$

   Find the analytical expression of $\mathcal{J}$.

9. Deduce the analytical expression of the caplet.

9.4.11 Pricing two-asset basket options

We assume that the risk-neutral dynamics of $S_1(t)$ and $S_2(t)$ are given by:

$$\begin{cases} \mathrm{d}S_1(t) = b_1 S_1(t)\,\mathrm{d}t + \sigma_1 S_1(t)\,\mathrm{d}W_1^{\mathbb{Q}}(t) \\ \mathrm{d}S_2(t) = b_2 S_2(t)\,\mathrm{d}t + \sigma_2 S_2(t)\,\mathrm{d}W_2^{\mathbb{Q}}(t) \end{cases}$$

where $W_1^{\mathbb{Q}}(t)$ and $W_2^{\mathbb{Q}}(t)$ are two correlated Brownian motions:

$$\mathbb{E}\left[W_1^{\mathbb{Q}}(t)\,W_2^{\mathbb{Q}}(t)\right] = \rho\,t$$

1. By considering the following payoff $(\alpha_1 S_1(T) + \alpha_2 S_2(T) - K)^+$, show that the price of the option can be expressed as a double integral.

2. We consider the computation of $I = \mathbb{E}\left[\left(Ae^{b+c\cdot\varepsilon} - D\right)^+\right]$ where $\varepsilon \sim \mathcal{N}(0,1)$, and A, b, c and D are four scalars.

   (a) Find the value of I when $A > 0$ and $D > 0$.

   (b) Deduce the value of I in the other cases.

3. We assume that $\alpha_1 < 0$, $\alpha_2 > 0$ and $K > 0$. Using the Cholesky decomposition, reduce the computation of the double integral to a single integral.

4. Extend this result to the case $\alpha_1 > 0$, $\alpha_2 < 0$ and $K > 0$.

5. Discuss the general case.

Chapter 10

Statistical Inference and Model Estimation

In this chapter, we present the statistical tools used in risk management. The first section concerns estimation methods that are essential to calibrate the parameters of a statistical model. This includes the linear regression, which is the standard statistical tool to investigate the relationships between data in empirical research, and the method of maximum likelihood (ML), whose goal is to estimate parameters of non-linear and non-Gaussian financial models. We also present the generalized method of moments (GMM), which is very popular in economics because we can calibrate non-reduced forms or structural models. Finally, the last part of the first section is dedicated to non-parametric estimators. In the second section, we study time series modeling, in particular ARMA processes and error correction models. We also investigate state-space models, which encompass many dynamic models. A focus is also done on volatility modeling, which is an important issue in risk management. Finally, we discuss the application of spectral analysis. Most of statistical tools presented in this chapter are used in the next chapters, for example the estimation of copula models, the calibration of stressed scenarios or the implementation of credit scoring.

10.1 Estimation methods

10.1.1 Linear regression

Let Y and X be two random vectors. We consider the conditional expectation problem:

$$y = \mathbb{E}\left[Y \mid X = x\right] = m\left(x\right) \tag{10.1}$$

The underlying idea is to find an estimate $\hat{m}\left(x\right)$ of the function $m\left(x\right)$. In the general case, this problem is extremely difficult to solve. However, if (Y, X) is a Gaussian random vector, the function $m\left(x\right)$ can then be determined by considering the Gaussian linear model:

$$Y = \beta^{\top} X + u \tag{10.2}$$

where $u \sim \mathcal{N}\left(0, \sigma^2\right)$. Most of the time, the joint distribution of (Y, X) is unknown. In this case, the linear model is estimated by applying least squares techniques to a given sample $(\mathbf{Y}, \mathbf{X})$:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

Remark 118 *In order to distinguish random variables and observations, we write matrices and vectors that are related to observations in bold style.*

10.1.1.1 Least squares estimation

Derivation of the OLS estimator We consider a training set of n *iid* samples (y_i, x_i). For the i^{th} observation, we have:

$$y_i = \sum_{k=1}^{K} \beta_k x_{i,k} + u_i \tag{10.3}$$

The least squares estimate of the parameter vector β is defined as follows:

$$\hat{\beta} = \arg\min \sum_{i=1}^{n} u_i^2$$

We introduce the following matrix notations: $\mathbf{Y}$ is the $n \times 1$ vector with elements $\mathbf{Y}_i = y_i$, $\mathbf{X}$ is the $n \times K$ matrix defined as follows:

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & & x_{1,K} \\ & \ddots & \\ x_{n,1} & & x_{n,K} \end{pmatrix}$$

and $\mathbf{U}$ is the $n \times 1$ vector with elements $\mathbf{U}_i = u_i$. In this case, the system of equations (10.3) becomes:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U} \tag{10.4}$$

Let $\text{RSS}(\beta)$ be the residual sum of squares. We have:

$$\begin{aligned} \text{RSS}(\beta) &= \sum_{i=1}^{n} u_i^2 \\ &= \mathbf{U}^\top \mathbf{U} \\ &= \mathbf{Y}^\top \mathbf{Y} - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \end{aligned}$$

The least squares estimator verifies the set of normal equations $\partial_\beta \mathbf{U}^\top \mathbf{U} = \mathbf{0}$ and we deduce that $-2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\hat{\beta} = \mathbf{0}$. The expression of the least squares estimator is then:

$$\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \tag{10.5}$$

To obtain the expression of $\hat{\beta}$, we only need the assumption that the rank of the matrix $\mathbf{X}$ is K. In this case, $\hat{\beta}$ is the solution of the least squares problem. To go further, we assume that (Y, X) is a Gaussian random vector. The solution of the conditional expectation problem $\mathbb{E}[Y \mid X = x] = m(x)$ is then:

$$\begin{aligned} \hat{m}(x) &= x^\top \hat{\beta} \\ &= x^\top \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \end{aligned}$$

It means that the prediction of Y given that $X = x$ is equal to $\hat{y} = x^\top \hat{\beta}$. If we consider the training data $\mathbf{X}$, we obtain:

$$\begin{aligned} \hat{Y} &= \hat{m}(\mathbf{X}) \\ &= \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= \mathbf{H}\mathbf{Y} \end{aligned}$$

where $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$ is called the '*hat*' matrix[1]. We notice that $\hat{m}(\mathbf{X})$ is a linear predictor of $\mathbf{Y}$.

Statistical inference Because (Y, X) is a Gaussian random vector, it implies that $u = Y - \beta^\top X$ is a Gaussian random variable. We notice that:

$$\begin{aligned}\hat{\beta} &= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\\ &= \beta + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{U}\end{aligned}$$

By assuming the exogeneity of the variables X – meaning that $\mathbb{E}[u \mid X = x] = 0$ – we deduce that $\hat{\beta}$ is an unbiased estimator:

$$\begin{aligned}\mathbb{E}\left[\hat{\beta}\right] &= \beta + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbb{E}\left[\mathbf{X}^\top\mathbf{U}\right]\\ &= \beta\end{aligned}$$

We recall that $\mathbf{U} \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$. It follows that:

$$\begin{aligned}\text{var}\left(\hat{\beta}\right) &= \mathbb{E}\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^\top\right]\\ &= \mathbb{E}\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{U}\mathbf{U}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right]\\ &= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbb{E}\left[\mathbf{U}\mathbf{U}^\top\right]\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\\ &= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\sigma^2 I_n\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\\ &= \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\end{aligned}$$

We conclude that:

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)$$

In most cases, σ^2 is unknown and we have to estimate it. The vector of residuals is:

$$\begin{aligned}\hat{\mathbf{U}} &= \mathbf{Y} - \hat{\mathbf{Y}}\\ &= \mathbf{Y} - \mathbf{X}\hat{\beta}\end{aligned}$$

We notice that $\mathbb{E}\left[\hat{\mathbf{U}}\right] = \mathbf{0}$ and $\text{var}\left(\hat{\mathbf{U}}\right) = \sigma^2\left(I_n - \mathbf{H}\right)$. Because $\text{RSS}\left(\hat{\beta}\right) = \hat{\mathbf{U}}^\top\left(I_n - \mathbf{H}\right)\hat{\mathbf{U}}$ is a quadratic form, we can show that:

$$\hat{\sigma}^2 = \frac{\text{RSS}\left(\hat{\beta}\right)}{n - K}$$

[1] We interpret $\mathbf{H}$ as the orthogonal projection matrix generated by $\mathbf{X}$ implying that $\mathbf{H}$ is idempotent, that is $\mathbf{H}\mathbf{H} = \mathbf{H}$. Indeed, we have:

$$\begin{aligned}\mathbf{H}\mathbf{H} &= \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\\ &= \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\\ &= \mathbf{H}\end{aligned}$$

is an unbiased estimator of σ^2 and $\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-K}$. In order to measure the model quality, we consider the coefficient of determination or R^2_c. It is defined as follows:

$$R^2_c = 1 - \frac{\text{RSS}\left(\hat{\beta}\right)}{\text{TSS}}$$

where $\text{TSS} = \sum_{i=1}^n (y_i - \bar{y})$ is the total sum of squares. We have $R^2_c \leq 1$. A high (resp. low) level indicates a good (resp. bad) goodness-of-fit of the regression model.

Example 100 *We consider the data given in Table 10.1. We would like to explain the dependent variable y_i by four explanatory variables x_1, x_2, x_3 and x_4. There are 10 observations and we note that x_1 is in fact a constant.*

TABLE 10.1: Data of the linear regression problem

i	y	x_1	x_2	x_3	x_4
1	1.5	1.0	2.4	3.6	0.3
2	20.4	1.0	1.1	3.8	5.9
3	17.1	1.0	5.1	6.3	6.1
4	30.9	1.0	2.7	2.4	9.5
5	22.2	1.0	3.3	3.0	7.4
6	9.1	1.0	1.0	5.4	4.9
7	39.2	1.0	9.6	2.8	8.1
8	3.1	1.0	2.9	4.4	1.0
9	7.2	1.0	4.2	5.6	1.7
10	27.6	1.0	8.1	1.7	5.4

We consider the linear regression model:

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + u_i$$

It follows that:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 10.000 & 40.400 & 39.000 & 50.300 \\ 40.400 & 235.980 & 143.660 & 224.830 \\ 39.000 & 143.660 & 172.460 & 179.170 \\ 50.300 & 224.830 & 179.170 & 339.790 \end{pmatrix}$$

and:

$$\mathbf{X}^\top \mathbf{Y} = \begin{pmatrix} 178.300 \\ 918.150 \\ 591.190 \\ 1209.440 \end{pmatrix}$$

We deduce that the estimates are:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 3.446 \\ 1.544 \\ -1.645 \\ 2.895 \end{pmatrix}$$

We can then compute the residuals. We obtain $\hat{u}_1 = -0.597$, $\hat{u}_2 = 4.427$, etc. The sum of squared residuals is equal to $\text{RSS}\left(\hat{\beta}\right) = 40.184$, which implies that $\hat{\sigma} = 2.588$. Therefore,

the estimate of the covariance matrix of $\hat{\beta}$ is:

$$\text{cov}\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} = \begin{pmatrix} 15.353 & -0.602 & -2.263 & -0.682 \\ -0.602 & 0.108 & 0.061 & -0.015 \\ -2.263 & 0.061 & 0.428 & 0.069 \\ -0.682 & -0.015 & 0.069 & 0.094 \end{pmatrix}$$

For this linear regression, the coefficient of determination is equal to:

$$R_c^2 = 1 - \frac{40.184}{1422.041} = 97.17\%$$

By construction, the standard errors $\sigma\left(\hat{\beta}_k\right)$ of the estimator $\hat{\beta}_k$ is the square root of the k^{th} diagonal element of $\text{cov}\left(\hat{\beta}\right)$. The assumption $\mathcal{H}_0 : \hat{\beta}_k = b_k$ is then tested by computing the t-statistic:

$$t = \frac{\hat{\beta}_k - b_k}{\sigma\left(\hat{\beta}_k\right)} \sim \mathbf{t}_{n-K}$$

and the associated p-value[2]:

$$p = 2\left(1 - \mathbf{t}_{n-K}\left(|t|\right)\right)$$

For instance, we report the t-statistic and the p-value associated to the hypothesis $\mathcal{H}_0 : \hat{\beta}_k = 0$ in Table 10.2. We cannot reject this assumption for the estimate $\hat{\beta}_1$ at the 10% confidence level, meaning that $\hat{\beta}_1$ is not significant.

TABLE 10.2: Results of the linear regression

Parameter	Estimate	Standard error	t-statistic	p-value
β_1	3.4461	3.9183	0.8795	0.4130
β_2	1.5442	0.3289	4.6943	0.0033
β_3	−1.6454	0.6543	−2.5146	0.0457
β_4	2.8951	0.3071	9.4264	0.0001

Gauss-Markov theorem Let $\tilde{\beta} = \mathbf{AY}$ be a linear estimator. The Gauss-Markov theorem states that, among all linear unbiased estimators of β, $\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y}$ has the smallest variance:

$$\text{var}\left(\tilde{\beta}\right) \geq \text{var}\left(\hat{\beta}\right)$$

In this case, we say that $\hat{\beta}$ is BLUE (best linear unbiased estimator). Let us write $\tilde{\beta}$ as follows:

$$\begin{aligned} \tilde{\beta} &= \mathbf{AY} \\ &= \mathbf{AX}\beta + \mathbf{AU} \end{aligned}$$

We have:

$$\begin{aligned} \mathbb{E}\left[\tilde{\beta}\right] &= \mathbb{E}\left[\mathbf{AY}\right] \\ &= \mathbf{AX}\beta + \mathbb{E}\left[\mathbf{AU}\right] \\ &= \mathbf{AX}\beta \end{aligned}$$

[2]The p-value is the estimated probability of rejecting the null hypothesis $\mathcal{H}_0$.

We deduce that $\tilde{\beta}$ is unbiased if $\mathbf{AX} = I_K$. We also notice that:

$$\begin{aligned}\text{var}\left(\tilde{\beta}\right) &= \mathbb{E}\left[\left(\tilde{\beta}-\beta\right)\left(\tilde{\beta}-\beta\right)^\top\right] \\ &= \mathbb{E}\left[\mathbf{A}\mathbf{U}\mathbf{U}^\top\mathbf{A}^\top\right] \\ &= \sigma^2\left(\mathbf{A}\mathbf{A}^\top\right)\end{aligned}$$

We set $\mathbf{A} = \mathbf{B} + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$. We have $\mathbf{BX} = \mathbf{0}$ because $\mathbf{AX} = I_K$. It follows that:

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= \left(\mathbf{B} + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)\left(\mathbf{B} + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)^\top \\ &= \mathbf{B}\mathbf{B}^\top + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{B}^\top + \mathbf{B}\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \\ &= \mathbf{B}\mathbf{B}^\top + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\end{aligned}$$

Because the matrix $\mathbf{B}\mathbf{B}^\top$ is positive semi-definite, we finally deduce that:

$$\begin{aligned}\text{var}\left(\tilde{\beta}\right) &= \sigma^2\left(\mathbf{A}\mathbf{A}^\top\right) \\ &= \sigma^2\left(\mathbf{B}\mathbf{B}^\top + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right) \\ &\geq \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\end{aligned}$$

10.1.1.2 Relationship with the conditional normal distribution

Let us consider a Gaussian random vector defined as follows:

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \Sigma_{y,y} & \Sigma_{y,x} \\ \Sigma_{x,y} & \Sigma_{x,x} \end{pmatrix}\right)$$

On page 1062, we show that the conditional distribution of Y given $X = x$ is a multivariate normal distribution where:

$$\begin{aligned}\mu_{y|x} &= \mathbb{E}\left[Y \mid X = x\right] \\ &= \mu_y + \Sigma_{y,x}\Sigma_{x,x}^{-1}\left(x - \mu_x\right)\end{aligned}$$

and:

$$\begin{aligned}\Sigma_{y,y|x} &= \sigma^2\left[Y \mid X = x\right] \\ &= \Sigma_{y,y} - \Sigma_{y,x}\Sigma_{x,x}^{-1}\Sigma_{x,y}\end{aligned}$$

We deduce that:

$$Y = \mu_y + \Sigma_{y,x}\Sigma_{x,x}^{-1}\left(x - \mu_x\right) + u$$

where u is a centered Gaussian random variable with variance $\sigma^2 = \Sigma_{y,y|x}$. It follows that:

$$Y = \underbrace{\left(\mu_y - \Sigma_{y,x}\Sigma_{x,x}^{-1}\mu_x\right)}_{\beta_0} + \underbrace{\Sigma_{y,x}\Sigma_{x,x}^{-1}}_{\beta^\top}x + u \tag{10.6}$$

We recognize the linear regression of Y on a constant and a set of exogenous variables X:

$$Y = \beta_0 + \beta^\top X + u$$

Moreover, we have:

$$\begin{aligned} R_c^2 &= 1 - \frac{\sigma^2}{\Sigma_{y,y}} \\ &= \frac{\Sigma_{y,x}\Sigma_{x,x}^{-1}\Sigma_{x,y}}{\Sigma_{y,y}} \end{aligned}$$

Example 101 *We consider a Gaussian random vector $X = (X_1, X_2, X_3, X_4)$. The expected values are equal to $\mu_1 = 2$, $\mu_2 = 5$, $\mu_3 = -4$ and $\mu_4 = 3$ whereas the standard deviations are equal to $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 0.5$ and $\sigma_4 = 1$. The correlation between the random variables is given by the following matrix:*

$$\rho = \begin{pmatrix} 1.00 & & & \\ 0.90 & 1.00 & & \\ 0.70 & 0.40 & 1.00 & \\ 0.60 & 0.50 & 0.30 & 1.00 \end{pmatrix}$$

For each random variable X_i, we can compute the conditional Gaussian regression using the previous formulas:

$$X_i = \beta_0 + \sum_{k \neq i} \beta_k X_k + u$$

Results are reported in Table 10.3. For example, it means the linear regression of X_1 on X_2, X_3 and X_4 is:

$$X_1 = 2.974 + 0.335 \cdot X_2 + 0.774 \cdot X_3 + 0.148 \cdot X_4 + u$$

where $u \sim \mathcal{N}\left(0, 0.19^2\right)$ and the associated R_c^2 is equal to 96.39%.

TABLE 10.3: Results of the conditional Gaussian regression

Y	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	σ	R_c^2
X_1	2.974		0.335	0.774	0.148	19.01%	96.39%
X_2	−7.205	2.667		−1.949	−0.308	53.59%	92.82%
X_3	−4.017	1.000	−0.317		−0.133	21.60%	81.33%
X_4	−4.273	2.091	−0.545	−1.455		71.35%	49.09%

Remark 119 *The previous analysis raises the question of the status of the variables Y and X in the linear regression framework. In this model, Y is called the dependent variable and X are called the independent (or explanatory) variables. This implies that there is a relationship **from** X **to** Y. In some way, linear regression has a strong connotation of an explicit directional (or causal) relationship. However, the previous example shows clearly that linear regression does not mean causality!*

Indeed, linear regression may be viewed as another way to interpret the correlation between random variables. Let us consider the case where X is a one-dimension random variable. We note $\rho_{x,y}$ the correlation between X and Y whereas σ_x and σ_y are their standard deviations. In this case, we have:

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \rho_{x,y}\sigma_x\sigma_y \\ \rho_{x,y}\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}$$

The conditional Gaussian regression (10.6) becomes:

$$Y = \beta_0 + \beta X + u$$

where:

$$\begin{aligned}\beta &= \frac{\rho_{x,y}\sigma_x\sigma_y}{\sigma_x^2} \\ &= \frac{\rho_{x,y}\sigma_y}{\sigma_x}\end{aligned}$$

and:

$$\begin{aligned}\beta_0 &= \mu_y - \frac{\rho_{x,y}\sigma_x\sigma_y}{\sigma_x^2}\mu_x \\ &= \mu_y - \beta\mu_x\end{aligned}$$

We also deduce that the expression of the R_c^2 statistic is:

$$\begin{aligned}R_c^2 &= \frac{(\rho_{x,y}\sigma_x\sigma_y)^2}{\sigma_x^2\sigma_y^2} \\ &= \rho_{x,y}^2\end{aligned}$$

The coefficient of determination is then the square of the correlation coefficient, meaning that their significance is not of the same magnitude. Indeed, a value of 50% for the R_c^2 statistic corresponds to a value of 70% for the correlation.

The previous analysis shows that:

$$\beta = \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}$$

The single risk factor model of Sharpe (1964) exploits this result since we have:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + u_{i,t}$$

or equivalently:

$$\beta_i = \frac{\operatorname{cov}(R_i, R_m)}{\operatorname{var}(R_i)}$$

where $R_{i,t}$ is the asset's return and $R_{m,t}$ is the market's return.

10.1.1.3 The intercept problem

Example 102 *We consider the following data with* 20 *observations:*

y_i	13.9	11.5	14.9	14.6	13.7	17.3	18.1	14.8	14.7	14.7
x_i	4.4	3.4	4.9	4.4	4.4	6.3	6.6	4.8	4.5	4.6
y_i	16.1	14.6	15.2	18.1	11.5	14.0	18.4	15.0	12.0	14.8
x_i	5.7	5.0	5.2	6.5	3.2	4.5	6.9	5.2	3.1	4.9

We want to explain the dependent variable Y by the explanatory variable X.

If we include a constant in the regression model, we obtain:

$$Y = 1.8291 + 5.8868 \cdot X + u \quad \text{with} \quad u \sim \mathcal{N}\left(0, 0.3522^2\right) \tag{10.7}$$

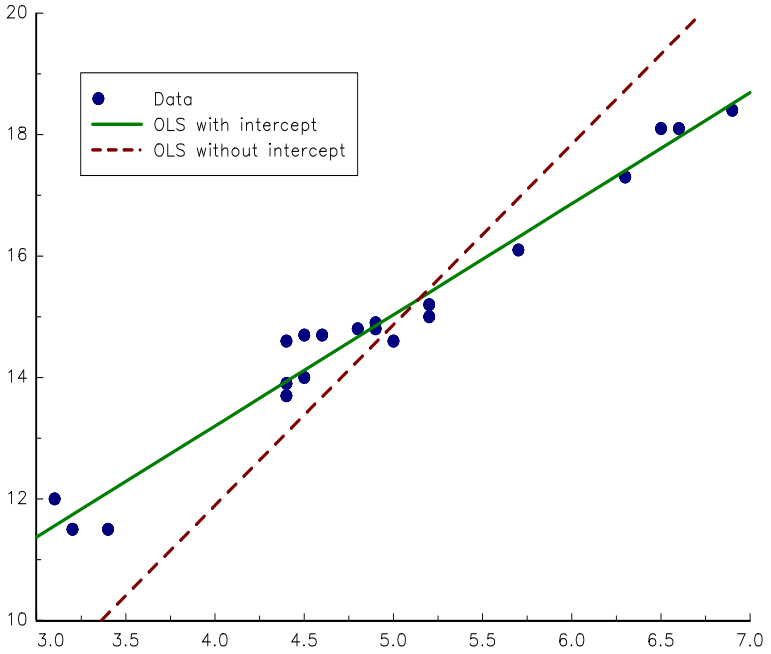


FIGURE 10.1: Illustration of the intercept problem

Without the intercept, the linear regression becomes:

$$Y = 2.9730 \cdot X + u \quad \text{with} \quad u \sim \mathcal{N}\left(0.2529, 1.2980^2\right) \tag{10.8}$$

The corresponding fitted curves are reported in Figure 10.1. By omitting the constant, we have overestimated the slope β of the curve.

The previous example shows that a linear regression is valid only if we include the intercept in the model. Indeed, without the intercept, the residuals are not centered: $\mathbb{E}[u] \neq 0$. If we consider the conditional Gaussian regression, we have the relationship $\beta_0 = \mu_y - \beta^\top \mu_x$. By omitting the constant – $Y = \beta^\top X + u$, we set $\beta = 0$ and the previous relationship does not hold any more. In this case, the residuals incorporate the mean effect of the data. Indeed, we have $\mathbb{E}[Y] = \beta^\top \mathbb{E}[X] + \mathbb{E}[u]$ meaning that $\mathbb{E}[u] = \mu_y - \beta^\top \mu_x$ is not necessarily equal to zero (see Exercise 10.3.2 on page 705).

Remark 120 *Let us consider the linear regression with a constant. We have* $Y = \beta_0 + \beta^\top X + u$. *It follows that* $\mu_y = \beta_0 + \beta^\top \mu_x$ *because the residuals are centered. We deduce that:*

$$Y - \mu_y = \beta^\top (X - \mu_x) + u$$

By considering the centered data instead of the original data, the intercept problem vanishes. This type of transformation is common in statistics and finance. Generally, raw data have to be analyzed and modified if we want to obtain more robust relationships. Normalizing, using logarithmic scale or creating dummy variables are some examples of data processing.

10.1.1.4 Coefficient of determination

We have defined the coefficient of determination as follows:

$$R_c^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

We can show that $R_c^2 \leq 1$. If one of the independent variables is a constant, the linear regression model becomes:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k} + u_i$$

In this case, we have $0 \leq R_c^2 \leq 1$. Testing the hypothesis $\mathcal{H}_0 : \beta_1 = \ldots = \beta_K = 0$ is equivalent to consider the Fisher test:

$$\frac{(n-K)\,R_c^2}{(K-1)\,(1-R_c^2)} \sim \mathfrak{F}\,(K-1, T-K)$$

Therefore, the use of R_c^2 is valid when there is a constant in the linear regression. It measures the significance of the model versus the naive model: $y_i = \bar{y} + u_i$. If the constant is omitted, R_c^2 can take negative value and is not pertinent. In this case, it is better to use the uncentered coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n {y_i}^2}$$

We can show that $0 \leq R^2 \leq 1$. One of the drawbacks when using R^2 or R_c^2 comes from the fact that the coefficient of determination increases with the number of exogenous variables. The more the number of explanatory variables, the more the R-squared. To correct this effect, we can use adjusted coefficients of determination:

$$\bar{R}^2 = 1 - \frac{\left(\sum_{i=1}^n u_i^2\right)/(n-K)}{\left(\sum_{i=1}^n y_i^2\right)/n}$$

and:

$$\bar{R}_c^2 = 1 - \frac{\left(\sum_{i=1}^n u_i^2\right)/(n-K)}{\left(\sum_{i=1}^n (y_i - \bar{y})^2\right)/(n-1)}$$

10.1.1.5 Extension to weighted least squares regression

Definition The weighted least squares (or WLS) estimator is defined by:

$$\hat{\beta} = \arg\min \sum_{i=1}^n w_i u_i^2$$

where w_i is the weight associated to the i^{th} observation. It is obvious that the analytical solution is:

$$\hat{\beta} = \left(\mathbf{X}^\top W \mathbf{X}\right)^{-1} \mathbf{X}^\top W \mathbf{Y}$$

where W is a diagonal matrix with $W_{i,i} = w_i$.

Robust regression Let us consider the least squares problem:

$$\hat{\beta} = \arg\min \sum_{i=1}^n \rho\,(u_i) = \arg\min \sum_{i=1}^n \rho\left(y_i - x_i^\top \beta\right) \tag{10.9}$$

where $\rho\,(u) = u^2$. Huber (1964) suggests to generalize this method by considering other functions $\rho\,(u)$. In this approach (called M-estimation), the function $\rho\,(u)$ satisfies some

properties: $\rho(u) \geq 0$, $\rho(0) = 0$, $\rho(u) = \rho(-u)$ and $\rho(u_1) \geq \rho(u_2)$ if $|u_1| \geq |u_2|$. If we note $\psi(u) = \rho'(u)$, the first-order conditions of Problem (10.9) are:

$$\sum_{i=1}^{n} \psi\left(y_i - x_i^\top \beta\right) x_{i,k} = 0 \qquad \text{for all } k = 1, \ldots, K$$

We deduce that:

$$\sum_{i=1}^{n} \frac{\psi\left(y_i - x_i^\top \beta\right)}{y_i - x_i^\top \beta} \left(y_i - x_i^\top \beta\right) x_{i,k} = 0$$

By writing $w_i = \psi(u_i)/u_i$, we finally obtain:

$$\sum_{i=1}^{n} w_i \left(y_i - x_i^\top \beta\right) x_{i,k} = 0$$

We notice that the system of equations corresponds exactly to the first-order conditions of the WLS problem. The only difference is the endogeneity of the weights w_i that depend on the residuals u_i. To solve this system, we use the following iterative algorithm:

1. we choose an initial value $\beta^{(0)}$;
2. we calculate the diagonal matrix $W^{(j-1)}$ with $w_i = \psi(u_i)/u_i$ where $u_i = y_i - x_i^\top \beta^{(j-1)}$;
3. at the j^{th} iteration, we calculate the WLS estimator:

$$\beta^{(j)} = \left(\mathbf{X}^\top W^{(j-1)} \mathbf{X}\right)^{-1} \mathbf{X}^\top W^{(j-1)} \mathbf{Y}$$

4. we repeat steps 2 and 3 until the convergence of the algorithm: $\left|\beta^{(j)} - \beta^{(j-1)}\right| \leq \varepsilon$;
5. the M-estimator $\hat{\beta}$ is equal to $\beta^{(\infty)}$.

The most well-known M-estimator is obtained by setting $\rho(u) = |u|$ and $\psi(u) = \text{sign}(u)$ and is called LAD (least absolute deviation). A variant is proposed by Huber (1964):

$$\rho(u) = \begin{cases} u^2 & \text{if } |u| \leq c \\ c|u| & \text{if } |u| > c \end{cases}$$

These two estimators are less sensitive to outliers than the OLS estimator. This is why we call them robust estimators.

Quantile regression Let us consider a random variable Y with probability distribution $\mathbf{F}$. The quantile of order α of Y is defined by:

$$Q(\alpha) = \inf\{y \mid \mathbf{F}(y) \geq \alpha\}$$

The estimator $\hat{q}_\alpha$ of $Q(\alpha)$ is given by:

$$\hat{q}_\alpha = \arg\min_{q \in \mathbb{R}} \sum_{y_i \geq q} \alpha |y_i - q| + \sum_{y_i < q} (1 - \alpha) |y_i - q|$$

or:

$$\hat{q}_\alpha = \arg\min_{q \in \mathbb{R}} \sum_{i=1}^{n} \chi_\alpha(y_i - q)$$

where[3] $\chi_\alpha(u) = u \cdot (\alpha - \mathbb{1}\{u < 0\})$. If we consider the Gaussian linear model $Y = X\beta + U$ and apply the previous approach to the random variable U, the estimator $\hat{\beta}_\alpha$ of the quantile regression of order α is:

$$\hat{\beta}_\alpha = \arg \min_{\beta \in \mathbb{R}^K} \sum_{i=1}^n \chi_\alpha\left(y_i - x_i^\top \beta\right)$$

In the case $\alpha = 50\%$, we obtain the median regression:

$$\hat{\beta}_{50\%} = \arg \min_{\beta \in \mathbb{R}^K} \sum_{i=1}^n \left|y_i - x_i^\top \beta\right|$$

It consists in minimizing the sum of absolute values of residuals.

Since we have:

$$\begin{aligned} y_i &= x_i^\top \beta_\alpha + u_i \\ &= x_i^\top \left(\beta_\alpha^+ - \beta_\alpha^-\right) + u_i^+ - u_i^- \end{aligned}$$

we obtain the linear programming problem:

$$\begin{aligned} z^\star &= \arg\min c^\top z \\ &\text{s.t.} \begin{cases} Az = b \\ z \geq 0 \end{cases} \end{aligned}$$

where $\mathbf{Y}$ and $\mathbf{X}$ are the vector of y_i's and the matrix of $x_{i,k}$'s, $z = \begin{pmatrix} \beta_\alpha^+ & \beta_\alpha^- & \mathbf{U}^+ & \mathbf{U}^- \end{pmatrix}^\top$, $A = \begin{pmatrix} \mathbf{X} & -\mathbf{X} & I_n & -I_n \end{pmatrix}$, $b = \mathbf{Y}$ and $c = \begin{pmatrix} \mathbf{0}_n & \mathbf{0}_n & \alpha \mathbf{1}_n & (1-\alpha)\mathbf{1}_n \end{pmatrix}^\top$. The standard approach to find the estimator $\hat{\beta}_\alpha = \hat{\beta}_\alpha^+ - \hat{\beta}_\alpha^-$ is to solve this LP program using interior points methods (Koenker, 2005).

An alternative method is to use the robust regression with:

$$\rho(u) = \chi_\alpha(u) = u \cdot (\alpha - \mathbb{1}\{u < 0\})$$

and:

$$\psi(u) = \alpha - \mathbb{1}\{u < 0\}$$

In the case $\alpha = 50\%$, we obtain $\rho(u) = u \cdot (0.5 - \mathbb{1}\{u < 0\}) = 0.5 \cdot |u|$. The estimator of the median regression is then the LAD estimator.

10.1.2 Maximum likelihood estimation

10.1.2.1 Definition of the estimator

We consider a sample $\mathbf{Y} = (y_1, \ldots, y_n)$ of n observations. We assume that the probability of the sample may be written as a parametric function:

$$\begin{aligned} \Pr\{Y_1 = y_1, \ldots, Y_n = y_n\} &= L(\mathbf{Y} \mid \theta) \\ &= L(\theta \mid \mathbf{Y}) \end{aligned}$$

[3] Because we have:

$$\begin{aligned} \chi_\alpha(u) &= u \cdot (\alpha - \mathbb{1}\{u < 0\}) \\ &= \begin{cases} (1-\alpha) \cdot |u| & \text{if } u < 0 \\ \alpha \cdot |u| & \text{if } u \geq 0 \end{cases} \\ &= \alpha \cdot |u| \cdot \mathbb{1}\{u \geq 0\} + (1-\alpha) \cdot |u| \cdot \mathbb{1}\{u < 0\} \end{aligned}$$

where θ is a $K \times 1$ vector of parameters to estimate. The function L is called the likelihood function whereas the maximum likelihood estimator (MLE) is defined as follows:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta \mid \mathbf{Y}) \tag{10.10}$$

where Θ is the parameter space[4]. The principle of maximum likelihood is to find the value of θ that maximizes the probability of the sample data $\mathbf{Y}$. This is an inverse probability problem because we do not want to calculate the probability $L(\mathbf{Y} \mid \theta)$ of the sample given a model and a parameter vector θ, but we want to estimate the implicit parameter θ given the sample and the model.

We have:

$$\begin{aligned} \Pr\{Y_1 = y_1, \ldots, Y_n = y_n\} &= \Pr\{Y_1 = y_1\} \cdot \\ &\quad \Pr\{Y_2 = y_2 \mid Y_1 = y_1\} \cdot \\ &\quad \cdots \qquad\qquad . \\ &\quad \Pr\{Y_n = y_n \mid Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}\} \end{aligned}$$

Assuming that the observations are independent simplifies the computation of the likelihood function:

$$\begin{aligned} L(\theta \mid \mathbf{Y}) &= \prod_{i=1}^{n} \Pr\{Y_i = y_i\} \\ &\equiv \prod_{i=1}^{n} L_i(\theta \mid y_i) \end{aligned}$$

where $L_i(\theta \mid y_i) = \Pr\{Y_i = y_i\}$. L_i is called the likelihood of the observation i and corresponds to its density. Generally, the optimization problem (10.10) is replaced by the following which is more tractable:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \boldsymbol{\ell}(\theta \mid \mathbf{Y}) \tag{10.11}$$

where $\boldsymbol{\ell}_i(\theta \mid Y_i) \equiv \ln L_i(\theta \mid Y_i)$ is the log-likelihood function for the observation i and $\boldsymbol{\ell}(\theta \mid \mathbf{Y}) = \sum_{i=1}^{n} \boldsymbol{\ell}_i(\theta \mid Y_i)$. The gradient of the log-likelihood function is called the score function:

$$\mathcal{S}(\theta) = \frac{\partial \boldsymbol{\ell}(\theta \mid \mathbf{Y})}{\partial \theta}$$

At the optimum, we have $\mathcal{S}\left(\hat{\theta}\right) = \mathbf{0}$.

Example 103 (Bernoulli distribution) *We consider the sample* $\mathbf{Y} = \{y_1, \ldots, y_n\}$ *where* y_i *takes the value* 1 *with probability* p *or* 0 *with probability* $1 - p$. *We note* n_0 *and* n_1 *the number of observations, whose values are respectively equal to* 0 *and* 1. *We have* $n_0 + n_1 = n$.

We have:

$$\Pr\{Y_i = y_i\} = (1-p)^{1-y_i} \cdot p^{y_i}$$

[4] It is generally equal to $\mathbb{R}^K$.

It follows that the log-likelihood function is:

$$\begin{aligned}\boldsymbol{\ell}(p) &= \sum_{i=1}^{n} \ln \Pr\{Y_i = y_i\} \\ &= \sum_{i=1}^{n} (1 - y_i) \cdot \ln(1 - p) + y_i \cdot \ln p \\ &= n_0 \cdot \ln(1 - p) + n_1 \cdot \ln p\end{aligned}$$

The first-order condition is:

$$\frac{\partial \boldsymbol{\ell}(p)}{\partial p} = 0 \Leftrightarrow \frac{n_1}{p} - \frac{n_0}{1 - p} = 0$$

We deduce the expression of the MLE:

$$\hat{p} = \frac{n_1}{n_0 + n_1} = \frac{n_1}{n}$$

In Figure 10.2, we have represented the log-likelihood function $\boldsymbol{\ell}(p)$ with respect to the parameter p when $n_0 = 46$ and $n_1 = 74$. We verify that the maximum is reached when p takes the value $74/120 \approx 0.6167$.

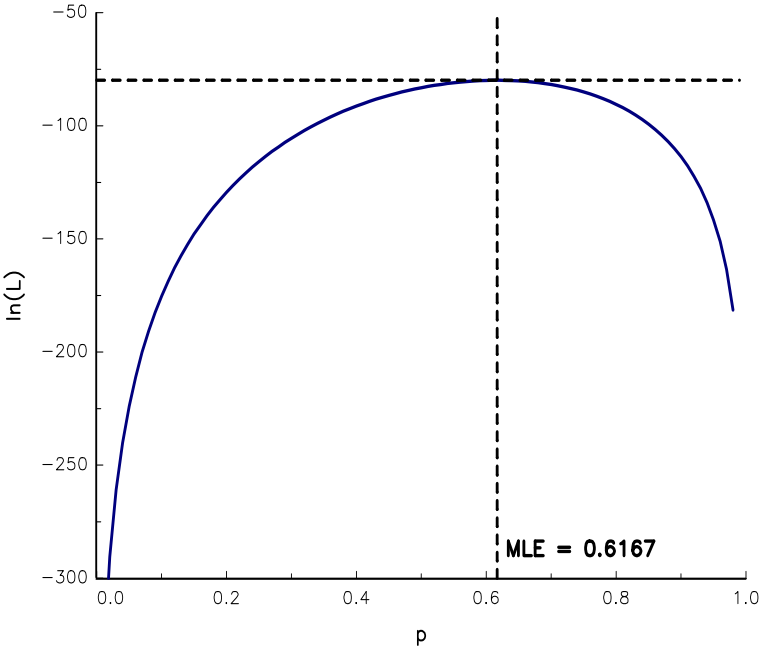


FIGURE 10.2: Log-likelihood function of the Bernoulli distribution

10.1.2.2 Asymptotic distribution

Let $\hat{\theta}_n$ denote the maximum likelihood estimator obtained with a sample of n observations. We can show that $\hat{\theta}_n$ is asymptotically normally distributed, unbiased and efficient:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \to \mathcal{N}\left(\mathbf{0}, \mathcal{J}^{-1}(\theta_0)\right)$$

where θ_0 is the true value and $\mathcal{J}(\theta_0)$ is the Fisher information matrix for an observation:

$$\mathcal{J}(\theta_0) = -\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}_i(\theta_0)}{\partial\theta\,\partial\theta^\top}\right]$$

Another useful result is the information matrix property:

$$\begin{aligned}\mathcal{I}(\theta_0) &= n\mathcal{J}(\theta_0)\\ &= -\mathbb{E}\left[\frac{\partial^2 \boldsymbol{\ell}(\theta_0)}{\partial\theta\,\partial\theta^\top}\right]\\ &= \mathbb{E}\left[\frac{\partial \boldsymbol{\ell}(\theta_0)}{\partial\theta}\frac{\partial \boldsymbol{\ell}(\theta_0)}{\partial\theta^\top}\right]\end{aligned}$$

This identity comes from the fact that we have:

$$\mathbb{E}[\mathcal{S}(\theta_0)] = \mathbb{E}\left[\frac{\partial \boldsymbol{\ell}(\theta_0)}{\partial\theta}\right] = \mathbf{0}$$

and:

$$\begin{aligned}\text{var}(\mathcal{S}(\theta_0)) &= \mathbb{E}\left[(\mathcal{S}(\theta_0) - \mathbb{E}[\mathcal{S}(\theta_0)])(\mathcal{S}(\theta_0) - \mathbb{E}[\mathcal{S}(\theta_0)])^\top\right]\\ &= \mathbb{E}\left[\frac{\partial \boldsymbol{\ell}(\theta_0)}{\partial\theta}\frac{\partial \boldsymbol{\ell}(\theta_0)}{\partial\theta}^\top\right]\\ &= \mathcal{I}(\theta_0)\end{aligned}$$

Remark 121 *Let $\tilde{\theta}$ be an unbiased estimator of θ_0. The Cramer-Rao theorem states that the variance of $\tilde{\theta}$ is bounded below by the inverse of the information matrix:*

$$\text{var}\left(\tilde{\theta}\right) \geq \mathcal{I}(\theta_0)^{-1}$$

It follows that the ML estimator is BUE (best unbiased estimator).

Let $h(\theta)$ be a function of the parameter vector θ. The invariance property states that $h\left(\hat{\theta}_n\right)$ converges almost surely to $h(\theta_0)$ and we have:

$$\sqrt{n}\left(h\left(\hat{\theta}_n\right) - h(\theta_0)\right) \to \mathcal{N}\left(\mathbf{0}, \frac{\partial h(\theta_0)}{\partial\theta^\top}\mathcal{J}^{-1}(\theta_0)\frac{\partial h(\theta_0)}{\partial\theta}\right)$$

Let us consider again the Bernoulli distribution (Example 103 on page 615). We recall that:

$$\frac{\partial \boldsymbol{\ell}(p)}{\partial p} = \frac{\sum_{i=1}^n y_i}{p} - \frac{\sum_{i=1}^n (1-y_i)}{1-p}$$

It follows that:

$$\frac{\partial^2 \boldsymbol{\ell}(p)}{\partial p^2} = -\frac{\sum_{i=1}^n y_i}{p^2} - \frac{\sum_{i=1}^n (1-y_i)}{(1-p)^2}$$

We have:

$$\begin{aligned}\mathcal{I}(p) &= -\mathbb{E}\left[-\frac{\sum_{i=1}^n Y_i}{p^2} - \frac{\sum_{i=1}^n (1-Y_i)}{(1-p)^2}\right]\\ &= \frac{\sum_{i=1}^n p}{p^2} + \frac{\sum_{i=1}^n (1-p)}{(1-p)^2}\\ &= \frac{n}{p(1-p)}\end{aligned}$$

and:

$$\mathcal{I}(p)^{-1} = \frac{p(1-p)}{n}$$

10.1.2.3 Statistical inference

Estimating the covariance matrix In practice, the covariance matrix of $\hat{\theta}$ is calculated as $\mathcal{I}\left(\hat{\theta}\right)^{-1} = \left(n\mathcal{J}\left(\hat{\theta}\right)\right)^{-1}$. Most of the time, the information matrix is however difficult to calculate analytically, and we prefer to estimate the covariance matrix as the inverse of the opposite of the Hessian matrix $\text{var}\left(\hat{\theta}\right) = \left(-H\left(\hat{\theta}\right)\right)^{-1}$ where:

$$H\left(\hat{\theta}\right) = \frac{\partial^2 \boldsymbol{\ell}\left(\hat{\theta}\right)}{\partial\,\theta\,\partial\,\theta^\top}$$

Two other estimators are very popular (Davidson and MacKinnon, 2004). The first one is the outer product of gradients (OPG) estimator:

$$\text{var}\left(\hat{\theta}\right) = \left(J\left(\hat{\theta}\right)^\top J\left(\hat{\theta}\right)\right)^{-1}$$

where $J\left(\hat{\theta}\right)$ is the Jacobian matrix of the log-likelihood function:

$$J\left(\hat{\theta}\right) = \left(\frac{\partial\,\boldsymbol{\ell}_i\left(\theta \mid y_i\right)}{\partial\,\theta}\right)$$

The second estimator is the sandwich estimator:

$$\text{var}\left(\hat{\theta}\right) = H\left(\hat{\theta}\right)^{-1} J\left(\hat{\theta}\right)^\top J\left(\hat{\theta}\right) H\left(\hat{\theta}\right)^{-1}$$

If the model is well-specified, the three estimators of the covariance matrix are equivalent. If it is not the case, it is better to use the sandwich estimator which is more robust to model misspecification.

From the covariance matrix $\text{var}\left(\hat{\theta}\right)$, we estimate the standard error of $\hat{\theta}_k$ by calculating the square root of the k^{th} diagonal element:

$$\sigma\left(\hat{\theta}_k\right) = \sqrt{\left(\text{var}\left(\hat{\theta}\right)\right)_{k,k}}$$

We can then test the hypothesis $\mathcal{H}_0 : \hat{\theta}_k = \xi_k$ by computing the t-statistic:

$$t = \frac{\hat{\theta}_k - \xi_k}{\sigma\left(\hat{\theta}_k\right)}$$

Asymptotically, we have $t \sim \mathcal{N}(0,1)$. In practice, we assume that $t \sim \mathbf{t}_{n-K}$ in the case of small samples.

Example 104 (Modeling LGD with the beta distribution) *We consider the following sample* $\mathbf{Y} = (y_1, \ldots, y_n)$ *of loss given default:*

$$\{68\%, 90\%, 22\%, 45\%, 17\%, 25\%, 89\%, 65\%, 75\%, 56\%, 87\%, 92\%, 46\%\}$$

We assume that the LGD parameter follows a beta distribution $\mathcal{B}(\alpha, \beta)$.

We have:

$$\begin{aligned}\boldsymbol{\ell}(\alpha,\beta) &= (\alpha-1)\sum_{i=1}^{n}\ln y_i+(b-1)\sum_{i=1}^{n}\ln(1-y_i)-\\ &\quad n\ln\mathfrak{B}(\alpha,\beta)\end{aligned}$$

The first-order conditions are:

$$\frac{\partial\boldsymbol{\ell}(\alpha,\beta)}{\partial\alpha}=-n\frac{\partial_a\mathfrak{B}(\alpha,\beta)}{\mathfrak{B}(\alpha,\beta)}+\sum_{i=1}^{n}\ln y_i=0$$

and:

$$\frac{\partial\boldsymbol{\ell}(\alpha,\beta)}{\partial\beta}=-n\frac{\partial_\beta\mathfrak{B}(\alpha,\beta)}{\mathfrak{B}(\alpha,\beta)}+\sum_{i=1}^{n}\ln(1-y_i)=0$$

Therefore, it is not possible to find the analytical expression of $\hat{\alpha}$ and $\hat{\beta}$. However, we can use numerical optimization for optimizing the log-likelihood function and we obtain $\hat{\alpha}=1.836$, $\hat{\beta}=1.248$ and $\boldsymbol{\ell}\left(\hat{\alpha},\hat{\beta}\right)=1.269$ The computation of the Hessian matrix gives:

$$\begin{aligned}H\left(\hat{\theta}\right) &= \begin{pmatrix}\partial^2_{\alpha}\boldsymbol{\ell}(\alpha,\beta) & \partial^2_{\alpha,\beta}\boldsymbol{\ell}(\alpha,\beta)\\ \partial^2_{\alpha,\beta}\boldsymbol{\ell}(\alpha,\beta) & \partial^2_{\beta}\boldsymbol{\ell}(\alpha,\beta)\end{pmatrix}\\ &= \begin{pmatrix}-4.3719 & 4.9723\\ 4.9723 & -10.6314\end{pmatrix}\end{aligned}$$

We deduce the covariance matrix:

$$\operatorname{var}\left(\hat{\theta}\right)=-H\left(\hat{\theta}\right)^{-1}=\begin{pmatrix}0.4887 & 0.2286\\ 0.2286 & 0.2010\end{pmatrix}$$

Finally, we obtain the results reported in Table 10.4.

TABLE 10.4: Results of the maximum likelihood estimation

Parameter	Estimate	Standard error	t-statistic	p-value
α	1.8356	0.6990	2.6258	0.0236
β	1.2478	0.4483	2.7834	0.0178

Hypothesis testing We now consider the general hypothesis $\mathcal{H}_0: C(\theta)=c$ where $C(\theta)$ is a non-linear function from $\mathbb{R}^K$ to $\mathbb{R}^g$, c is a vector of dimension g and g is the number of restrictions. We note $\hat{\theta}$ the unconstrained estimator and $\hat{\theta}_c$ the constrained estimator:

$$\begin{aligned}\hat{\theta}_c &= \arg\max\boldsymbol{\ell}(\theta)\\ &\text{s.t.}\quad C(\theta)=c\end{aligned}$$

$\mathcal{H}_0$ can be tested using Wald, likelihood ratio (LR) and Lagrange multiplier (LM) tests. The Wald statistic is defined as:

$$W=\left(C\left(\hat{\theta}\right)-c\right)^{\top}\left(\frac{\partial C\left(\hat{\theta}\right)}{\partial\theta^{\top}}\operatorname{var}\left(\tilde{\theta}\right)\frac{\partial C\left(\hat{\theta}\right)^{\top}}{\partial\theta}\right)^{-1}\left(C\left(\hat{\theta}\right)-c\right)$$

Under $\mathcal{H}_0$, the Wald test is:

$$W \sim \chi^2(g)$$

The second approach is based on the likelihood ratio:

$$\Lambda = \frac{L\left(\hat{\theta}_c \mid \mathbf{Y}\right)}{L\left(\hat{\theta} \mid \mathbf{Y}\right)}$$

Under $\mathcal{H}_0$, the LR test is:

$$-2\ln\Lambda = -2\left(\boldsymbol{\ell}\left(\hat{\theta}_c \mid \mathbf{Y}\right) - \boldsymbol{\ell}\left(\hat{\theta} \mid \mathbf{Y}\right)\right) \sim \chi^2(g)$$

The third approach uses the Lagrange multiplier statistic:

$$\text{LM} = \frac{\partial\,\boldsymbol{\ell}\left(\hat{\theta}_c\right)}{\partial\,\theta^\top}\mathcal{I}\left(\hat{\theta}_c\right)^{-1}\frac{\partial\,\boldsymbol{\ell}\left(\hat{\theta}_c\right)}{\partial\,\theta}$$

Under $\mathcal{H}_0$, the distribution of the LM statistic is the chi-squared distribution $\chi^2(g)$. We notice that the Wald test uses the unconstrained estimator $\hat{\theta}$ whereas the LM test uses the restricted estimator $\hat{\theta}_c$.

10.1.2.4 Some examples

Multivariate normal distribution We assume that $Y_i \sim \mathcal{N}_p(\mu, \Sigma)$. We have:

$$\boldsymbol{\ell}(\mu, \Sigma) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \mu)^\top \Sigma^{-1}(Y_i - \mu)$$

The first-order condition with respect to μ is:

$$\partial_\mu\,\boldsymbol{\ell}(\mu, \Sigma) = \sum_{i=1}^{n}\Sigma^{-1}(Y_i - \mu) = \mathbf{0}$$

Since $\sum_{i=1}^{n}\Sigma^{-1}(Y_i - \mu) = \Sigma^{-1}\sum_{i=1}^{n}(Y_i - \mu)$, we deduce that $\hat{\mu}$ is the empirical mean:

$$\hat{\mu} = \bar{Y} = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} Y_{i,1} \\ \vdots \\ n^{-1}\sum_{i=1}^{n} Y_{i,p} \end{pmatrix}$$

By using the properties of the trace function, the concentrated log-likelihood function becomes:

$$\begin{aligned}
\boldsymbol{\ell}(\hat{\mu}, \Sigma) &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^\top \Sigma^{-1}(Y_i - \hat{\mu}) \\
&= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\operatorname{tr}\left(\left(Y_i - \bar{Y}\right)^\top \Sigma^{-1}(Y_i - \hat{\mu})\right) \\
&= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\operatorname{tr}\left(\Sigma^{-1}\left(Y_i - \bar{Y}\right)\left(Y_i - \bar{Y}\right)^\top\right) \\
&= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}S\right)
\end{aligned}$$

where S is the $p \times p$ matrix defined in the following way:

$$S = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right) \left(Y_i - \bar{Y}\right)^\top$$

We deduce the first-order condition:

$$\frac{\partial \boldsymbol{\ell}\left(\hat{\mu}, \Sigma\right)}{\partial \Sigma^{-1}} = \frac{n}{2}\Sigma - \frac{1}{2}S = 0$$

It follows that the ML estimator of Σ is the empirical covariance matrix:

$$\hat{\Sigma} = \frac{1}{n}S = \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right) \left(Y_i - \bar{Y}\right)^\top$$

Extension of the linear regression By assuming that $y_i = x_i^\top \beta + u_i$ with $u_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$, we obtain:

$$\begin{aligned}
\boldsymbol{\ell}\left(\beta, \sigma\right) &= \sum_{i=1}^{n} \ln \Pr\left\{Y_i = y_i\right\} \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i} \cdot \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i^\top \beta}{\sigma_i}\right)^2\right)\right) \\
&= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n} \ln \sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - x_i^\top \beta}{\sigma_i}\right)^2
\end{aligned}$$

In the homoscedastic case $\sigma_i = \sigma$, we can show that the estimators $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_{\text{OLS}}$ are the same[5]. Let us now assume that:

$$\sigma_i^2 = \sigma^2 + z_i^\top \gamma$$

We obtain:

$$\boldsymbol{\ell}\left(\beta, \sigma, \gamma\right) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n} \ln\left(\sigma^2 + z_i^\top \gamma\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - x_i^\top \beta\right)^2}{\sigma^2 + z_i^\top \gamma}$$

This is an example of linear models with heteroscedastic residuals.

When the model is non-linear $y_i = g\left(x_i, \beta\right) + u_i$ where $u_i \sim \mathcal{N}\left(0, \sigma^2\right)$, the log-likelihood function becomes:

$$\boldsymbol{\ell}\left(\beta, \sigma\right) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - g\left(x_i, \beta\right)\right)^2}{\sigma^2}$$

Some examples of non-linear models are given in Table 10.5.

10.1.2.5 EM algorithm

The expectation–maximization (EM) algorithm is an iterative method to find the maximum likelihood estimate when the statistical model depends on unobserved latent variables.

[5] See Exercise 10.3.4 on page 706.

TABLE 10.5: Non-linear models

Model	Function $g(x, \beta)$
Exponential (gowth)	$y_i = \beta_1 e^{\beta_2 x_i} + u_i$
Exponential (decay)	$y_i = y_- + (y_+ - y_-) e^{-\beta x_i} + u_i$
Hyperbola	$y_i = (\beta_1 x_i) / (\beta_2 + x_i)$
Sine	$y_i = \beta_1 + \beta_2 \sin(\beta_3 x_i + \beta_4) + u_i$
Boltzmann	$y_i = y_- + (y_+ - y_-) / \left(1 + e^{\frac{\beta_1 - x_i}{\beta_2}}\right)$
Damped sine	$y_i = \beta_1 e^{-\beta_2 x_i} \sin(\beta_3 x_i + \beta_4) + u_i$

We note $\mathbf{Y}$ the sample of observed data and $\mathbf{Z}$ the sample of unobservable data. We have:

$$\begin{aligned}
\boldsymbol{\ell}(\mathbf{Y}, \mathbf{Z}; \theta) &= \sum_{i=1}^{n} \ln f(y_i, z_i; \theta) \\
&= \sum_{i=1}^{n} \ln \left(f(y_i; \theta) f(z_i \mid y_i; \theta) \right) \\
&= \sum_{i=1}^{n} \ln f(y_i; \theta) + \sum_{i=1}^{n} \ln f(z_i \mid y_i; \theta) \\
&= \boldsymbol{\ell}(\mathbf{Y}; \theta) + \sum_{i=1}^{n} \ln f(z_i \mid y_i; \theta)
\end{aligned}$$

We deduce that:

$$\boldsymbol{\ell}(\mathbf{Y}; \theta) = \boldsymbol{\ell}(\mathbf{Y}, \mathbf{Z}; \theta) - \sum_{i=1}^{n} \ln f(z_i \mid y_i; \theta)$$

Dempster *et al.* (1977) define the expected value of the log likelihood function as follows:

$$Q\left(\theta; \theta^{(k)}\right) = \mathbb{E}\left[\boldsymbol{\ell}(\mathbf{Y}, \mathbf{Z}; \theta) \middle| \theta^{(k)}\right]$$

where $\theta^{(k)}$ is the vector of parameters at iteration k. They show that under some conditions the sequence of maxima $\theta^{(k+1)} = \arg\max_\theta Q\left(\theta; \theta^{(k)}\right)$ tends to a global maximum $\hat{\theta}_{\text{EM}} = \theta^{(\infty)}$ implying that $\boldsymbol{\ell}\left(\mathbf{Y}; \hat{\theta}_{\text{EM}}\right) \geq \boldsymbol{\ell}(\mathbf{Y}; \theta)$ for all $\theta \in \Theta$. The EM algorithm consists in iteratively applying the two steps:

(E-Step) we calculate the expected value of the log-likelihood function $\mathbb{E}\left[\boldsymbol{\ell}(\mathbf{Y}, \mathbf{Z}; \theta) \middle| \theta^{(k)}\right]$ with respect to the parameter vector $\theta^{(k)}$;

(M-Step) we estimate $\theta^{(k+1)}$ by maximizing $Q\left(\theta; \theta^{(k)}\right)$:

$$\theta^{(k+1)} = \arg\max_\theta Q\left(\theta; \theta^{(k)}\right)$$

The EM algorithm is used to solve many statistical problems: missing data, grouping, censoring and truncation models, finite mixtures, variance components, factor analysis, hidden Markov models, switching Markov processes, etc.

TABLE 10.6: Insulation life in hours at various test temperatures

Motorette	1	2	3	4	5
150°	8 064*	8 064*	8 064*	8 064*	8 064*
170°	1 764	2 772	3 444	3 542	3 780
190°	408	408	1 344	1 344	1 440
220°	408	408	504	504	504
Motorette	6	7	8	9	10
150°	8 064*	8 064*	8 064*	8 064*	8 064*
170°	4 860	5 196	5 448*	5 448*	5 448*
190°	1 680*	1 680*	1 680*	1 680*	1 680*
220°	528*	528*	528*	528*	528*

Source: Schmee and Hahn (1979).
An asterisk * indicates that the test has been stopped without the failure of the motorette, implying that the observation is censored.

Censored data Table 10.6 gives the results of temperature accelerated life tests on electrical insulation in 40 motorettes. Ten motorettes were tested at each of four temperatures (150°, 170°, 190° and 220°). The results are the following: all 10 motorettes at 150° are still on test without failure at 8 064 hours; 3 motorettes at 170° are still on test without failure at 5 448 hours; 5 motorettes at 190° are still on test without failure at 1 680 hours; 5 motorettes at 220° are still on test without failure at 528 hours. We assume the following model:

$$y_i = \beta_0 + \beta_1 x_i + \sigma \varepsilon_i$$

where $y_i = \log_{10} d_i$, d_i is the failure time, $x_i = 1\,000/(t_i + 273.2°)$, t_i is the temperature and $\varepsilon_i \sim \mathcal{N}(0,1)$. We cannot use the linear regression because some values of y_i are censured. Let $\mathcal{A}$ and $\mathcal{B}$ be the sets of non-censored and censored data. The expression of the log-likelihood function is then:

$$\begin{aligned} \ell(\theta) &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \\ &\quad \frac{1}{2\sigma^2}\left(\sum_{i\in\mathcal{A}}(y_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i\in\mathcal{B}}(Z_i - \beta_0 - \beta_1 x_i)^2\right) \end{aligned}$$

where Z_i is the failure time of the motorette i that we do not have observed. However, we know that $Z_i \geq c_i$ where c_i is the censured failure time. We deduce that:

$$\begin{aligned} \ell(\theta) &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i\in\mathcal{A}}(y_i - \beta_0 - \beta_1 x_i)^2 \\ &\quad -\frac{1}{2\sigma^2}\sum_{i\in\mathcal{B}}\mathbb{E}\left[(Z_i - \beta_0 - \beta_1 x_i)^2 \mid Z_i \geq c_i\right] \end{aligned} \tag{10.12}$$

where:

$$\begin{aligned} \mathbb{E}\left[(Z_i - \beta_0 - \beta_1 x_i)^2 \mid Z_i \geq c_i\right] &= \mathbb{E}\left[Z_i^2 \mid Z_i \geq c_i\right] - \\ &\quad 2(\beta_0 + \beta_1 x_i)\,\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right] + \\ &\quad (\beta_0 + \beta_1 x_i)^2 \end{aligned} \tag{10.13}$$

Tanner (1993) showed that:

$$\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right] = \mu_i + \sigma \frac{\phi\left(\sigma^{-1}\left(c_i - \mu_i\right)\right)}{1 - \Phi(\sigma^{-1}\left(c_i - \mu_i\right))}$$

and:

$$\mathbb{E}\left[Z_i^2 \mid Z_i \geq c_i\right] = \mu_i^2 + \sigma^2 + \sigma\left(c_i + \mu_i\right) \frac{\phi\left(\sigma^{-1}\left(c_i - \mu_i\right)\right)}{1 - \Phi(\sigma^{-1}\left(c_i - \mu_i\right))}$$

where $\mu_i = \beta_0 + \beta_1 x_i$. The EM algorithm is then:

(E-Step) we calculate $\mathbb{E}\left[Z_i^2 \mid Z_i \geq c_i\right]$ and $\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right]$ using the values $\beta_0^{(k)}$, $\beta_1^{(k)}$ and $\sigma^{(k)}$; we deduce the conditional expectation (10.13);

(M-Step) we estimate $\beta_0^{(k+1)}$, $\beta_1^{(k+1)}$ and $\sigma^{(k+1)}$ by maximizing the conditional log-likelihood function (10.12).

Starting from the initial values $\beta_0^{(0)} = \beta_1^{(0)} = \sigma^{(0)} = 1$, we obtain $\beta_0^{(1)} = -5.087$, $\beta_1^{(1)} = 4.008$ and $\sigma^{(1)} = 0.619$ at the first iteration, $\beta_0^{(2)} = -6.583$, $\beta_1^{(2)} = 4.670$ and $\sigma^{(2)} = 0.515$ at the second iteration, etc. Finally, the algorithm converges after 33 iterations and the EM estimates are $\hat{\beta}_0 = -6.019$, $\hat{\beta}_1 = 4.311$ and $\hat{\sigma} = 0.259$. In Table 10.7, we also report the value taken by the expected failure time $\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right]$ at the last iteration.

TABLE 10.7: Expected failure time $\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right]$ obtained with the EM algorithm

Motorette	1	2	3	4	5
150°	17 447*	17 447*	17 447*	17 447*	17 447*
170°	1 764	2 772	3 444	3 542	3 780
190°	408	408	1 344	1 344	1 440
220°	408	408	504	504	504
Motorette	6	7	8	9	10
150°	17 447*	17 447*	17 447*	17 447*	17 447*
170°	4 860	5 196	8 574*	8 574*	8 574*
190°	2 862*	2 862*	2 862*	2 862*	2 862*
220°	850*	850*	850*	850*	850*

The censored data represented by an asterisk * are replaced by the value of $\mathbb{E}\left[Z_i \mid Z_i \geq c_i\right]$ calculated by the EM algorithm at the last iteration.

Multivariate Gaussian mixture model The probability density function of the random vector Y of dimension K is defined as a weighted sum of Gaussian distributions:

$$f\left(y\right) = \sum_{j=1}^{m} \pi_j \phi_K\left(y; \mu_j, \Sigma_j\right)$$

where m is the number of mixture components, μ_j and Σ_j are the mean vector and the covariance matrix of the Gaussian distribution associated with the j^{th} component, and π_j is the mixture weight such that $\sum_{j=1}^{m} \pi_j = 1$. The log-likelihood function of the sample $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ is:

$$\ell\left(\theta\right) = \sum_{i=1}^{n} \ln \sum_{j=1}^{m} \pi_j \phi_K\left(Y_i; \mu_j, \Sigma_j\right)$$

The derivative of $\boldsymbol{\ell}(\theta)$ with respect to μ_j is equal to:

$$\frac{\partial \boldsymbol{\ell}(\theta)}{\partial \mu_j} = \sum_{i=1}^{n} \frac{\pi_j \phi_K(Y_i; \mu_j, \Sigma_j)}{\sum_{s=1}^{m} \pi_s \phi_K(Y_i; \mu_s, \Sigma_s)} \Sigma_j^{-1}(Y_i - \mu_j)$$

Therefore, the first-order condition is:

$$\sum_{i=1}^{n} \pi_{j,i} \Sigma_j^{-1}(Y_i - \mu_j) = \mathbf{0}$$

where:

$$\pi_{j,i} = \frac{\pi_j \phi_K(Y_i; \mu_j, \Sigma_j)}{\sum_{s=1}^{m} \pi_s \phi_K(Y_i; \mu_s, \Sigma_s)}$$

We deduce the expression of the estimator $\hat{\mu}_j$:

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n} \pi_{j,i} Y_i}{\sum_{i=1}^{n} \pi_{j,i}} \tag{10.14}$$

For the derivative with respect to Σ_j, we consider the function $g\left(\Sigma_j^{-1}\right)$ defined as follows:

$$\begin{aligned} g\left(\Sigma_j^{-1}\right) &= \frac{1}{(2\pi)^{K/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}(Y_i-\mu_j)^\top \Sigma_j^{-1}(Y_i-\mu_j)} \\ &= \frac{\left|\Sigma_j^{-1}\right|^{1/2}}{(2\pi)^{K/2}} e^{-\frac{1}{2}\operatorname{trace}\left(\Sigma_j^{-1}(Y_i-\mu_j)(Y_i-\mu_j)^\top\right)} \end{aligned}$$

We note:

$$\varphi\left(\Sigma_j^{-1}\right) = \exp\left(-\frac{1}{2}\operatorname{trace}\left(\Sigma_j^{-1}(Y_i - \mu_j)(Y_i - \mu_j)^\top\right)\right)$$

It follows that[6]:

$$\begin{aligned} \frac{\partial g\left(\Sigma_j^{-1}\right)}{\partial \Sigma_j^{-1}} &= \frac{1}{2} \frac{\left|\Sigma_j^{-1}\right|^{-1/2} \left|\Sigma_j^{-1}\right| \Sigma_j}{(2\pi)^{K/2}} \varphi\left(\Sigma_j^{-1}\right) \\ &\quad -\frac{1}{2}(Y_i - \mu_j)(Y_i - \mu_j)^\top \frac{\left|\Sigma_j^{-1}\right|^{1/2}}{(2\pi)^{K/2}} \varphi\left(\Sigma_j^{-1}\right) \\ &= \frac{1}{(2\pi)^{K/2} |\Sigma_j|^{1/2}} \varphi\left(\Sigma_j^{-1}\right) \frac{\left(\Sigma_j - (Y_i - \mu_j)(Y_i - \mu_j)^\top\right)}{2} \\ &= \frac{1}{2} g\left(\Sigma_j^{-1}\right) \left(\Sigma_j - (Y_i - \mu_j)(Y_i - \mu_j)^\top\right) \end{aligned}$$

We deduce that:

$$\frac{\partial \boldsymbol{\ell}(\theta)}{\partial \Sigma_j^{-1}} = \frac{1}{2} \sum_{i=1}^{n} \frac{\pi_j \phi_K(Y_i; \mu_j, \Sigma_j)}{\sum_{s=1}^{m} \pi_s \phi_K(Y_i; \mu_s, \Sigma_s)} \left(\Sigma_j - (Y_i - \mu_j)(Y_i - \mu_j)^\top\right)$$

[6] We use the following results:

$$\begin{aligned} \frac{\partial |A|}{\partial A} &= |A| \left(A^{-1}\right)^\top \\ \frac{\partial \operatorname{trace}\left(A^\top B\right)}{\partial A} &= B \end{aligned}$$

The first-order condition is then:

$$\sum_{i=1}^{n} \pi_{j,i} \left(\Sigma_j - (Y_i - \mu_j)(Y_i - \mu_j)^\top \right) = \mathbf{0}$$

It follows that the estimator $\hat{\Sigma}_j$ is equal to:

$$\hat{\Sigma}_j = \frac{\sum_{i=1}^{n} \pi_{j,i} (Y_i - \mu_j)(Y_i - \mu_j)^\top}{\sum_{i=1}^{n} \pi_{j,i}} \tag{10.15}$$

Regarding the mixture probabilities π_j, the first-order condition implies:

$$\sum_{i=1}^{n} \frac{\phi_K(Y_i; \mu_j, \Sigma_j)}{\sum_{s=1}^{m} \pi_s \phi_K(Y_i; \mu_s, \Sigma_s)} = \lambda$$

where λ is the Lagrange multiplier associated to the constraint $\sum_{j=1}^{m} \pi_j = 1$. We deduce that $\lambda = n$. We conclude that it is not possible to directly define the estimator $\hat{\pi}_j$. This is why we have to use another route to obtain the ML estimator.

We introduce the estimator $\hat{\pi}_{j,i}$:

$$\hat{\pi}_{j,i} = \frac{\pi_j \phi_K(Y_i; \mu_j, \Sigma_j)}{\sum_{s=1}^{m} \pi_s \phi_K(Y_i; \mu_s, \Sigma_s)} \tag{10.16}$$

$\hat{\pi}_{j,i}$ is the posterior probability of the regime index for the observation i. Knowing $\hat{\pi}_{j,i}$, the estimator $\hat{\pi}_j$ is given by:

$$\hat{\pi}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}_{j,i} \tag{10.17}$$

The EM algorithm consists in the following iterations:

1. we set $k = 0$ and initialize the algorithm with starting values $\pi_j^{(0)}$, $\mu_j^{(0)}$ and $\Sigma_j^{(0)}$;
2. using Equation (10.16), we calculate the posterior probabilities $\pi_{j,i}$:

$$\pi_{j,i}^{(k)} = \frac{\pi_j^{(k)} \phi_K\left(Y_i; \mu_j^{(k)}, \Sigma_j^{(k)}\right)}{\sum_{s=1}^{m} \pi_s^{(k)} \phi_K\left(Y_i; \mu_s^{(k)}, \Sigma_s^{(k)}\right)}$$

3. using Equations (10.14), (10.15) and (10.17), we update the estimators $\hat{\pi}_j$, $\hat{\mu}_j$ and $\hat{\Sigma}_j$:

$$\begin{aligned}
\pi_j^{(k+1)} &= \frac{\sum_{i=1}^{n} \pi_{j,i}^{(k)}}{n} \\
\mu_j^{(k+1)} &= \frac{\sum_{i=1}^{n} \pi_{j,i}^{(k)} Y_i}{\sum_{i=1}^{n} \pi_{j,i}^{(k)}} \\
\Sigma_j^{(k+1)} &= \frac{\sum_{i=1}^{n} \pi_{j,i}^{(k)} \left(Y_i - \mu_j^{(k+1)}\right)\left(Y_i - \mu_j^{(k+1)}\right)^\top}{\sum_{i=1}^{n} \pi_{j,i}^{(k)}}
\end{aligned}$$

4. we iterate steps 2 and 3 until convergence;
5. finally, we have $\hat{\pi}_j = \pi_j^{(\infty)}$, $\hat{\mu}_j = \pi_j^{(\infty)}$ and $\hat{\Sigma}_j = \Sigma_j^{(\infty)}$.

Let us consider the monthly returns of the S&P 500 index from January 2000 to December 2015. Using a Gaussian model, we obtain $\hat{\mu} = 0.49\%$ and $\hat{\sigma} = 4.90\%$. If we consider a bivariate mixture model, we obtain the following estimates:

Regime	$\hat{\pi}_j$	$\hat{\mu}_j$	$\hat{\sigma}_j$
$j = 1$	72.01%	1.40%	3.13%
$j = 2$	27.99%	−1.84%	7.28%

We have represented the corresponding probability density functions in Figure 10.3. We notice that the Gaussian and mixture pdfs are very different, even if they have the same mean and variance[7]. However, the skewness coefficients are very different. For the Gaussian distribution, γ_1 is equal to zero, whereas we have $\gamma_1 = -0.745$ for the mixture distribution[8].

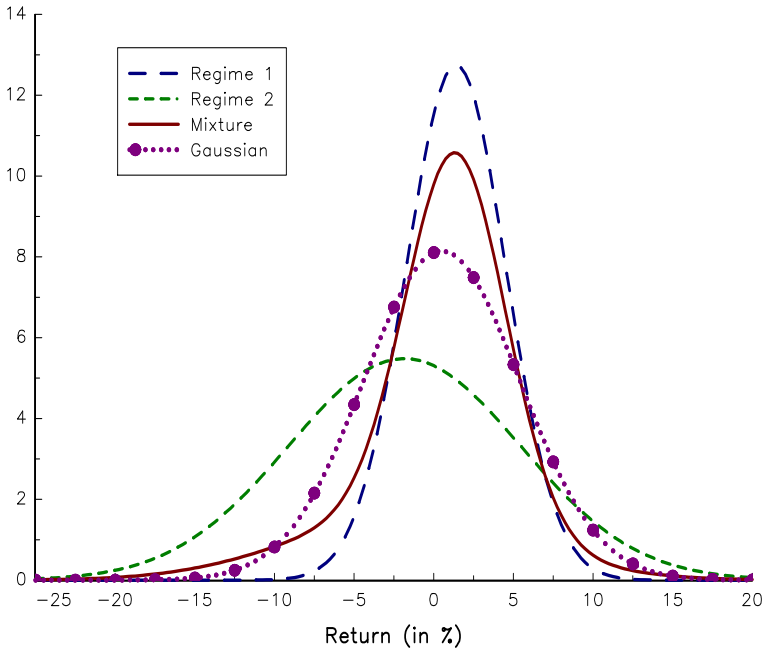


FIGURE 10.3: Probability density function of the monthly returns of the S&P 500 index

[7]The first two moments of the mixture distribution are:

$$\mathbb{E}[Y] = \pi_1\mu_1 + \pi_2\mu_2$$

and:

$$\text{var}(Y) = \pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1 - \mu_2)^2$$

[8]The expression of the skewness coefficient is:

$$\gamma_1(Y) = \frac{\pi_1\pi_2\left((\pi_2 - \pi_1)(\mu_1 - \mu_2)^3 + 3(\mu_1 - \mu_2)\left(\sigma_1^2 - \sigma_2^2\right)\right)}{\left(\pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1 - \mu_2)^2\right)^{3/2}}$$

10.1.3 Generalized method of moments

The method of moments is another approach for estimating a statistical model. While the objective of the method of maximum likelihood is to maximize the probability of the sample data, the method of moments estimates the parameters by fitting the empirical moments of the sample data. The choice of one approach rather than another generally depends on the computational facilities associated to the probability density function and the statistical moments. The method of moments is particularly suitable for some financial models that cannot be described by an analytical probability distribution.

10.1.3.1 Method of moments

Let Y be a random variable, whose probability distribution $\mathbf{F}(y;\theta)$ depends on some parameters θ. We assume that we can calculate the first m statistical moments:

$$m_j(\theta) = \mathbb{E}\left[Y^j\right] = \int y^j \,\mathrm{d}\mathbf{F}(y;\theta)$$

We consider a sample $\mathbf{Y} = \{y_1, \ldots, y_n\}$ and we note $g(\theta)$ the $m \times 1$ vector, whose elements are equal to:

$$\begin{aligned} g_j(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i^j - m_j(\theta)\right) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} y_i^j\right) - m_j(\theta) \end{aligned}$$

Let K be the dimension of θ. If K is exactly equal to m, the method of moments (MM) estimator is defined by:

$$g\left(\hat{\theta}\right) = \mathbf{0}$$

If $K < m$, the MM estimator minimizes the quadratic criterion:

$$\hat{\theta} = \arg\min Q(\theta)$$

where:

$$Q(\theta) = g(\theta)^\top W g(\theta)$$

and W is a $m \times m$ matrix.

Example 105 *We assume that* $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

We have:

$$m_1(\theta) = \mathbb{E}[Y] = \mu$$

and:

$$m_2(\theta) = \mathbb{E}\left[Y^2\right] = \mu^2 + \sigma^2$$

It follows that the MM estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ satisfies:

$$\begin{cases} n^{-1}\sum_{i=1}^{n} y_i - \hat{\mu} = 0 \\ n^{-1}\sum_{i=1}^{n} y_i^2 - \left(\hat{\mu}^2 + \hat{\sigma}^2\right) = 0 \end{cases}$$

We deduce that:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

and:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

The MM estimators $\hat{\mu}$ and $\hat{\sigma}$ correspond to the empirical mean and the empirical standard deviation.

Example 106 *We now assume that* $Y = X \cdot U$ *where* $X \sim \mathcal{N}(\mu, \sigma^2)$, $U \sim \mathcal{U}_{[0,1]}$ *and* $X \perp U$. *We want to estimate the parameters* μ *and* σ *for the sample* $\mathbf{Y} = \{-0.320, -0.262, -0.284, -0.296, 0.636, 0.547, 0.024, 0.483, -1.045, -0.030\}$.

We have:

$$\mathbb{E}[Y] = \mathbb{E}[X \cdot U] = \mathbb{E}[X] \cdot \mathbb{E}[U] = \frac{\mu}{2}$$

and:

$$\mathbb{E}\left[Y^2\right] = \mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[U^2\right] = \left(\mu^2 + \sigma^2\right)\left(\frac{1}{12} + \frac{1}{2^2}\right) = \frac{1}{3}\left(\sigma^2 + \mu^2\right)$$

We deduce that the MM estimators $\hat{\mu}$ and $\hat{\sigma}$ are:

$$\hat{\mu} = \frac{2}{n}\sum_{i=1}^{n} y_i = 2\bar{y}$$

and:

$$\hat{\sigma}^2 = \frac{3}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}^2 = \frac{3}{n}\sum_{i=1}^{n} y_i^2 - 4\bar{y}^2$$

Using the sample $\mathbf{Y}$, we obtain $\hat{\mu} = -0.109$ and $\hat{\sigma} = 0.836$.

Example 107 (Modeling LGD with the beta distribution) *We consider Example 104 on page 618, but we now want to estimate the parameters of the beta distribution by the method of moments.*

If $Y \sim \mathcal{B}(\alpha, \beta)$, we have:

$$\mathbb{E}[Y] = \frac{\alpha}{\alpha + \beta}$$

and:

$$\text{var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Let $\hat{\mu}_{\text{LGD}}$ and $\hat{\sigma}_{\text{LGD}}$ be the empirical mean and standard deviation of the LGD sample. We deduce that the MM estimators are:

$$\hat{\alpha} = \frac{\hat{\mu}_{\text{LGD}}^2(1 - \hat{\mu}_{\text{LGD}})}{\hat{\sigma}_{\text{LGD}}^2} - \hat{\mu}_{\text{LGD}}$$

and:

$$\hat{\beta} = \frac{\hat{\mu}_{\text{LGD}}(1 - \hat{\mu}_{\text{LGD}})^2}{\hat{\sigma}_{\text{LGD}}^2} - (1 - \hat{\mu}_{\text{LGD}})$$

Using our sample, we obtain $\hat{\mu}_{\text{LGD}} = 59.77\%$ and $\hat{\sigma}_{\text{LGD}} = 27.02\%$. Therefore, the MM estimates are $\hat{\alpha}_{\text{MM}} = 1.371$ and $\hat{\beta}_{\text{MM}} = 0.923$. We recall that the ML estimates were $\hat{\alpha}_{\text{ML}} = 1.836$ and $\hat{\beta}_{\text{ML}} = 1.248$. If we compare the two calibrated probability distributions, we observe that their shape is very different (see Figure 10.4).

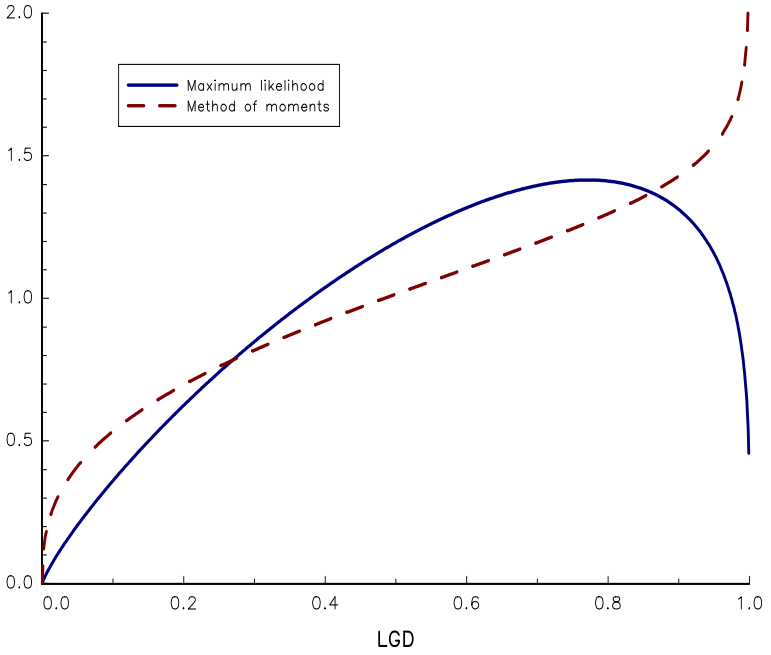


FIGURE 10.4: Calibrated density function of the loss given default

Let us assume that $Y \sim \mathcal{E}(\lambda)$. We have $m_1(\lambda) = \lambda^{-1}$ and $m_2(\lambda) = 2\lambda^{-2}$. We deduce that:

$$\begin{aligned} Q(\lambda) &= w_1 \left(\frac{1}{n}\sum_{i=1}^n y_i - \frac{1}{\lambda} \right)^2 + \\ &\quad 2w_2 \left(\frac{1}{n}\sum_{i=1}^n y_i - \frac{1}{\lambda} \right) \left(\frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{2}{\lambda^2} \right) + \\ &\quad w_3 \left(\frac{1}{n}\sum_{i=1}^n y_i^2 - \frac{2}{\lambda^2} \right)^2 \end{aligned}$$

where:

$$W = \begin{pmatrix} w_1 & w_2 \\ w_2 & w_3 \end{pmatrix}$$

If $w_2 = w_3 = 0$, the MM estimator is:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n y_i} = \frac{1}{\bar{y}}$$

If $w_1 = w_2 = 0$, the MM estimator becomes:

$$\hat{\lambda} = \sqrt{\frac{n}{2\sum_{i=1}^n y_i^2}}$$

In the other cases, we have to use a numerical optimization algorithm to find $\hat{\lambda}$. We consider the following sample: 0.08, 0.14, 0.00, 0.06, 0.11, 0.22, 0.11, 0.09, 0.02 and 0.26. Using the

weighting scheme $w_1 = w^2$, $w_2 = w(1-w)$ and $w_3 = (1-w)^2$ where $w \in [0,1]$, we obtain the following MM estimates:

w	0.00	0.20	0.50	0.70	1.00
$\hat{\lambda}$	10.59	9.99	9.54	9.35	9.17

10.1.3.2 Extension to the GMM approach

The classical method of moments assumes that the number m of moments is equal to the number K of parameters and the weight matrix W is the identity matrix. Hansen (1982) extends this approach in two ways. First, he assumes that $g(\theta)$ does not necessarily correspond to the first m statistical moments, but can also include orthogonal conditions. Second, the matrix W is chosen in order to obtain an estimate $\hat{\theta}$ with the smallest variance. Therefore, the generalized method of moments (GMM) is a direct extension of the method of moments.

Statistical inference Like the MM approach, the GMM approach considers m empirical centered moments that depend on the parameter vector θ. We note $h_{i,j}(\theta)$ the j^{th} moment condition for the observation i such that $\mathbb{E}[h_{i,j}(\theta_0)] = 0$ where θ_0 is the true parameter vector. We note $g(\theta)$ the vector whose elements are:

$$g_j(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_{i,j}(\theta)$$

The GMM estimator is defined as:

$$\hat{\theta} = \arg\min g(\theta)^\top W g(\theta)$$

where W is the weighting symmetric matrix. Hansen (1982) shows that $\hat{\theta}$ is asymptotically normally distributed:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}(\mathbf{0}, V)$$

where:

$$V = \left(D^\top W D\right)^{-1} D^\top W S W D \left(D^\top W D\right)^{-1}$$

D is the Jacobian matrix of $g(\theta)$, and S is the covariance matrix of empirical moments:

$$\begin{aligned} S &= \lim_{n\to\infty} n \cdot \mathbb{E}\left[g(\theta_0)\, g(\theta_0)^\top\right] \\ &= \lim_{n\to\infty} \mathbb{E}\left[h(\theta_0)\, h(\theta_0)^\top\right] \end{aligned}$$

We can also show that the optimal weighting matrix corresponds to the case $W = S^{-1}$. The underlying idea is that moments with small variance are more informative than moments with high variance. Therefore, they should have a larger weight. We then deduce that $V = \left(D^\top S^{-1} D\right)^{-1}$ and $\text{var}\left(\hat{\theta}\right) = \left(nD^\top S^{-1} D\right)^{-1}$.

We notice that the quadratic form $Q(\theta) = g(\theta)^\top W g(\theta)$ is particular since the weighting matrix $W = S^{-1}$ depends on the parameter vector θ. A direct optimization of $Q(\theta)$ does not generally converge. This is why we can use the following iterative algorithm:

1. let $W^{(0)}$ be the initial weighting matrix;
2. at iteration k, we find the optimal value $\hat{\theta}^{(k)}$:

$$\hat{\theta}^{(k)} = \arg\min g(\theta)^\top W^{(k-1)} g(\theta)$$

3. we update the weighting matrix $W^{(k)} = {S^{(k)}}^{-1}$ where:

$$S^{(k)} = \mathbb{E}\left[h\left(\hat{\theta}^{(k)}\right) h\left(\hat{\theta}^{(k)}\right)^\top\right]$$

4. we repeat steps 2 and 3 until convergence: $\left|\hat{\theta}^{(k)} - \hat{\theta}^{(k-1)}\right| \leq \varepsilon$.

Remark 122 *The two-step GMM procedure of Hansen (1982) consists in setting $W^{(0)}$ to the identity matrix and to stop at the second iteration. In this case, we obtain $\hat{\theta}_{\text{GMM}} = \hat{\theta}^{(2)}$.*

We consider the previous example. We have $h_{i,1}(\lambda) = y_i - \lambda^{-1}$ and $h_{i,2}(\lambda) = y_i^2 - 2\lambda^{-2}$. At the first iteration, we set $W^{(0)} = I_2$ and we obtain $\hat{\lambda}^{(1)} = 9.36$. We deduce that:

$$S^{(1)} = \frac{1}{n} h\left(\hat{\lambda}^{(1)}\right) h\left(\hat{\lambda}^{(1)}\right)^\top = 10^{-4} \cdot \begin{pmatrix} 59.54 & 15.55 \\ 15.55 & 4.78 \end{pmatrix}$$

At the second iteration, the weighting matrix becomes $W^{(1)} = {S^{(1)}}^{-1}$, implying that the solution is $\hat{\lambda}^{(2)} = 11.49$. Finally, we obtain $\hat{\lambda}^{(3)} = 12.04$, $\hat{\lambda}^{(4)} = 12.13$, $\hat{\lambda}^{(5)} = 12.14$ and $\hat{\lambda}^{(6)} = 12.14$. The algorithm has then converged after 6 iterations and we have $\hat{\lambda}_{\text{GMM}} = 12.14$. We also obtain:

$$D = \begin{pmatrix} 67.87 \\ 22.36 \end{pmatrix}$$

and

$$S = \frac{1}{n} h\left(\hat{\lambda}_{\text{GMM}}\right) h\left(\hat{\lambda}_{\text{GMM}}\right)^\top = 10^{-4} \cdot \begin{pmatrix} 66.57 & 16.79 \\ 19.79 & 4.71 \end{pmatrix}$$

It follows that the standard deviation of $\hat{\lambda}_{\text{GMM}}$ is equal to 2.81.

If $m > K$, then there are more orthogonality conditions than parameters to estimate. We don't have $g\left(\hat{\theta}\right) = 0$, meaning that the model is over-identified. In order to test the over-identifying constraints, we consider the J-test. Under the null hypothesis $\mathcal{H}_0 : \mathbb{E}\left[h_{i,j}\left(\hat{\theta}\right)\right] = 0$, we have:

$$J = nQ\left(\hat{\theta}\right) = n g\left(\hat{\theta}\right)^\top W g\left(\hat{\theta}\right) \to \chi^2(g)$$

where $g = m - K$ is the number of over-identifying conditions. In the case of the previous example, we have $Q\left(\hat{\theta}\right) = 0.2336$ and $J = 2.336$. It follows that the p-value is equal to 12.64%. At the 90% confidence level, we then reject the null hypothesis.

The method of instrumental variables Let us consider the linear regression:

$$y_i = \sum_{k=1}^{K} \beta_k x_{i,k} + u_i$$

Since we have $\mathbb{E}[u_i] = 0$ and $\text{var}(u_i) = \sigma^2$, we deduce that:

$$h_{i,1}(\theta) = u_i = y_i - \sum_{k=1}^{K} \beta_k x_{i,k} \tag{10.18}$$

and:

$$h_{i,2}(\theta) = u_i^2 - \sigma^2 = \left(y_i - \sum_{k=1}^{K} \beta_k x_{i,k}\right)^2 - \sigma^2 \tag{10.19}$$

where the vector of parameters is equal to (β, σ). The number of parameters is then equal to $K+1$. If $K>1$, we need to find other orthogonal conditions. We recall that the linear model assumes that the residuals are orthogonal to exogenous variables. This implies that:

$$h_{i,2+k}(\theta) = u_i x_{i,k} = \left(y_i - \sum_{k=1}^{K} \beta_k x_{i,k}\right) x_{i,k} \tag{10.20}$$

We obtain a system of $K+2$ moments for $K+1$ parameters. The model is over-identified except if the linear model contains an intercept. In this case, the first and third moments are the same because $x_{i,1}=1$, and we obtain:

$$\hat{\theta}_{\text{GMM}} = \hat{\theta}_{\text{ML}} = \begin{pmatrix} \hat{\beta}_{\text{OLS}} \\ \hat{\sigma}_{\text{ML}} \end{pmatrix}$$

If the assumption $\mathbb{E}[u \mid X=x] = 0$ is not verified – or $\mathbb{E}\left[\mathbf{X}^\top \mathbf{U}\right] = 0$, we estimate the parameter β by the method of instrumental variables (IV). The underlying idea is to find a set of variables Z such that $\mathbb{E}[u \mid Z=z]=0$. Since we have:

$$\mathbf{Z}^\top \mathbf{Y} = \mathbf{Z}^\top \mathbf{X}\beta + \mathbf{Z}^\top \mathbf{U}$$

where $\mathbf{Z}$ is the $n \times K$ matrix of instrumental variables $z_{i,k}$, we deduce that:

$$\hat{\beta}_{\text{IV}} = \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbf{Z}^\top \mathbf{Y}$$

If follows that the estimator $\hat{\beta}_{\text{IV}}$ is unbiased:

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_{\text{IV}}\right] &= \mathbb{E}\left[\left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbf{Z}^\top \mathbf{Y}\right] \\
&= \mathbb{E}\left[\left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbf{Z}^\top (\mathbf{X}\beta + \mathbf{U})\right] \\
&= \beta + \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbb{E}\left[\mathbf{Z}^\top \mathbf{U}\right] \\
&= \beta
\end{aligned}$$

and the expression of its variance is:

$$\begin{aligned}
\text{var}\left(\hat{\beta}_{\text{IV}}\right) &= \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbf{Z}^\top \sigma^2 I_n \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \\
&= \sigma^2 \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1} \mathbf{Z}^\top \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{X}\right)^{-1}
\end{aligned}$$

Let us now consider the GMM estimator $\hat{\theta}_{\text{GMM}}$ defined by the two moments (10.18) and (10.19) and the following orthogonality conditions:

$$h_{i,2+k}(\theta) = u_i z_{i,k} = \left(y_i - \sum_{k=1}^{K} \beta_k x_{i,k}\right) z_{i,k} \tag{10.21}$$

We can show that $\hat{\beta}_{\text{GMM}} = \hat{\beta}_{\text{IV}}$. The method of instrumental variables is then a special case of the generalized method of moments[9].

[9] We also notice that the previous analysis is also valid for non-linear models:

$$y_i = f(x_i, \beta) + u_i$$

We just have to replace u_i by the expression $y_i - f(x_i, \beta)$ in Equations (10.18), (10.19) and (10.21).

Some examples The general form of a ARCH model is (Engle, 1982):

$$\Phi(L) y_t = x_t^\top \beta + u_t$$

where $u_t \sim \mathcal{N}\left(0, h_t^2\right)$ and $h_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j u_{t-j}^2$. The conditional variance is then an autoregressive process[10]. The first two moments are $\mathbb{E}[u_t] = 0$ and $\mathbb{E}\left[u_t^2 - h_t^2\right] = 0$. We can also impose the K orthogonality conditions $\mathbb{E}[u_t x_{t,k}] = 0$ for all $k = 1, \ldots, K$. However, these different moments are not enough for estimating the parameters α_j. This is why we have to impose the q orthogonality conditions between the centered innovations $u_t^2 - h_t^2$ and the lagged residuals u_{t-j}^2:

$$\mathbb{E}\left[\left(u_t^2 - h_t^2\right) u_{t-j}^2\right] = 0$$

for $j = 1, \ldots, q$. This estimation method based on GMM has been suggested by Mark (1988).

In the standard linear time series model $y_t = x_t^\top \beta + u_t$, we assume that $u_t \sim \mathcal{N}\left(0, \sigma^2\right)$. If we also assume that the residuals are autocorrelated, we obtain:

$$\begin{aligned}
\mathbb{E}[u_t u_{t-1}] &= \mathbb{E}[(\rho u_{t-1} + \varepsilon_t) u_{t-1}] \\
&= \rho \mathbb{E}\left[u_{t-1}^2\right] + \mathbb{E}[\varepsilon_t u_{t-1}] \\
&= \rho \sigma^2
\end{aligned}$$

The GMM estimator consists in using Equations (10.18), (10.19) and (10.20) and adding the following orthogonality moment:

$$h_{t,K+3}(\theta) = u_t u_{t-1} - \rho \sigma^2$$

where the vector of parameters θ becomes (β, σ, ρ).

Let $h_t(\theta) = (h_{t,1}(\theta), \ldots, h_{t,m}(\theta))$ be the $m \times 1$ vector of empirical moments for the observation date t. We have seen that the estimate of $S = \mathbb{E}\left[h\left(\hat{\theta}\right) h\left(\hat{\theta}\right)^\top\right]$ is equal to:

$$\hat{S} = \frac{1}{T} \sum_{t=1}^T h_t\left(\hat{\theta}\right) h_t\left(\hat{\theta}\right)^\top$$

This implies the assumption that the empirical moments are not autocorrelated. However, when dealing with time series, this hypothesis is generally not satisfied and it is better to use a heteroscedasticity and autocorrelation consistent (HAC) estimator:

$$\hat{S} = \Gamma(0) + \sum_{j=1}^\ell w_j \left(\Gamma(j) + \Gamma(j)^\top\right)$$

where:

$$\Gamma(j) = \frac{1}{T} \sum_{t=j+1}^T h_t\left(\hat{\theta}\right) h_{t-j}\left(\hat{\theta}\right)^\top$$

and w_j is the weight function. Newey and West (1987) showed that the Bartlett kernel defined as $w_j = 1 - j/(\ell + 1)$ is a simple method that is consistent under fairly general conditions.

[10] The term ARCH means *autoregressive conditional heteroscedasticity*.

10.1.3.3 Simulated method of moments

When a model is complex, it may be difficult to find an analytical expression for each moment condition. The basic idea behind the simulated method of moments (SMM) is then to simulate the model and to replace the theoretical moments by the simulated moments. The theory of SMM has been formulated by McFadden (1989), Pakes and Pollard (1989) and Duffie and Singleton (1993), and this approach is particularly popular for dynamic asset pricing models.

If we consider the method of moments, we have:

$$g_j(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^j - \hat{m}_j(\theta)\right)$$

where $\hat{m}_j(\theta)$ is the j^{th} simulated moment computed with n_S simulations. For the generalized method of moments, we obtain:

$$g_j(\theta) = \frac{1}{n}\sum_{i=1}^{n}\hat{h}_{i,j}(\theta)$$

where $\hat{h}_{i,j}(\theta)$ is the j^{th} simulated orthogonal condition for the i^{th} observation. The SMM estimator is then defined exactly as the GMM estimator:

$$\hat{\theta} = \arg\min g(\theta)^{\top} W g(\theta)$$

Like the GMM estimator, the SMM estimator $\hat{\theta}$ is asymptotically normally distributed (Duffie and Singleton, 1993):

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}\left(\mathbf{0}, \left(1 + \frac{n}{n_S}\right)V\right)$$

where $V = \left(D^{\top}S^{-1}D\right)^{-1}$ is the GMM covariance matrix and n_S is the number of simulations. Therefore, the covariance matrix of $\hat{\theta}$ depends on the ratio $\tau = n/n_S$:

$$\text{var}\left(\hat{\theta}\right) = \frac{1+\tau}{n}\left(D^{\top}S^{-1}D\right)^{-1} \tag{10.22}$$

This implies that the number of simulations n_S must be larger than the number of observations:

$$n_S \gg n$$

Remark 123 *The key point when considering the simulated method of moments is to use the same random numbers at each iteration of the optimization algorithm in order to ensure the convergence of the SMM estimator.*

In order to illustrate the simulated method of moments, we consider Example 106 on page 629. The simulated moments are:

$$g_1(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \frac{1}{n_s}\sum_{s=1}^{n_s}y_{(s)}\right)$$

and:

$$g_2(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^2 - \frac{1}{n_s}\sum_{s=1}^{n_s}y_{(s)}^2\right)$$

TABLE 10.8: Comparison of GMM and SMM estimates

Method	n_S	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\sigma}(\hat{\mu})$	$\hat{\sigma}(\hat{\sigma})$	$\hat{\sigma}_{\text{MC}}(\hat{\mu})$	$\hat{\sigma}_{\text{MC}}(\hat{\sigma})$
GMM		−0.109	0.836	0.306	0.165		
SMM	25	−0.115	0.880	0.373	0.218	0.214	0.340
SMM	200	−0.102	0.847	0.315	0.172	0.070	0.062

where $y_{(s)}$ is a simulated value of $Y = XU$ where $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $U \sim \mathcal{U}_{[0,1]}$ and $X \perp U$. We have:

$$y_{(s)} = x_{(s)} \cdot u_{(s)} = \left(\mu + \sigma n_{(s)}\right) \cdot u_{(s)}$$

where $u_{(s)}$ and $n_{(s)}$ are uniform and standard normal random numbers. In Table 10.8, we compare the results obtained with the GMM and the SMM approaches. For the simulated method of moments, we consider 20 000 Monte Carlo replications and report the average values. We notice that when the number of simulations is low, the estimator can be biased. For example, $\hat{\sigma}$ is equal to 0.880 on average when n_S is equal to 25, whereas the GMM estimate is equal to 0.836. We also notice that the standard errors $\hat{\sigma}(\hat{\mu})$ and $\hat{\sigma}(\hat{\sigma})$ of the estimated parameters are higher for the SMM estimator than for the GMM estimator because of the factor τ. However, these results are based on the asymptotic theory. When the number of observations is low ($n = 10$), the approximation of the covariance matrix by Equation (10.22) is not valid. For instance, if we calculate the standard errors by calculating the standard deviation of the MC estimates, we obtain the values $\hat{\sigma}_{\text{MC}}(\hat{\mu})$ and $\hat{\sigma}_{\text{MC}}(\hat{\sigma})$ given in the last two columns. When the number of simulations is large and the number of observations is small, the asymptotic theory then overestimates the standard errors. Indeed, in the case where $n_S = 200$, the MC standard errors are $\hat{\sigma}_{\text{MC}}(\hat{\mu}) = 0.070$ and $\hat{\sigma}_{\text{MC}}(\hat{\sigma}) = 0.062$ whereas we obtain $\hat{\sigma}(\hat{\mu}) = 0.315$ and $\hat{\sigma}(\hat{\sigma}) = 0.172$ calculated with Equation (10.22).

In Chapter 5 dedicated to the operational risk, we have seen how to estimate the parameters θ of the severity distribution with the method of maximum likelihood or the generalized method of moments. We recall that $\{x_1, \ldots, x_T\}$ is the sample of losses collected for a given cell of the operational risk matrix. If we assume that the losses are log-normal distributed, the orthogonal conditions are:

$$\begin{cases} h_{i,1}(\theta) = x_i - e^{\mu + \frac{1}{2}\sigma^2} \\ h_{i,2}(\theta) = \left(x_i - e^{\mu + \frac{1}{2}\sigma^2}\right)^2 - e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \end{cases}$$

Let us now assume that we do not collect individual losses but aggregated losses. In this case, the sample is defined by $\{(n_1, x_1), \ldots, (n_T, x_T)\}$ where n_i is the number of individual losses and x_i is the aggregated loss for the i^{th} observation. Since the individual losses are independent, the orthogonal conditions become:

$$\begin{cases} h_{i,1}(\theta) = x_i - n_i e^{\mu + \frac{1}{2}\sigma^2} \\ h_{i,2}(\theta) = \left(x_i - n_i e^{\mu + \frac{1}{2}\sigma^2}\right)^2 - n_i e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \end{cases}$$

We have seen that data collection in operational risk is impacted by truncation, because data are recorded only when their amounts are higher than a threshold H. On page 320, we were able to calculate the theoretical moments of truncated individual losses. However, it is impossible to find the theoretical moments of truncated aggregated losses and the generalized method of moments cannot be applied. Nevertheless, we can consider the simulated

method of moments with the following orthogonal conditions:

$$\begin{cases} h_{i,1}(\theta) = x_i - \hat{m}_{i,1}(\theta) \\ h_{i,2}(\theta) = (x_i - \hat{m}_{i,1}(\theta))^2 - \hat{m}_{i,2}(\theta) \end{cases}$$

where:

$$\hat{m}_{i,1}(\theta) = \frac{1}{\sum_{s=1}^{n_S} \mathbb{1}\{X_i(s) \geq H\}} \sum_{s=1}^{n_S} \mathbb{1}\{X_i(s) \geq H\} \cdot X_i(s)$$

and:

$$\hat{m}_{i,2}(\theta) = \frac{1}{\sum_{s=1}^{n_S} \mathbb{1}\{X_i(s) \geq H\}} \sum_{s=1}^{n_S} \mathbb{1}\{X_i(s) \geq H\} \cdot X_i^2(s) - \hat{m}_{i,1}^2(\theta)$$

In the case of the log-normal distribution, each aggregated loss is simulated as follows:

$$X_i(s) = \sum_{j=1}^{n_i} e^{\mu + \sigma \cdot u_{(s)}}$$

where $u_{(s)} \sim \mathcal{N}(0,1)$. Let us consider an example with 20 observations: 1 404, 1 029, 2 607, 2 369, 2 163, 2 730, 4 045, 1 147, 2 319, 2 521, 2 021, 1 528, 1 715, 2 547, 1 039, 1 853, 3 515, 1 273, 2 048 and 2 744. Each observation corresponds to an aggregated sum of 5 individual losses[11]. If individual losses are log-normal distributed, we obtain the following results: $\hat{\mu}_{\text{GMM}} = 5.797$, $\hat{\sigma}_{\text{GMM}} = 0.718$, $\hat{\mu}_{\text{SMM}} = 5.796$ and $\hat{\sigma}_{\text{SMM}} = 0.722$. We notice that the SMM estimates are close to the GMM estimates[12]. If we now assume that the aggregated losses have been collected above the threshold $H = 1\,000$, we obtain $\hat{\mu}_{\text{SMM}} = 5.763$ and $\hat{\sigma}_{\text{SMM}} = 0.745$. Therefore, the effect of truncation has changed the estimated parameters.

10.1.4 Non-parametric estimation

In the previous paragraphs, we specify a parametric model, that is a statistical model which depends on a parameter vector θ. Therefore, the estimation of the model consists in estimating the parameter vector θ. Using $\hat{\theta}$, we can determine some quantity of interest, for example probabilities, quantiles and expectations. In the case of non-parametric models, we directly estimate the quantity of interest without specifying a probability distribution or a statistical model.

10.1.4.1 Non-parametric density estimation

Histogram estimator Let X be a random variable with continuous distribution function $\mathbf{F}(x)$. Using the sample $\{x_1, \ldots, x_n\}$, we can estimate $\mathbf{F}(x)$ by the empirical distribution function:

$$\hat{\mathbf{F}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\}$$

where $\hat{\mathbf{F}}(x)$ is the percentage of observations that are lower than x. If we now consider the estimation of the density $f(x)$ from a sample $\{x_1, \ldots, x_n\}$, we have $\mathrm{d}\hat{\mathbf{F}}(x) = \hat{f}(x)\,\mathrm{d}x$ and

[11] We have $n_i = 5$.

[12] Of course, the SMM estimates depend on the number of simulations and the seed of the random number generators. Here, the results have been obtained with 2 000 simulations.

we deduce that:

$$\begin{aligned}\hat{f}(x) &= \frac{\mathrm{d}\hat{\mathbf{F}}(x)}{\mathrm{d}x} \\ &\simeq \frac{\hat{F}(x+h)-\hat{F}(x-h)}{2h} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2h}\cdot\mathbb{1}\{x-h\leq x_i\leq x+h\}\end{aligned}$$

The density estimator $\hat{f}(x)$ is known as the histogram. It counts the percentage of observations that belong to the interval $[x-h, x+h]$. The issue with the histogram estimator is that the density function $\hat{f}(x)$ is not smooth and is sensitive to the bandwidth h. To obtain a continuous density, we can specify a parametric density function $f(x;\theta)$, estimate the parameter θ by maximum likelihood and assume that $\hat{f}(x) = f\left(x;\hat{\theta}_{\mathrm{ML}}\right)$. By construction, the estimator $f\left(x;\hat{\theta}_{\mathrm{ML}}\right)$ is continuous, but it is biased because the statistical model $f(x;\theta)$ is not necessarily the right model.

Kernel estimator We notice that the previous estimator $\hat{f}(x)$ can be written as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}\left(\frac{x-x_i}{h}\right) \tag{10.23}$$

where $\mathcal{K}$ is the uniform density function on $[-1,1]$:

$$\mathcal{K}(u) = \frac{1}{2}\cdot\mathbb{1}\{-1\leq u\leq 1\}$$

$\mathcal{K}$ is also called a rectangular kernel. The idea is then to replace this function by other window functions that are sufficiently smoothed and satisfy some properties (Silverman, 1986):

- $\mathcal{K}(u)\geq 0$ to ensure the positivity of the density;
- $\int_{-\infty}^{\infty}\mathcal{K}(u)\,\mathrm{d}u = 1$ to verify that $\hat{\mathbf{F}}(\infty) = 1$.

Moreover, the symmetry property $\mathcal{K}(u) = \mathcal{K}(-u)$ is generally added. We can then show that:

$$\mathbb{E}\left[\hat{f}_n(x) - f(x)\right] \approx \frac{h^2}{2}f''(x)\int_{-\infty}^{\infty}u^2\mathcal{K}(u)\,\mathrm{d}u$$

and:

$$\mathrm{var}\left(\hat{f}_n(x)\right) \approx \frac{1}{nh}f(x)\int_{-\infty}^{\infty}\mathcal{K}^2(u)\,\mathrm{d}u$$

The bias of $\hat{f}_n(x)$ depends then on the second moment of the kernel $\mu_2'(\mathcal{K}) = \int_{-\infty}^{\infty}u^2\mathcal{K}(u)\,\mathrm{d}u$ whereas the variance of $\hat{f}_n(x)$ is related to the roughness of the kernel $R(\mathcal{K}) = \int_{-\infty}^{\infty}\mathcal{K}^2(u)\,\mathrm{d}u$. We also notice that the bias is proportional to the curvature $f''(x)$ and the variance is inversely proportional to the number of observations n. The most popular kernel functions are the Gaussian kernel — we have $\mathcal{K}(u) = \phi(u)$ and $\mathcal{I}(u) = \Phi(u)$ — and the Epanechnikov kernel[13] — we have $\mathcal{K}(u) = \frac{3}{4}\cdot\left(1-u^2\right)\cdot\mathbb{1}\{|u|\leq 1\}$

[13] The Epanechnikov kernel is often called the optimal kernel because it minimizes the mean squared error.

and $\mathcal{I}(u) = \min\left(\frac{1}{4} \cdot \left(3u - u^3 + 2\right) \cdot \mathbb{1}\{u > -1\}, 1\right)$. The difficulty is the choice of the bandwidth h since there is an arbitrage between bias and variance (Jones *et al.*, 1996). For the Gaussian kernel, a rule-of-thumb is $h = 1.06 \cdot \hat{\sigma} \cdot n^{-1/5}$ where $\hat{\sigma}$ is the standard deviation of the sample $\{x_1, \ldots, x_n\}$.

Remark 124 *From Equation (10.23), we deduce that:*

$$\begin{aligned}\hat{\mathbf{F}}(x) &= \int_{-\infty}^{x} \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}\left(\frac{t - x_i}{h}\right) \mathrm{d}t \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}\left(\frac{x - x_i}{h}\right) \end{aligned} \tag{10.24}$$

where $\mathcal{I}(u) = \int_{-\infty}^{u} \mathcal{K}(t)\, \mathrm{d}t$ *is the integrated kernel.*

In the case of a multivariate distribution function, we have:

$$\hat{f}(x_1, \ldots, x_m) = \frac{1}{nh^m} \sum_{i=1}^{n} \mathcal{K}\left(\frac{x_1 - x_{1,i}}{h}, \ldots, \frac{x_m - x_{m,i}}{h}\right)$$

where $\mathcal{K}(u)$ is now a multidimensional kernel function[14] that satisfies $\mathcal{K}(u) \geq 0$, $\int \mathcal{K}(u)\, \mathrm{d}u = 1$, $\int u\mathcal{K}(u)\, \mathrm{d}u = 0$ and $\int uu^\top \mathcal{K}(u)\, \mathrm{d}u = I_m$. Generally, the multidimensional kernel $\mathcal{K}(u)$ is defined as the product of univariate kernels $\mathcal{K}(u) = \prod_{j=1}^{m} \mathcal{K}_j(u_j)$ where $\mathcal{K}_j$ is the kernel for the j^{th} marginal distribution.

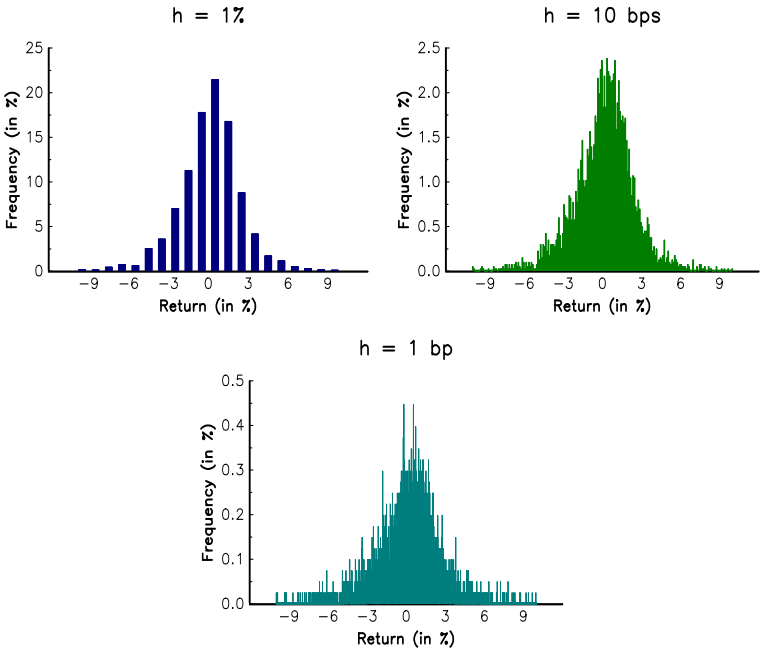


FIGURE 10.5: Histogram of the weekly returns of the S&P 500 index

[14] We have $u = (u_1, \ldots, u_m)$.

An example In Figure 10.5, we have reported the histogram of weekly returns R_t of the S&P 500 index[15] for different values of h. We verify that the histogram is very sensitive to the parameter h. We now consider the estimation of the probability density function of R_t. For that, we consider 4 statistical models:

1. the first model assumes that weekly returns are Gaussian;
2. the second model assumes that weekly returns are distributed according to the Student's t_4 distribution;
3. the third model is a variant of the second model, but the Student's t distribution has 2 degrees of freedom;
4. finally, we use a Gaussian kernel with $h = 1\%$ for the fourth model.

Results are reported in Figure 10.6. In order to illustrate the model choice on distribution tails, we consider the order statistic $1:52$, meaning that we estimate the density function of the worst weekly return over one year. On page 755, we show that the distribution function of the random variable $X_{1:n}$ is equal to $\mathbf{F}_{1:n}(x) = 1 - (1 - \mathbf{F}(x))^n$. We deduce that the estimated probability density function is equal to:

$$\hat{f}_{1:n}(x) = n\left(1 - \hat{\mathbf{F}}(x)\right)^{n-1} \hat{f}(x)$$

Results are given in Figure 10.7. We notice that we observe significant differences between the four models. Since the non-parametric estimation is less biased, we conclude that the Student's t_4 distribution is the more appropriate parametric distribution function.

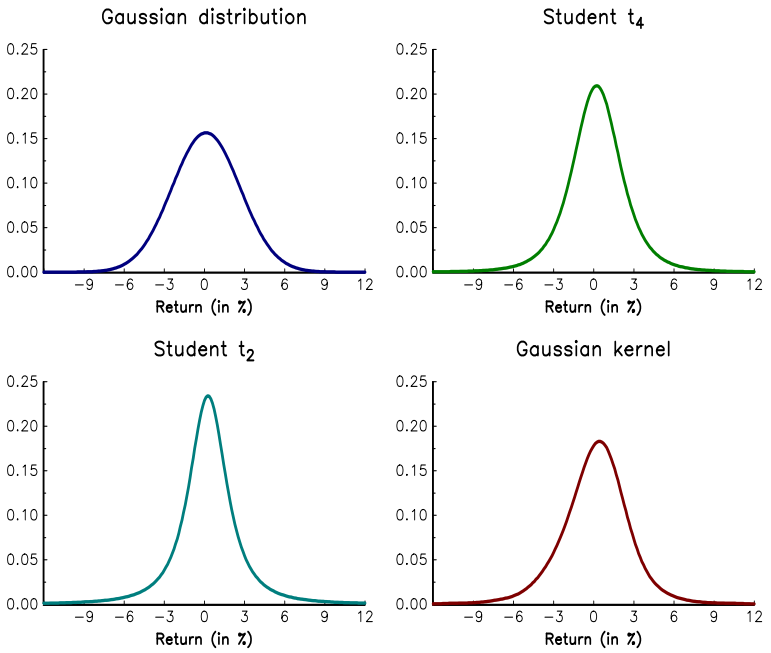


FIGURE 10.6: Density estimation of the weekly returns of the S&P 500 index

[15]The study period is January 2000 – December 2015.

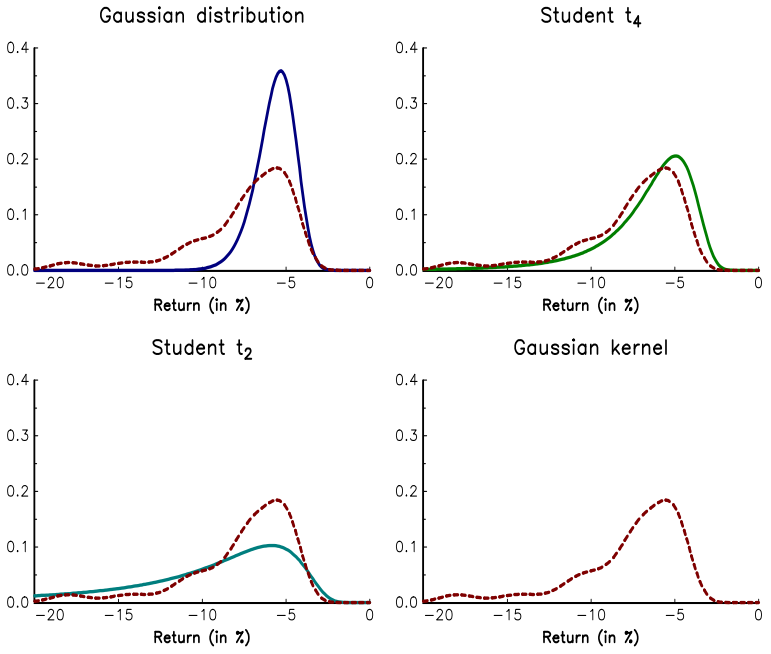


FIGURE 10.7: Density estimation of the worst weekly return over one year

10.1.4.2 Non-parametric regression

Nadaraya-Watson regression On page 603, we have shown that the conditional expectation problem is to find the function $m(x)$ such that:

$$y = \mathbb{E}[Y \mid X = x] = m(x)$$

In the case where X is a scalar random variable, it follows that:

$$\begin{aligned} m(x) &= \int_{\mathbb{R}} y f_{y|x}(y;x)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} y \frac{f_{x,y}(x,y)}{f_x(x)}\,\mathrm{d}y \\ &= \frac{\int_{\mathbb{R}} y f_{x,y}(x,y)\,\mathrm{d}y}{f_x(x)} \end{aligned}$$

where $f_{x,y}$ is the joint density of (X,Y), f_x is the density of X and $f_{y|x}$ is the conditional density of Y given that $X = x$. We deduce that an estimator of $m(x)$ is:

$$\begin{aligned} \hat{m}(x) &= \frac{\int_{\mathbb{R}} y \hat{f}_{x,y}(x,y)\,\mathrm{d}y}{\hat{f}_x(x)} \\ &= \frac{\sum_{i=1}^n \mathcal{K}_x\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^n \mathcal{K}_x\left(\frac{x-x_i}{h}\right)} \end{aligned}$$

because we have:

$$\begin{aligned}\int_{\mathbb{R}} y\hat{f}_{x,y}(x,y)\,\mathrm{d}y &= \frac{1}{nh^2}\int_{\mathbb{R}} y\sum_{i=1}^{n}\mathcal{K}_{x,y}\left(\frac{x-x_i}{h},\frac{y-y_i}{h}\right)\mathrm{d}y \\ &= \frac{1}{nh^2}\int_{\mathbb{R}} y\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)\mathcal{K}_y\left(\frac{y-y_i}{h}\right)\mathrm{d}y \\ &= \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)\frac{1}{h}\int_{\mathbb{R}} y\,\mathcal{K}_y\left(\frac{y-y_i}{h}\right)\mathrm{d}y \\ &= \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)y_i\end{aligned}$$

and:

$$\hat{f}_x(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)$$

Finally, we deduce that $\hat{m}(x)$ is a weighted sum of the sample data $\{y_1,\ldots,y_n\}$ of Y:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} w_i\cdot y_i}{\sum_{i=1}^{n} w_i}$$

where the weights are the kernel values applied to the sample data $\{x_1,\ldots,x_n\}$ of X:

$$w_i = \mathcal{K}_x\left(\frac{x-x_i}{h}\right)$$

Local polynomial regression We consider the following least squares problem:

$$\hat{\beta}_0(x) = \underset{\beta_0\in\mathbb{R}}{\arg\min}\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)(y_i-\beta_0)^2$$

The first-order condition is:

$$\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)(y_i-\beta_0) = 0$$

It follows that $\hat{\beta}_0(x) = \hat{m}(x)$. Therefore, the Nadaraya-Watson regression is a weighted regression with a constant:

$$y_i = \beta_0 + u_i$$

where $w_i(x) = \mathcal{K}_x\left(\frac{x-x_i}{h}\right)$. Since the weights $w_i(x)$ depend on the value x, $\hat{\beta}_0$ depends also on the value x and we use the notation $\hat{\beta}_0(x)$.

We can extend the local constant model to the local polynomial model:

$$y_i = \beta_0 + \sum_{j=1}^{p}\beta_j(x_i-x)^j + u_i$$

The least squares problem becomes:

$$\hat{\beta}(x) = \arg\min\sum_{i=1}^{n}\mathcal{K}_x\left(\frac{x-x_i}{h}\right)\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j(x_i-x)^j\right)^2$$

We know that the WLS estimate is:

$$\hat{\beta}(x) = \left(\mathbf{X}^\top W \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y}$$

where W is a diagonal matrix with $W_{i,i} = w_i(x)$, $\mathbf{Y}$ is the column vector $(y_1, \ldots, y_n)$ and $\mathbf{X}$ is the design matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & \cdots & (x_1 - x)^p \\ 1 & (x_2 - x) & (x_2 - x)^2 & & (x_2 - x)^p \\ & & & \ddots & \\ 1 & (x_n - x) & (x_n - x)^2 & & (x_n - x)^p \end{pmatrix}$$

Moreover, we have:

$$\begin{aligned} \mathbb{E}[Y \mid X = x] &= \mathbb{E}\left[\beta_0 + \sum_{j=1}^{p} \beta_j (X - x)^j + U \,\middle|\, X = x\right] \\ &= \beta_0 \end{aligned}$$

We deduce that the conditional expectation is again equal to the intercept of the weighted polynomial regression:

$$\hat{m}(x) = \hat{\beta}_0(x)$$

Application to quantile regression On page 613, we have seen that the quantile regression:

$$\Pr\{Y \leq q_\alpha(x) \mid X = x\} = \alpha$$

can be formulated as a M-estimator:

$$\hat{q}_\alpha(x) = \hat{\beta}_0 + \hat{\beta} x$$

where:

$$\left(\hat{\beta}_0, \hat{\beta}\right) = \arg\min \sum_{i=1}^{n} \rho(y_i - \beta_0 - \beta x_i)$$

and $\rho(u) = \chi_\alpha(u) = u \cdot (\alpha - \mathbb{1}\{u < 0\})$. Let us consider the local polynomial regression:

$$\hat{\beta}(x) = \arg\min \sum_{i=1}^{n} \mathcal{K}_x\left(\frac{x - x_i}{h}\right) \rho\left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j (x_i - x)^j\right)$$

We deduce that:

$$\hat{q}_\alpha(x) = \hat{\beta}_0(x)$$

To estimate the conditional quantile $\hat{q}_\alpha(x)$, we can then use a classical quantile regression procedure by considering the regressor values $(x_i - x)^j$ for $j = 1, \ldots, p$ and the weighting matrix based on the kernel function.

Examples We consider the additive noise model:

$$y = \sin(9x) + u \tag{10.25}$$

where $u \sim \mathcal{U}_{[-0.5,0.5]}$. The conditional expectation function is then $m(x) = \sin(9x)$ whereas the conditional quantile function is $q_\alpha(x) = \sin(9x) + \alpha - \frac{1}{2}$. We simulate 1 000 realizations

of Model (10.25) and estimate the functions $m(x)$ and $q_\alpha(x)$. In Figure 10.8, we have reported the estimated functions $\hat{m}(x)$ and $\hat{q}_\alpha(x)$ for different models. We notice that the Nadaraya-Watson estimator is less efficient than the local quadratic estimator. In the case of the quantile regression, we find that the local linear regression gives correct results when the function is quasi-affine. When the second derivative is high enough, results are much more questionable, and local quadratic regression seems more suitable. We obtain a similar conclusion with the multiplicative noise model:

$$y = (\cos(2\pi x - \pi) + 1)\, u \tag{10.26}$$

where $u \sim \mathcal{U}_{[0,1]}$. We have $m(x) = \frac{1}{2}(\cos(2\pi x - \pi) + 1)$ and $q_\alpha(x) = \alpha(\cos(2\pi x - \pi) + 1)$. In Figure 10.9, we report the estimates of the quantile regression based on 1 000 simulations. We notice that the local quadratic quantile regression may also present some bias. This is due to the small size of the sample ($n = 1\,000$) compared to the quantile level ($\alpha = 90\%$). Indeed, non-parametric quantile regression may need a big sample size to converge.

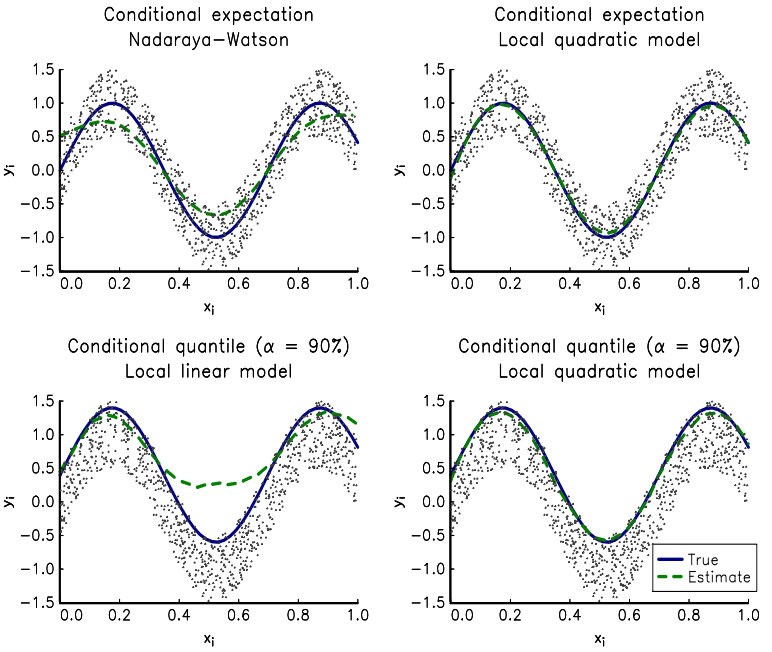


FIGURE 10.8: Non-parametric regression of the additive model

10.2 Time series modeling

10.2.1 ARMA process

10.2.1.1 The VAR(1) process

Let y_t be a n-dimensional process and $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$. We define the vector autoregression model (VAR) of order one as follows:

$$y_t = \mu + \Phi y_{t-1} + \varepsilon_t \tag{10.27}$$

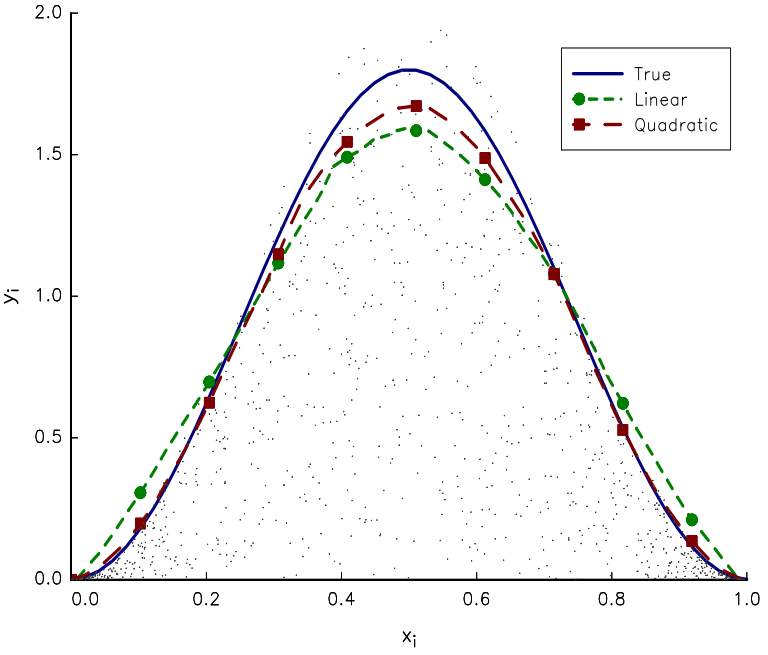


FIGURE 10.9: Non-parametric regression of the multiplicative model

The VaR(1) model is very interesting from a computational viewpoint because matrix calculus is simple. Moreover, the computations done in this paragraph are straightforward to extend to more complex processes like VARMA or state space models.

We say that the process is stable if the eigenvalues of Φ have modulus less than one or the characteristic polynomial of $|I_n - \Phi z|$ has its roots outside the unit circle. Under the stability assumption[16], the process has an infinite vector moving average (VMA) representation[17]:

$$y_t = (I_n - \Phi)^{-1} \mu + \sum_{i=0}^{\infty} \Phi^i \varepsilon_{t-i}$$

To compute the covariance matrix $\Gamma_y(k) = \mathbb{E}\left[y_t y_{t-k}^\top\right] - \mathbb{E}[y_t]\,\mathbb{E}\left[y_{t-k}^\top\right]$, we first calculate $\Gamma_y(0)$ by using the relationship:

$$\operatorname{vec}(\Gamma_y(0)) = (I_{n^2} - \Phi \otimes \Phi)^{-1} \operatorname{vec}(\Sigma)$$

Then, we calculate $\Gamma_y(k)$ by recursion:

$$\begin{aligned}
\Gamma_y(k) &= \mathbb{E}\left[y_t y_{t-k}^\top\right] - \mathbb{E}[y_t]\,\mathbb{E}\left[y_{t-k}^\top\right] \\
&= \mathbb{E}\left[\mu y_{t-k}^\top + \Phi y_{t-1} y_{t-k}^\top + \varepsilon_t y_{t-k}^\top\right] - \mu\mu^\top \\
&= \Phi \mathbb{E}\left[y_{t-1} y_{t-k}^\top\right] \\
&= \Phi \Gamma_y(k-1)
\end{aligned}$$

[16] We can show that the stability assumption implies the stationarity assumption (Lütkepohl, 2005).

[17] We have:

$$\lim_{i \to \infty} \Phi^i = (I_n - \Phi)^{-1}$$

Impulse response analysis describes the dynamic evolution of the system after a shock in one variable. For instance, the responses to forecast errors are "*the effect of an innovation of one variable to another variable*" (Lütkepohl, 2005). The matrix of responses after k periods is $\Lambda_k = \Phi^k$ and we note $\Psi_k = \sum_{j=0}^{k} \Lambda_j$ the matrix of cumulated responses. Sometimes, it is better to consider normalized innovations than forecast errors. In this case, we define the responses to orthogonal impulses by $\Delta_k = \Phi^k P(\Sigma)$ and the matrix of cumulated responses by $\Xi_k = \sum_{j=0}^{k} \Delta_k$ where $P(\Sigma)$ is the Cholesky decomposition matrix of Σ.

10.2.1.2 Extension to ARMA models

A vector ARMA(p,q) process is defined by:

$$y_t - \sum_{i=1}^{p} \Phi_i y_{t-i} = \beta x_t + \varepsilon_t - \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i} \tag{10.28}$$

where y_t is a process of dimension n and $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$. x_t is a $K \times 1$ vector of exogenous variables. The dimension of matrices Φ_i and Θ_i is $n \times n$ whereas β is a matrix with dimension $n \times K$. The parameters p and q are the orders of the autoregressive and moving average polynomials.

Remark 125 *In the econometric literature, the process (10.28) is known as the VARMAX model because it is a vector process and it contains exogenous variables. If the order q is equal to zero – or $\Theta_i = \mathbf{0}$, then we obtain a VARX model. A VMAX model corresponds to the case $p = 0$ or $\Phi_i = \mathbf{0}$. When there is no exogenous variable except a constant, we use the terms VARMA, VAR and VMA. The terms ARMAX, ARMA, AR and MA are often reserved for the one-dimensional case $n = 1$.*

Let us consider the case where exogenous variables reduce to a constant. We have:

$$\left(I_n - \sum_{i=1}^{p} \Phi_i L^i\right) y_t = \mu + \left(I_n - \sum_{i=1}^{q} \Theta_i L^i\right) \varepsilon_t \tag{10.29}$$

where L is the lag operator. We can write this process as a VAR(1) model:

$$\alpha_t = c + T\alpha_{t-1} + u_t \tag{10.30}$$

where $\alpha_t = (y_t, \ldots, y_{t-p+1}, \varepsilon_t, \ldots, \varepsilon_{t-q+1})$ is a process of dimension $n(p+q)$. The residuals u_t are equal to $R\varepsilon_t$ where the matrix R is equal to $\begin{pmatrix} I_n & \mathbf{0} & \cdots & \mathbf{0} & I_n & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}^\top$ and has the dimension $n(p+q) \times n$. The vector c is equal to $\begin{pmatrix} \mu & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}^\top$. The dimension of the matrix T is $n(p+q) \times n(p+q)$ and we have:

$$T = \begin{pmatrix}
\Phi_1 & \cdots & \Phi_{p-1} & \Phi_p & \Theta_1 & \cdots & \Theta_{q-1} & \Theta_q \\
I_n & & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \\
 & \ddots & & \vdots & \vdots & & \mathbf{0} & \\
\mathbf{0} & \cdots & I_n & \mathbf{0} & \mathbf{0} & & & \\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
 & & & \mathbf{0} & I_n & & \mathbf{0} & \mathbf{0} \\
 & \mathbf{0} & & \vdots & & \ddots & & \vdots \\
 & & & \mathbf{0} & \mathbf{0} & & I_n & \mathbf{0}
\end{pmatrix}$$

We also notice that $y_t = Z\alpha_t$ where the matrix $Z = \begin{pmatrix} I_n & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}$ has the dimension $n \times n\,(p+q)$. Using the VAR(1) representation, we deduce easily the expression of autocovariance matrices, the responses to forecast errors or the responses to orthogonal impulses of α_t and y_t.

We can also write the previous VARMA process as a state space model:

$$\begin{cases} y_t = Z\alpha_t \\ \alpha_t = T\alpha_{t-1} + c + R\varepsilon_t \end{cases} \tag{10.31}$$

This representation is very useful since the analysis of state space models also applies to VARMA process. For instance, it is standard to estimate the parameters of VARMA models by using the Kalman filter, which is presented below. For a VAR(p) process, it is better to use closed-form formulas. The VAR(p) process is defined as follows:

$$y_t = \mu + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \varepsilon_t$$

Using the notations $Y = \begin{pmatrix} y_1 & \cdots & y_T \end{pmatrix}$, $B = \begin{pmatrix} \mu & \Phi_1 & \cdots & \Phi_p \end{pmatrix}$, $X_t = (1, y_t, \ldots, y_{t-p+1})$ and $X = \begin{pmatrix} X_0 & X_1 & \cdots & X_{T-1} \end{pmatrix}$, Lütkepohl (2005) showed that:

$$\hat{B} = YX^\top \left(XX^\top\right)^{-1}$$

and:

$$\hat{\Sigma} = \frac{1}{T - Kp - 1} Y \left(I_T - X^\top \left(XX^\top\right)^{-1} X\right) Y^\top$$

10.2.2 State space models

10.2.2.1 Specification and estimation of state space models

A state space model (SSM) includes a measurement equation and a transition equation. In the measurement equation, we define the relationship between an observable system and state variables, whereas the transition equation describes the dynamics of state variables. Generally, the state vector α_t is generated by a Markov linear process[18]:

$$\alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t \tag{10.32}$$

where α_t is a $m \times 1$ vector, T_t is a $m \times m$ matrix, c_t is a $m \times 1$ vector and R_t is a $m \times p$ matrix. In the case of a linear SSM, the measurement equation is given by:

$$y_t = Z_t\alpha_t + d_t + \epsilon_t \tag{10.33}$$

where y_t is a n-dimensional time series, Z_t is a $n \times m$ matrix, d_t is a $n \times 1$ vector. We also assume that η_t and ϵ_t are two independent white noise processes of dimension p and n with covariance matrices Q_t and H_t.

Kalman filtering In the state space model, the variable y_t is observable, but it is generally not the case of the state vector α_t. The Kalman filter is a statistical tool to estimate the distribution function of α_t. Let $\alpha_0 \sim \mathcal{N}(\hat{\alpha}_0, P_0)$ the initial position of the state vector. We note $\hat{\alpha}_{t|t}$ (or $\hat{\alpha}_t$) and $\hat{\alpha}_{t|t-1}$ the optimal estimators of α_t given the available information until time t and $t-1$:

$$\begin{aligned} \hat{\alpha}_{t|t} &= \mathbb{E}\left[\alpha_t \mid \mathcal{F}_t\right] \\ \hat{\alpha}_{t|t-1} &= \mathbb{E}\left[\alpha_t \mid \mathcal{F}_{t-1}\right] \end{aligned}$$

[18] The presentation is based on the book of Harvey (1990).

$P_{t|t}$ (or P_t) and $P_{t|t-1}$ are the covariance matrices associated to $\hat{\alpha}_{t|t}$ and $\hat{\alpha}_{t|t-1}$:

$$\begin{aligned} P_{t|t} &= \mathbb{E}\left[\left(\hat{\alpha}_{t|t}-\alpha_t\right)\left(\hat{\alpha}_{t|t}-\alpha_t\right)^\top\right] \\ P_{t|t-1} &= \mathbb{E}\left[\left(\hat{\alpha}_{t|t-1}-\alpha_t\right)\left(\hat{\alpha}_{t|t-1}-\alpha_t\right)^\top\right] \end{aligned}$$

These different quantities are calculated thanks to the Kalman filter, which consists in a recursive algorithm[19] (Harvey, 1990):

$$\begin{cases} \hat{\alpha}_{t|t-1} = T_t\hat{\alpha}_{t-1|t-1} + c_t \\ P_{t|t-1} = T_tP_{t-1|t-1}T_t^\top + R_tQ_tR_t^\top \\ \hat{y}_{t|t-1} = Z_t\hat{\alpha}_{t|t-1} + d_t \\ v_t = y_t - \hat{y}_{t|t-1} \\ F_t = Z_tP_{t|t-1}Z_t^\top + H_t \\ \hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + P_{t|t-1}Z_t^\top F_t^{-1}v_t \\ P_{t|t} = \left(I_m - P_{t|t-1}Z_t^\top F_t^{-1}Z_t\right)P_{t|t-1} \end{cases}$$

where $\hat{y}_{t|t-1} = \mathbb{E}\left[y_t \mid \mathcal{F}_{t-1}\right]$ is the best estimator of y_t given the available information until time $t-1$, v_t is the innovation process and F_t is the associated covariance matrix.

Remark 126 *Harvey (1990) showed that we can directly calculate $\hat{\alpha}_{t+1|t}$ from $\hat{\alpha}_{t|t-1}$:*

$$\hat{\alpha}_{t+1|t} = \left(T_{t+1} - K_tZ_t\right)\hat{\alpha}_{t|t-1} + K_ty_t + c_{t+1} - K_td_t$$

where $K_t = T_{t+1}P_{t|t-1}Z_t^\top F_t^{-1}$ is the gain matrix. It follows that:

$$\hat{\alpha}_{t+1|t} = T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1} + K_t\left(y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right)$$

By recognizing the innovation process v_t, we obtain the following innovation representation:

$$\begin{cases} y_t = Z_t\hat{\alpha}_{t|t-1} + d_t + v_t \\ \hat{\alpha}_{t+1|t} = T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1} + K_tv_t \end{cases}$$

Kalman smoothing Let $t^\star \leq T$ be a date before the final observation of the sample. We note:

$$\hat{\alpha}_{t|t^\star} = \mathbb{E}\left[\alpha_t \mid \mathcal{F}_{t^\star}\right]$$

and:

$$P_{t|t^\star} = \mathbb{E}\left[\left(\hat{\alpha}_{t|t^\star}-\alpha_t\right)\left(\hat{\alpha}_{t|t^\star}-\alpha_t\right)^\top\right]$$

for all $t \leq t^\star$. By construction, $\hat{\alpha}_{t^\star|t^\star}$ and $P_{t^\star|t^\star}$ are exactly equal to the quantities calculated with the Kalman filter. We can show that the smoothed estimates for $t < t^\star$ are given by the Kalman smoother algorithm:

$$\begin{cases} S = P_{t|t}T_{t+1}^\top P_{t+1|t}^{-1} \\ \hat{\alpha}_{t|t^\star} = \hat{\alpha}_{t|t} + S\left(\hat{\alpha}_{t+1|t^\star} - \hat{\alpha}_{t+1|t}\right) \\ P_{t|t^\star} = P_{t|t} + S\left(P_{t+1|t^\star} - P_{t+1|t}\right)S^\top \end{cases}$$

While the Kalman filter is a forward algorithm[20], the Kalman smoother is a backward algorithm[21].

[19] The algorithm is initialized with values $\hat{\alpha}_{0|0} = \hat{\alpha}_0$ and $P_{0|0} = P_0$.

[20] The algorithm proceeds recursively, starting with values $\hat{\alpha}_0$ and P_0 at the starting date t_0, and moving forward in time toward the ending date.

[21] The algorithm proceeds recursively, starting with values $\hat{\alpha}_{t^\star|t^\star}$ and $P_{t^\star|t^\star}$ at the ending date $t^\star$, and moving backward in time toward the initial date.

Estimation of unknown parameters In many cases, the state space model depends on certain parameters that are unknown. Given a set θ of values for these unknown parameters, the Kalman filter may be applied to estimate the state vector α_t. We have:

$$v_t \sim \mathcal{N}(\mathbf{0}, F_t)$$

where $v_t = y_t - \hat{y}_{t|t-1}$ is the innovation at time t and $F_t = Z_t P_{t|t-1} Z_t^\top + H_t$ is the covariance matrix. If we change θ and we run the Kalman filter, we will obtain other values of v_t and F_t, meaning that v_t and F_t depend on θ. This is why we can write $v_t(\theta)$ and $F_t(\theta)$. We deduce that the likelihood function of the sample $\{y_1, \ldots, y_T\}$ is equal to:

$$\boldsymbol{\ell}(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\ln|F_t(\theta)| + v_t(\theta)^\top F_t(\theta)^{-1} v_t(\theta)\right)$$

We can then estimate the vector θ of unknown parameters by the method of maximum likelihood:

$$\hat{\theta} = \arg\max \boldsymbol{\ell}(\theta)$$

Once the ML estimate $\hat{\theta}$ is found, we can run again[22] the Kalman filter to estimate the other quantities $\hat{\alpha}_{t|t-1}$, $\hat{\alpha}_{t|t}$, $P_{t|t-1}$ and $P_{t|t}$.

Time-invariant state space model We consider the time-invariant model:

$$\begin{cases} y_t = Z\alpha_t + d + \epsilon_t \\ \alpha_t = T\alpha_{t-1} + c + R\eta_t \end{cases} \tag{10.34}$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, H)$ and $\eta_t \sim \mathcal{N}(\mathbf{0}, Q)$. If the state space model converges to a steady state, the estimators $(\hat{\alpha}_\infty, P_\infty)$ must satisfy the following equations:

$$\begin{cases} \hat{\alpha}_\infty = T\hat{\alpha}_\infty + c \\ P_\infty = TP_\infty T^\top + RQR^\top \end{cases}$$

It follows that the solution is:

$$\begin{cases} \hat{\alpha}_\infty = (I_m - T)^{-1} c \\ \text{vec}(P_\infty) = (I_{m^2} - T \otimes T)^{-1} \text{vec}\left(RQR^\top\right) \end{cases} \tag{10.35}$$

where $\hat{\alpha}_\infty$ and P_∞ are the unconditional mean and covariance matrix of α_t. Without any knowledge of the initial position α_0, the best way to define $\hat{\alpha}_0$ and P_0 is then to use the steady state:

$$\begin{cases} \hat{\alpha}_0 = \hat{\alpha}_\infty \\ P_0 = P_\infty \end{cases}$$

In many state space models, the matrices T, c, R and Q depend on unknown parameters θ, implying that $\hat{\alpha}_\infty$ and P_∞ also depend on θ. This means that when maximizing the log-likelihood function, the Kalman filter is initialized by values of $\hat{\alpha}_0$ and P_0 that depend on θ. This is the main difference with time-varying state space model since the Kalman filter is initialized by fixed values of $\hat{\alpha}_0$ and P_0.

We consider the AR(1) process $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. The state space form is:

$$\begin{cases} y_t = Z\alpha_t \\ \alpha_t = T\alpha_{t-1} + c + R\eta_t \end{cases}$$

[22] We say again, because computing the log-likelihood function requires one Kalman filter run, implying that many Kalman filter runs are used for maximizing the log-likelihood function.

where $\alpha_t = (y_t, \varepsilon_t)$, $Z = \begin{pmatrix} 0 & 1 \end{pmatrix}$, $T = \begin{pmatrix} \phi_1 & 0 \\ 0 & 0 \end{pmatrix}$, $c = \begin{pmatrix} \mu \\ 0 \end{pmatrix}$, $R = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $Q = \sigma_\varepsilon^2$. It follows that:

$$\begin{aligned} \hat{\alpha}_\infty &= (I_2 - T)^{-1} c \\ &= \begin{pmatrix} 1-\phi_1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\ &= \frac{1}{1-\phi_1} \begin{pmatrix} \mu \\ 0 \end{pmatrix} \end{aligned}$$

and:

$$\begin{aligned} \operatorname{vec}(P_\infty) &= (I_4 - T \otimes T)^{-1} \operatorname{vec}\left(RQR^\top\right) \\ &= \begin{pmatrix} 1-\phi_1^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix} \\ &= \begin{pmatrix} \left(1-\phi_1^2\right)^{-1} \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 \end{pmatrix} \end{aligned}$$

We deduce that the Kalman filter must be initialized with:

$$\hat{\alpha}_0 = \begin{pmatrix} \dfrac{\mu}{1-\phi_1} \\ 0 \end{pmatrix}$$

and:

$$P_0 = \sigma_\varepsilon^2 \begin{pmatrix} \dfrac{1}{1-\phi_1^2} & 1 \\ 1 & 1 \end{pmatrix}$$

More generally, for an ARMA(p,q) process, we calculate $\hat{\alpha}_0$ and P_0 by using Equation (10.35) with the SSM form given on page 647.

Since the SSM (10.34) can be viewed as a VaR(1) process, we can easily calculate the autocovariance matrices $\Gamma_y(k) = \operatorname{cov}(y_t, y_{t-k})$. We have $\Gamma_y(k) = ZT^k \Gamma_y(0) Z^\top$ where $\Gamma_y(0) = P_\infty$. We also deduce that the responses to forecast errors and orthogonal impulses are equal to $\Lambda_k = ZT^k R$ and $\Delta_k = ZT^k RP(Q)$ where $P(Q)$ is the Cholesky decomposition matrix of Q. It follows that the long-term multipliers are $\Psi_\infty = Z(I-T)^{-1} R$ and $\Xi_\infty = Z(I-T)^{-1} RP(Q)$.

10.2.2.2 Some applications

The recursive least squares We consider the linear model $y_t = x_t^\top \beta + u_t$ where x_t is a vector of K exogenous variables and $u_t \sim \mathcal{N}\left(0, \sigma^2\right)$. This model corresponds to the following state space model:

$$\begin{cases} y_t = x_t^\top \beta_t + u_t \\ \beta_t = \beta_{t-1} \end{cases}$$

where β_t is the state vector. Using the Kalman filter, we obtain $\hat{\beta}_{t|t-1} = \hat{\beta}_{t-1|t-1}$, $P_{t|t-1} = P_{t-1|t-1}$, $\hat{y}_{t|t-1} = x_t^\top \hat{\beta}_{t|t-1}$, $v_t = y_t - \hat{y}_{t|t-1}$ and $F_t = x_t^\top P_{t|t-1} x_t + \sigma^2$. The updating

equations are:

$$\begin{cases} \hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + P_{t|t-1} x_t F_t^{-1} v_t \\ P_{t|t} = \left(I_K - P_{t|t-1} x_t F_t^{-1} x_t^\top\right) P_{t|t-1} \end{cases}$$

We consider the $t \times 1$ vector $Y_t = (y_1, \ldots, y_t)$ and the $t \times K$ matrix $X_t = \left(x_1^\top, \ldots, x_t^\top\right)$. We have $X_t^\top X_t = X_{t-1}^\top X_{t-1} + x_t x_t^\top$ and $X_t^\top Y_t = X_{t-1}^\top Y_{t-1} + x_t y_t$. If we assume that $P_{t-1|t-1} = \sigma^2 \left(X_{t-1}^\top X_{t-1}\right)^{-1}$, we deduce that:

$$F_t = \sigma^2 \left(1 + x_t^\top \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t\right)$$

It follows that[23]:

$$\begin{aligned} P_{t|t} &= \left(I_K - P_{t|t-1} x_t F_t^{-1} x_t^\top\right) P_{t|t-1} \\ &= \sigma^2 \left(X_{t-1}^\top X_{t-1}\right)^{-1} - \frac{\sigma^2 \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t x_t^\top \left(X_{t-1}^\top X_{t-1}\right)^{-1} \sigma^2}{\sigma^2 \left(1 + x_t^\top \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t\right)} \\ &= \sigma^2 \left(\left(X_{t-1}^\top X_{t-1}\right)^{-1} - \frac{\lambda_t \lambda_t^\top}{1 + \lambda_t^\top x_t}\right) \\ &= \sigma^2 \left(X_t^\top X_t\right)^{-1} \end{aligned}$$

where $\lambda_t = \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t$. This proves that the assumption $P_t = \sigma^2 \left(X_t^\top X_t\right)^{-1}$ is true. Since we have:

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t \frac{v_t}{1 + \lambda_t^\top x_t}$$

the Kalman filter reduces to the following set of equations:

$$\begin{cases} v_t = y_t - x_t^\top \hat{\beta}_{t-1} \\ \lambda_t = \left(X_{t-1}^\top X_{t-1}\right)^{-1} x_t \\ F_t = \sigma^2 \left(1 + \lambda_t^\top x_t\right) \\ \hat{\beta}_t = \hat{\beta}_{t-1} + P_{t-1} x_t F_t^{-1} v_t \\ P_t = P_{t-1} - \left(1 + \lambda_t^\top x_t\right)^{-1} \lambda_t \lambda_t^\top \end{cases}$$

where $\hat{\beta}_t = \hat{\beta}_{t|t}$ and $P_t = P_{t|t}$. These equations define exactly the system of the recursive least squares (Spanos, 1986) and avoid the inversion of matrices $X_t^\top X_t$ to compute the RLS estimates $\hat{\beta}_t = \left(X_t^\top X_t\right)^{-1} X_t^\top Y_t$ for $t = 1, \ldots, T$.

At the terminal date, we have:

$$\hat{\beta}_T = \hat{\beta}_{T|T} = \hat{\beta}_{\text{OLS}}$$

Since we have $S = P_{t|t} T_{t+1}^\top P_{t+1|t}^{-1} = P_t P_{t+1|t}^{-1} = I_K$, the Kalman smoother gives:

$$\hat{\beta}_{t|T} = \hat{\beta}_{t|t} + \left(\hat{\beta}_{t+1|T} - \hat{\beta}_{t+1|t}\right) = \hat{\beta}_{t+1|T}$$

and:

$$P_{t|T} = P_{t|t} + \left(P_{t+1|T} - P_{t+1|t}\right) = P_{t+1|T}$$

[23] The last identity is obtained thanks to the Sherman-Morrison-Woodbury formula. We have:

$$\left(X_t^\top X_t\right)^{-1} = \left(X_{t-1}^\top X_{t-1}\right)^{-1} - \frac{\lambda_t \lambda_t^\top}{1 + \lambda_t^\top x_t}$$

We conclude that $\hat{\beta}_{t|T} = \hat{\beta}_{\text{OLS}} = \left(X_T^\top X_T\right)^{-1} X_T^\top Y_T$ and $P_{t|T} = \text{cov}\left(\hat{\beta}_{\text{OLS}}\right) = \sigma^2 \left(X_T^\top X_T\right)^{-1}$. These results are easy to understand because the best estimator given all the information is the OLS estimator applied to the full sample.

In order to illustrate the benefits of recursive least squares, we simulate the following model:

$$y_t = \begin{cases} 2 + t + 3x_t + u_t & \text{if } t \leq 100 \\ 10 + t + 3x_t + u_t & \text{if } t > 100 \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$. The exogenous variable x_t is simulated from the distribution function $\mathcal{U}_{[0,5]}$. In Figure 10.10, we report the RLS estimated values of the constant c_t of the linear model $y_t = c_t + \beta_t x_t + u_t$. We observe a behavioral change of the estimate when $t > 100$. If we consider the evolution of the innovation process v_t and the corresponding 99% confidence interval, we verify a structural change in the model. We generally identify trend breaks by using CUSUM and CUSUMSQ tests (Brown *et al.*, 1975). Let $w_t = \left(1 + \lambda_t^\top x_t\right)^{-1/2} v_t$ be the normalized innovation. Under the assumption $\mathcal{H}_0 : \beta_t = \beta_{t-1}$, the CUSUM statistic defined by $W_t = s^{-1} \sum_{i=1}^t w_i$ where $s^2 = (T-K)^{-1} \sum_{t=1}^T \left(y_t - x_t^\top \hat{\beta}_t\right)^2$ follows the distribution function $\mathcal{N}(0, t-K)$. The CUSUMSQ statistic corresponds to $V_t = \sum_{i=1}^t w_i^2 \Big/ \sum_{i=1}^T w_i^2$ and has a beta distribution $\mathcal{B}((T-t)/2, (t-K)/2)$ under $\mathcal{H}_0$.

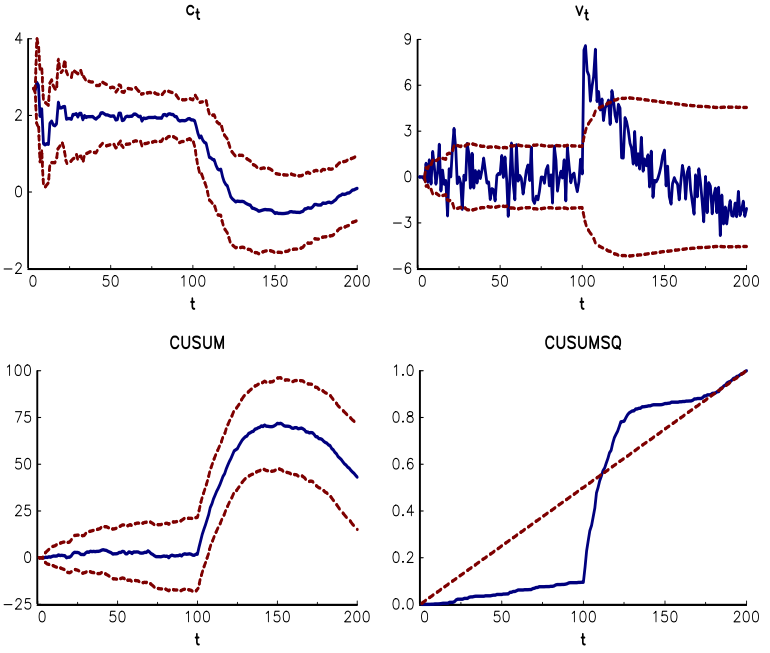


FIGURE 10.10: CUMSUM test and recursive least squares

Structural time series models With the Kalman filter, we can estimate unobserved components. Let us consider the deterministic trend model:

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \beta \cdot t \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. Estimating the trend μ_t is equivalent to estimate the parameter β by ordinary least squares and set $\hat{\mu}_t = \hat{\beta} \cdot t$. We notice that the previous model can be written as:

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta \end{cases}$$

A way to introduce a stochastic trend is to add a noise η_t in the trend equation:

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta + \eta_t \end{cases}$$

where $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$. This model is called the local level (LL) model. Using the SSM notations, we have the following correspondence: $Z_t = 1$, $d_t = 0$, $H_t = \sigma_\varepsilon^2$, $\alpha_t = \mu_t$, $T_t = 1$, $c_t = \beta$ and $Q_t = \sigma_\eta^2$. Once the parameters β, σ_ε and σ_η are estimated by the method of maximum likelihood, we estimate $\hat{\mu}_{t|t}$ and $\hat{\mu}_{t|t-1}$ by using the Kalman filter. Let us now assume that the slope of the trend is also stochastic:

$$\begin{cases} \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

where $\zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$. We obtain the local linear trend (LLT) model. The corresponding SSM form is:

$$\begin{cases} y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \varepsilon_t \\ \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix} \end{cases}$$

where $H_t = \sigma_\varepsilon^2$ and $Q_t = \text{diag}\left(\sigma_\eta^2, \sigma_\zeta^2\right)$. With the Kalman filter, we both estimate the stochastic trend μ_t and the stochastic slope β_t. We now consider that two ARMA(1,1) time series $y_{1,t}$ and $y_{2,t}$ have a common component c_t:

$$\begin{cases} y_{1,t} = \phi_1 y_{1,t-1} + \beta_1 c_{t-1} + \varepsilon_{1,t} - \theta_1 \varepsilon_{1,t-1} \\ y_{2,t} = \phi_2 y_{2,t-1} + \beta_2 c_{t-1} + \varepsilon_{2,t} - \theta_2 \varepsilon_{2,t-1} \\ c_t = c_{t-1} + \varepsilon_{3,t} \end{cases}$$

We obtain the following state space model:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ c_t \\ \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

where:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ c_t \\ \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_1 & 0 & \beta_1 & -\theta_1 & 0 \\ 0 & \phi_2 & \beta_2 & 0 & -\theta_2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ c_{t-1} \\ \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}$$

We notice that the state vector contains two observed variables ($y_{1,t}$ and $y_{2,t}$) and three unobserved variables (c_t, $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$).

Remark 127 *These different models have been popularized by Harvey (1990) and are known as structural time series model (STSM). The reader will find additional materials on these models on page 679.*

Time-varying parameters If we assume that the beta coefficients are time-varying, the linear regression model has the following SSM form:

$$\begin{cases} y_t = x_t^\top \beta_t + u_t \\ \beta_t = \beta_{t-1} + \eta_t \end{cases}$$

where $u_t \sim \mathcal{N}\left(0, \sigma_u^2\right)$ and $\eta_t \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$. When the initial position is $\beta_0 \sim \mathcal{N}\left(\hat{\beta}_0, P_0\right)$, the Kalman filter equations are $\hat{\beta}_{t|t-1} = \hat{\beta}_{t-1|t-1}$, $P_{t|t-1} = P_{t-1|t-1} + \Sigma$, $v_t = y_t - x_t^\top \hat{\beta}_{t|t-1}$, $F_t = x_t^\top P_{t|t-1} x_t + \sigma_u^2$, $\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + \left(\frac{P_{t|t-1}}{F_t}\right) x_t v_t$ and $P_{t|t} = \left(I_m - \left(\frac{P_{t|t-1}}{F_t}\right) x_t x_t^\top\right) P_{t|t-1}$. Let us now multiply the parameters σ_u^2, Σ and P_0 by the scalar λ. The Kalman filter becomes: $\hat{\beta}'_{t|t-1} = \hat{\beta}'_{t-1|t-1}$, $P'_{t|t-1} = P'_{t-1|t-1} + \lambda\Sigma$, $v'_t = y_t - x_t^\top \hat{\beta}'_{t|t-1}$, $F'_t = x_t^\top P'_{t|t-1} x_t + \lambda\sigma_u^2$, $\hat{\beta}'_{t|t} = \hat{\beta}'_{t|t-1} + \left(\frac{P'_{t|t-1}}{F'_t}\right) x_t v'_t$, $P'_{t|t} = \left(I_m - \left(\frac{P'_{t|t-1}}{F'_t}\right) x_t x_t^\top\right) P'_{t|t-1}$. We obtain the following matching: $\hat{\beta}'_{t|t-1} = \hat{\beta}_{t|t-1}$, $P'_{t|t-1} = \lambda P_{t|t-1}$, $v'_t = v_t$, $F'_t = \lambda F_t$, $\hat{\beta}'_{t|t} = \hat{\beta}_{t|t}$ and $P'_{t|t} = \lambda P_{t|t}$. This means that the scaling parameter λ has no impact on the state estimates $\hat{\beta}_{t|t-1}$ and $\hat{\beta}_{t|t}$.

In this model, the parameters σ_u^2 and Σ are unknown and can be estimated by the method of maximum likelihood. The log-likelihood function is equal to:

$$\boldsymbol{\ell}(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\ln F_t + \frac{v_t^2}{F_t}\right)$$

where $\theta = \left(\sigma^2, \Sigma\right)$. Maximizing the log-likelihood function requires specifying the initial conditions $\hat{\beta}_0$ and P_0, which are not necessarily known. It is a bad idea to consider that $\hat{\beta}_0$ and P_0 are also unknown parameters and to perform the maximization with respect to $\theta = \left(\sigma^2, \Sigma, \hat{\beta}_0, P_0\right)$. Generally, this approach does not converge. It is better to try different initial conditions and to test their impact on the ML estimates.

Let us consider an illustration provided by Roncalli and Weisang (2009). They suppose that a fund manager allocates between the MSCI USA index and the MSCI EMU index. The monthly performance R_t of his portfolio is equal to:

$$R_t = w_t^{(\text{USA})} \cdot R_t^{(\text{USA})} + \left(1 - w_t^{(\text{USA})}\right) \cdot R_t^{(\text{EMU})}$$

where $R_t^{(\text{USA})}$ and $R_t^{(\text{EMU})}$ are the monthly returns of MSCI USA and EMU indices, and $w_t^{(\text{USA})}$ and $w_t^{(\text{EMU})} = 1 - w_t^{(\text{USA})}$ are the corresponding monthly allocations. In Figure 10.11, we report an example of the dynamic allocation and the cumulative performance of the portfolio. The investor generally does not know the allocation process $\left(w_t^{(\text{USA})}, w_t^{(\text{EMU})}\right)$ since he only observes the performance of the fund and the different components. Estimating the implied exposures is known as a tracking problem. In our example, we can specify the

following state space model:

$$\begin{cases} R_t = \begin{pmatrix} R_t^{(\text{USA})} & R_t^{(\text{EMU})} \end{pmatrix} \begin{pmatrix} \beta_t^{(\text{USA})} \\ \beta_t^{(\text{EMU})} \end{pmatrix} + u_t \\ \begin{pmatrix} \beta_t^{(\text{USA})} \\ \beta_t^{(\text{EMU})} \end{pmatrix} = \begin{pmatrix} \beta_{t-1}^{(\text{USA})} \\ \beta_{t-1}^{(\text{EMU})} \end{pmatrix} + \eta_t \end{cases}$$

where $u_t \sim \mathcal{N}\left(0, \sigma_u^2\right)$ and $\eta_t \sim \mathcal{N}\left(\mathbf{0}, \Sigma\right)$. Since we have simulated the allocation model in Figure 10.11, we can estimate the parameters σ_u^2 and Σ. By construction, $\hat{\sigma}_u^2$ is equal to zero and $\hat{\Sigma}$ is the empirical covariance matrix[24] between $w_t^{(\text{USA})} - w_{t-1}^{(\text{USA})}$ and $w_t^{(\text{EMU})} - w_{t-1}^{(\text{EMU})}$. In the first panel in Figure 10.12, we have reported the allocation $\hat{\beta}_t^{(\text{USA})}$ estimated by the Kalman filter[25]. Generally, Σ is specified as a diagonal matrix implying that the parameter changes are independent:

$$\mathbb{E}\left[\left(\beta_{j,t} - \beta_{j,t-1}\right)\left(\beta_{k,t} - \beta_{k,t-1}\right)\right] = 0 \qquad \text{if } j \neq k$$

This is the standard approach when specifying a time-varying parameter model. In this case, we obtain the second panel. The results are less precise, because we don't take into account the negative correlation between $w_t^{(\text{USA})} - w_{t-1}^{(\text{USA})}$ and $w_t^{(\text{EMU})} - w_{t-1}^{(\text{EMU})}$. In the third panel, we estimate σ_u^2 and Σ by the method of maximum likelihood. Again, the results are less robust because we specify a diagonal matrix for Σ. In the last panel, we consider a variant of the previous model:

$$\begin{cases} R_t - R_t^{(\text{EMU})} = \left(R_t^{(\text{USA})} - R_t^{(\text{EMU})}\right) \beta_t^{(\text{USA})} + u_t \\ \beta_t^{(\text{USA})} = \beta_{t-1}^{(\text{USA})} + \eta_t \end{cases}$$

The idea is to transform the set of correlated exogenous variables into a set of quasi-independent exogenous variables. This technique is frequently used in financial tracking problems. In this case, the results are very good (see the fourth panel in Figure 10.12).

10.2.3 Cointegration and error correction models

10.2.3.1 Nonstationarity and spurious regression

Let us consider a random walk process:

$$y_t = y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. y_t is a nonstationary AR(1) process. Since the lag polynomial of y_t is equal to $A(L) = 1 - L$, we deduce that the root associated to the equation $A(z) = 0$ is equal to one, and we say that y_t has a unit root. We also note $y_t \sim I(1)$.

Let us now consider two time series (x_t, y_t) generated by the following bivariate process:

$$\begin{cases} x_t = x_{t-1} + \eta_t \\ y_t = y_{t-1} + \varepsilon_t \end{cases}$$

[24] We obtain:

$$\hat{\Sigma} = \begin{pmatrix} 57.9848 & -57.9848 \\ -57.9848 & 57.9848 \end{pmatrix} \times 10^{-4}$$

[25] We assume that the initial conditions are $\hat{\beta}_0 = (50\%, 50\%)$ and $P_0 = \mathbf{0}$.

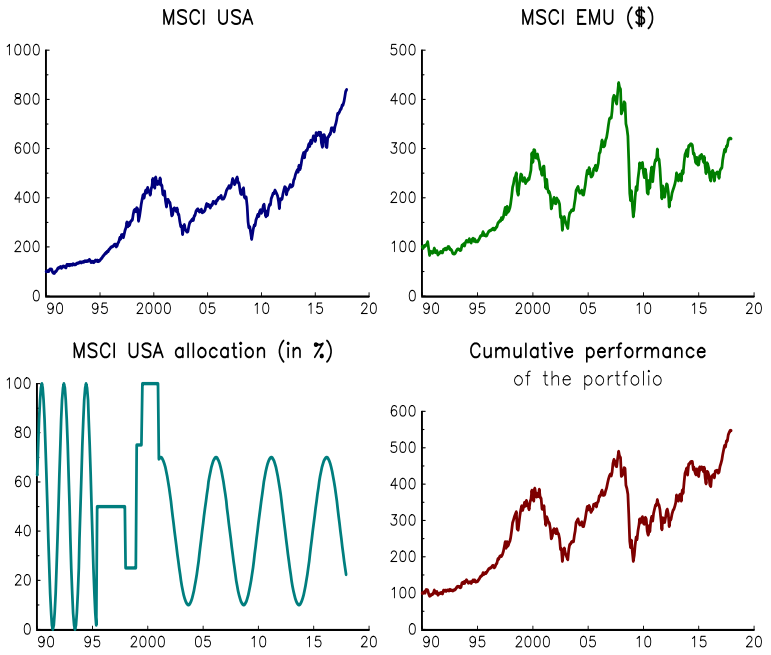


FIGURE 10.11: The tracking problem

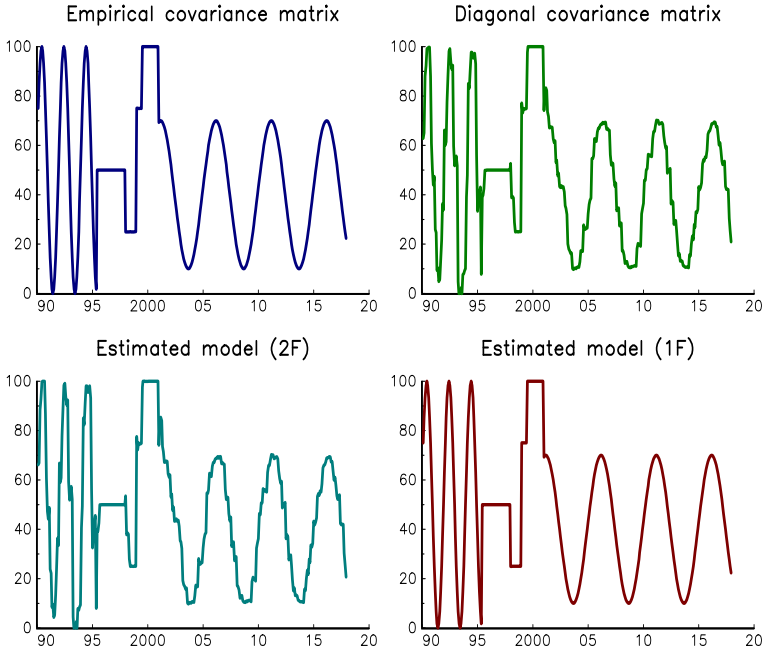


FIGURE 10.12: Estimation of the dynamic allocation by Kalman filtering

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ and $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ are two independent white noise processes. We are interested in the linear regression:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

By construction, we know that β_1 is equal to zero because x_t and y_t are two independent processes. We can also consider the alternative linear model:

$$y_t - y_{t-1} = \beta_0 + \beta_1 \left(x_t - x_{t-1}\right) + u_t$$

In Figure 10.13, we have reported the probability density function[26] of the OLS estimator $\hat{\beta}_1$. We notice that the linear regression using the level or the first-difference does not lead to the same results. For instance, there is a 20% probability that $\hat{\beta}_1$ exceeds 0.5 when using variables in levels. In these cases, we can conclude that y_t is related to x_t, but this relationship is spurious (Granger and Newbold, 1974). More generally, we can have the impression that there is a strong relationship between x_t and y_t because they both exhibit a trend even if they are not correlated. The spurious regression phenomenon particularly occurs when the processes are not stationary.

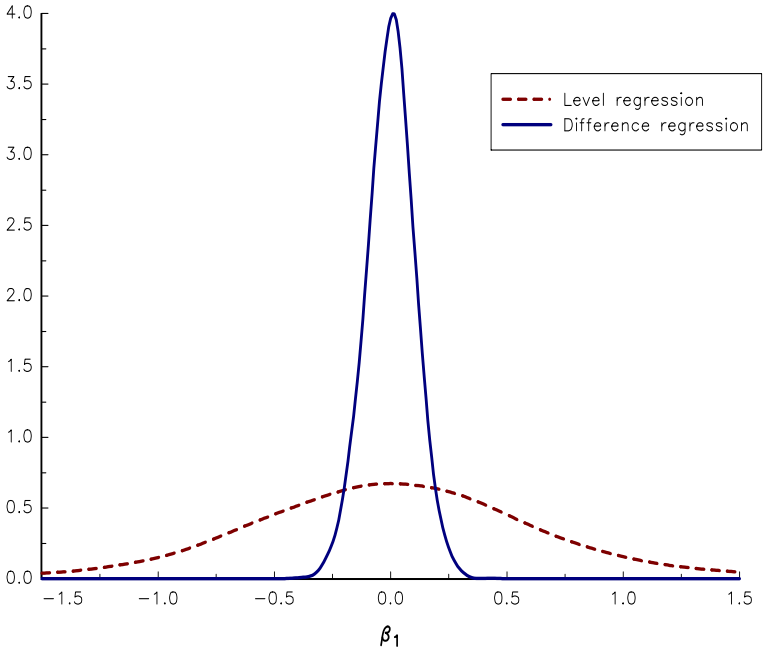


FIGURE 10.13: Probability density function of $\hat{\beta}_1$ in the case of a spurious regression

10.2.3.2 The concept of cointegration

Nonstationary processes are not limited to the random walk without drift. For example, the process can incorporate a deterministic or a stochastic trend. If we consider the previous process $y_t = y_{t-1} + \varepsilon_t$, the first difference $y_t - y_{t-1}$ is stationary and we will say that $y_t - y_{t-1}$

[26] We simulate (x_t, y_t) for $t = 1, \ldots, 100$ by assuming that $x_0 = y_0 = 100$, and $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$. For each simulation, we compute the OLS estimate for the two models. We run 5 000 Monte Carlo replications and estimate the probability density function of $\hat{\beta}_1$ by using the non-parametric kernel approach.

is integrated of order zero: $y_t - y_{t-1} \sim I(0)$. More generally, a process y_t is called integrated of order one if $y_t - y_{t-1}$ is stationary. Therefore, we have the following property:

$$y_t \sim I(1) \Rightarrow y_t - y_{t-1} \sim I(0)$$

We can extend the concept of integration to an order $d > 1$:

$$y_t \sim I(d) \Rightarrow (1-L)^d y_t \sim I(0)$$

For instance, y_t is integrated of order two if $y_t - y_{t-1}$ is integrated of order one and $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ is integrated of order zero.

A n-dimensional time series $y_t = (y_{1,t}, \ldots, y_{n,t})$ is said to be cointegrated of order (d_1, d_2) and we note $y_t \sim CI(d_1, d_2)$ if each component is $I(d_1)$ and there are linear combinations[27] $\beta^\top y_t$ that are integrated of an order d_2 where $d_2 < d_1$. In what follows, we restrict the analysis to $d_1 = 1$ and $d_2 = 0$, which is the most frequent case in econometrics:

$$\begin{cases} y_t \sim I(1) \\ \beta^\top y_t \sim I(0) \end{cases} \Rightarrow y_t \sim CI(1,0)$$

From an economic point of view, this implies that $\beta^\top y_t$ forms a long-run equilibrium: the variables y_t are nonstationary, but are related because a linear combination is stationary. An example is given in Figure 10.14. The three variables $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ are integrated of order one and exhibit some trends. However, the linear combination $z_t = 2y_{1,t} - y_{2,t} - y_{3,t}$ is stationary and moves around its mean, which is equal to 10. Therefore, it is certainly difficult to predict the three univariate time series, because they are nonstationary. It is easier to forecast the combination z_t, because it is stationary and returns to its mean in the long-run.

Let us consider the bivariate process (x_t, y_t) where $x_t \sim I(1)$ and $y_t \sim I(1)$. The linear regression $y_t = \beta_0 + \beta_1 x_t + u_t$ implies that $z_t = y_t - \beta_1 x_t = \beta_0 + u_t$ can be integrated of order zero or one. If $z_t \sim I(1)$, there is no long-run relationship between x_t and y_t because z_t is nonstationary. The knowledge of the process x_t does not help to understand where the process y_t is located. On the contrary, if $z_t \sim I(0)$, there is a long-run stationary relationship between x_t and y_t. On the long-run, shocks to x_t and y_t have permanent effects, but z_t can only deviate far from its mean in the short-run. z_t is also called the equilibrium error. In economics, a famous example is the theory of the purchasing power parity (PPP). According to the law of one price, commodity prices must be the same in two different countries:

$$P_i = SP_i^\star$$

where P_i and $P_i^\star$ are the price of commodity i in the home and foreign economy, and S is the nominal exchange rate. Purchasing power parity is the application of the law of one price to a large basket of goods:

$$P_t = S_t P_t^\star$$

where P_t and $P_t^\star$ are the domestic and foreign prices of the basket, and S_t is the nominal exchange rate at time t. By taking the logarithm, we obtain:

$$p_t = s_t + p_t^\star$$

where $p_t = \ln P_t$, $p_t^\star = \ln P_t^\star$ and $s_t = \ln S_t$. In practice, market frictions, transportation costs, taxes, etc. explain that short-run deviations from PPP are large, but exchange rates tend toward PPP in the long run. To test this theory, econometricians demonstrate that $p_t \sim I(1)$, $s_t \sim I(1)$, and $p_t^\star \sim I(1)$, but $(p_t, s_t, p_t^\star)$ is cointegrated.

[27] We have $\beta \neq \mathbf{0}$.

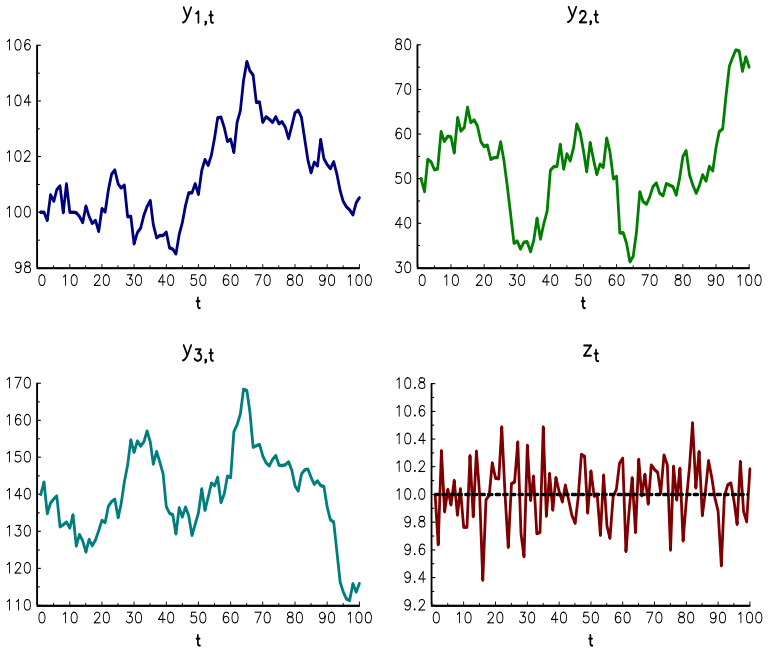


FIGURE 10.14: Illustration of the cointegration

Remark 128 *If y_t has $n > 2$ components, we can have several cointegration relationships. If there are r independent cointegrated vectors with $r \leq n-1$, y_t is said to be cointegrated of order r. β is then a matrix of dimension $n \times r$, whose columns form a basis for the space of cointegrating vectors. We consider the following process:*

$$\begin{cases} y_{1,t} = y_{2,t} - 0.5 \cdot y_{3,t} + \varepsilon_{1,t} \\ y_{2,t} = 0.4 \cdot y_{3,t} + \varepsilon_{2,t} \\ y_{3,t} = y_{3,t-1} + \varepsilon_{3,t} \end{cases}$$

where $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ are three independent white noise processes. It is obvious that $y_{1,t} \sim I(1)$, $y_{2,t} \sim I(1)$, and $y_{3,t} \sim I(1)$. Moreover, we have:

$$z_{1,t} = y_{1,t} - y_{2,t} + 0.5 \cdot y_{3,t} \sim I(0)$$

and:

$$z_{2,t} = y_{2,t} - 0.4 \cdot y_{3,t} \sim I(0)$$

The rank r is then equal to 2 and we have:

$$\beta = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0.5 & -0.4 \end{pmatrix}$$

10.2.3.3 Error correction model

We consider the bivariate example:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \mu + b_1 x_t + b_2 x_{t-1} + \varepsilon_t$$

where y_t is an AR(2) process that depends on the lags of x_t. We deduce that:

$$\begin{aligned}
\Delta y_t &= (a_1 - 1)\, y_{t-1} + a_2 y_{t-2} + \mu + b_1 x_t + b_2 x_{t-1} + \varepsilon_t \\
&= (a_1 + a_2 - 1)\, y_{t-1} - a_2 \Delta y_{t-1} + \mu + b_1 \Delta x_t + (b_1 + b_2)\, x_{t-1} + \varepsilon_t \\
&= -a_2 \Delta y_{t-1} + \mu + b_1 \Delta x_t + \\
&\quad (a_1 + a_2 - 1) \left(y_{t-1} - \frac{(b_1 + b_2)}{1 - a_1 - a_2} x_{t-1} \right) + \varepsilon_t \\
&= -a_2 \Delta y_{t-1} + \mu + b_1 \Delta x_t + \alpha z_{t-1} + \varepsilon_t
\end{aligned}$$

where $\alpha = a_1 + a_2 - 1$ and $z_{t-1} = y_{t-1} + \alpha^{-1}(b_1 + b_2)\, x_{t-1}$. We notice that the short-run dynamics of y_t depends on the long-run equilibrium z_t. More generally, if $A'(L) y_t = \mu + B'(L)\, x_t + \varepsilon_t$, Engle and Granger (1987) show that:

$$A(L)\Delta y_t = \mu + B(L)\, \Delta x_t + \alpha\, (y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + \varepsilon_t$$

where $\alpha < 0$. This model incorporates an error correction mechanism (Salmon, 1982):

- if $z_{t-1} > 0$, then y_{t-1} is greater than its long-run target $\beta_0 + \beta_1 x_{t-1}$, which implies that the error correction is negative ($\alpha z_{t-1} < 0$) and y_t tends to decrease;
- if $z_{t-1} < 0$, then y_{t-1} is less than its long-run target $\beta_0 + \beta_1 x_{t-1}$, which implies that the error correction is positive ($\alpha z_{t-1} > 0$) and y_t tends to increase.

The Engle-Granger model assumes that the variable y_t is endogenous and the other variable x_t is exogenous. This is a restricted hypothesis. In the general n-dimensional case, we have $y_t \sim I(1)$ and $\Delta y_t \sim I(0)$. The vector error correction model (VECM) defines the short-run dynamics of y_t and corresponds to a VAR process that incorporates an error correction mechanism[28]:

$$\Phi(L)\, \Delta y_t = \mu_t + \alpha z_{t-1} + \varepsilon_t \tag{10.36}$$

where $z_t = \beta^\top y_t$ is the long-run equilibrium.

10.2.3.4 Estimation of cointegration relationships

The first step before estimating cointegration relationships is to verify that all the components of y_t are integrated of order one. The second step consists in testing the existence of cointegration relationships between the different components of y_t. In few cases, the cointegration vector β is known and given by the economic theory. If we reject the assumption that the residuals $z_t = \beta^\top y_t$ are integrated of order one, we can then conclude that the process y_t is cointegrated and β is a cointegration vector. In other cases, we have to estimate the cointegration relationships using the least squares approach or the method of maximum likelihood.

[28] Suppose that y_t is a VAR(p) process:

$$\Phi'(L)\, y_t = \mu_t + \varepsilon_t$$

If y_t has a unit root, we have $\Phi'(L) = \Phi'(1) + (1 - L)\, \Phi(L)$ and:

$$\Phi(L)\, \Delta y_t = \mu_t - \Phi'(1)\, y_{t-1} + \varepsilon_t$$

We deduce that $\alpha z_{t-1} = -\Phi'(1)\, y_{t-1}$ or $\alpha \beta^\top = -\Phi'(1)$. This result is known as the *Granger representation theorem*.

Unit root tests The most popular test is the augmented Dickey-Fuller test (ADF) proposed by Dickey and Fuller (1979, 1981). It consists in estimating the following regression model:

$$\Delta y_t = c_t + \phi y_{t-1} + \sum_{k=1}^{p} \phi_k \Delta y_{t-k} + \varepsilon_t \tag{10.37}$$

where ε_t is a white noise process. The original Dickey-Fuller test (DF) corresponds to the case $p = 0$ whereas the ADF test includes an autoregressive component ($p \geq 1$). Generally, we consider three specifications of Equation (10.37):

1. the linear regression does not include the term c_t;
2. c_t is a constant: $c_t = \mu$;
3. c_t is a constant plus a deterministic trend: $c_t = \mu + \lambda t$.

The hypotheses are $\mathcal{H}_0 : \phi = 0$ ($y_t \sim I(1)$) and $\mathcal{H}_1 : \phi < 0$ ($y_t \sim I(0)$). It is important to test the three models and to specify the null and alternative hypotheses appropriately. For instance, the third specification means that the null hypothesis is $y_t \sim I(1)$ and the alternative hypothesis is $y_t \sim I(0)$ with a deterministic trend. The ADF test is a Student's t test:

$$\text{ADF} = t_\phi = \frac{\hat{\phi}}{\sigma\left(\hat{\phi}\right)}$$

Depending on the specification, the critical values are noted τ (no c_t), τ_μ ($c_t = \mu$) and τ_τ ($c_t = \mu + \lambda t$). In Table 10.9, we report the Monte Carlo critical values obtained by David Dickey in his Ph.D. dissertation. We notice that they are different than the values obtained for the standard t statistic. Today, critical values are calculated using the response surface estimation approach proposed by MacKinnon (1996). This approach is very fast, very accurate for small sample size and implemented in most econometric softwares.

TABLE 10.9: Critical values of the ADF test

Significance level	10%	5%	1%
τ	−1.62	−1.94	−2.56
τ_μ	−2.57	−2.86	−3.43
τ_τ	−3.13	−3.41	−3.96

Remark 129 *The unit root tests above are only valid if ε_t is white noise. The purpose of the ADF test ($p \geq 1$) is then to correct the DF test ($p = 0$) when the residuals are correlated and heteroscedastic. An alternative to the ADF test is to consider the Phillips-Perron test (PP), which is based on the linear regression associated to the Dickey-Fuller test:*

$$\Delta y_t = c_t + \phi y_{t-1} + \varepsilon_t \tag{10.38}$$

Phillips and Perron (1988) propose a modified statistic Z_t of the Student's t statistic that takes into account the Newey-West long-run standard error of ε_t.

While ADF and PP unit root tests consider that the null hypothesis is $y_t \sim I(1)$, the KPSS test supposes that the null hypothesis is $y_t \sim I(0)$. This stationary test is based on the state space model:

$$\begin{cases} y_t = c_t + \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \omega_t \end{cases}$$

where $\omega_t \sim \mathcal{N}\left(0, \sigma_\omega^2\right)$. In this case, the stationary hypothesis is $\mathcal{H}_0 : \sigma_\omega^2 = 0$. The KPSS test is given by:

$$\text{KPSS} = T^{-2} \frac{\sum_{t=1}^{T} S_t^2}{s^2(l)}$$

where $S_t = \sum_{j=1}^{t} \hat{\varepsilon}_t$, $s^2(l)$ is the Newey-West long-run variance of ε_t with lag number l and $\hat{\varepsilon}_t$ is the residual $y_t - \hat{c}_t$. As for the ADF test, Kwiatkowski *et al.* (1992) propose two tests η_μ and η_τ depending on the specification of c_t. The critical values of these tests are reported in Table 10.10.

TABLE 10.10: Critical values of the KPSS test

Significance level	10%	5%	1%
η_μ	0.347	0.463	0.739
η_τ	0.119	0.146	0.216

Least squares estimation If the n-dimensional process y_t is cointegrated, there exists a non-zero vector $\beta = (\beta_1, \ldots, \beta_n)$ such that $\beta^\top y_t$ is stationary. Without loss of generality, we can assume that β_1 is not equal to zero, implying that β/β_1 is also a cointegration vector. Therefore, we can estimate the cointegration relationship with the following linear regression:

$$y_{1,t} = c_t + \beta_2 y_{2,t} + \cdots + \beta_n y_{n,t} + u_t \tag{10.39}$$

and verifying the integration order of the residuals. If $u_t \sim I(1)$, then y_t is not cointegrated. In the other cases, $u_t \sim I(0)$ and the normalized cointegration vector is $\hat{\beta} = \left(1, -\hat{\beta}_2, \ldots, -\hat{\beta}_n\right)$. We can then estimate the associated ECM:

$$\Phi_{(1)}(L)\,\Delta y_{1,t} = \mu_t + \sum_{j=2}^{n} \Phi_{(j)}(L)\,\Delta y_{j,t} + \alpha \hat{z}_{t-1} + \varepsilon_t$$

where $\hat{z}_t = \hat{\beta}^\top y_t$. As said previously, this two-step approach (also called the Engle-Granger method) has two main limitations. First, it implicitly assumes that $y_{1,t}$ is endogenous and $(y_{2,t}, \ldots, y_{n,t})$ are exogenous. Second, it is not valid if there are multiple cointegration relationships.

Maximum likelihood estimation If y_t is a VAR(p) process:

$$y_t = \mu_t + \sum_{i=1}^{p} \Phi_i' y_{t-i} + \varepsilon_t$$

the associated VECM is:

$$\Delta y_t = \mu_t + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \varepsilon_t \tag{10.40}$$

where $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$, μ_t is the vector component containing deterministic terms[29] (constant and/or trends) and:

$$\underset{(n\times n)}{\Pi} = \underset{(n\times r)}{\alpha}\ \underset{(r\times n)}{\beta^\top} \tag{10.41}$$

[29] We have $\mu_t = \mu_0$ or $\mu_t = \mu_0 + \mu_1 \cdot t$ where μ_0 and μ_1 are $n \times 1$ vectors.

Since $0 < r = \operatorname{rank}\Pi < n$, y_t is $I(1)$ with r linearly independent cointegrating relationships and $n-r$ common stochastic trends. Given a sample $\{y_1, \ldots, y_T\}$, the log-likelihood function is:

$$\ell(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^\top \Sigma^{-1}\varepsilon_t \tag{10.42}$$

The ML estimate $\hat{\theta} = \left(\hat{\mu}_0, \hat{\mu}_1, \hat{\Phi}_1, \ldots, \hat{\Phi}_{p-1}, \hat{\Sigma}, \hat{\Pi}\right)$ is obtained by maximizing the objective function (10.42) under the constraints (10.41).

Johansen (1988, 1991) proposes two tests to determine the rank of Π. These tests are based on the eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ of $\hat{\Pi}$. Let $\mathcal{H}_0$ be the null hypothesis that there are r cointegration relationships ($\operatorname{rank}\Pi = r$) and $\mathcal{H}_1$ be the alternative hypothesis that there are more than r cointegration relationships ($\operatorname{rank}\Pi > r$). The trace test of Johansen is defined by the likelihood ratio test:

$$\begin{aligned}
\mathrm{LR}_{\mathrm{trace}}(r) &= -2\ln\Lambda \\
&= -2\ln\frac{\sup_{\mathcal{H}_0}\ell(\theta)}{\sup\ell(\theta)} \\
&= -T\sum_{i=r+1}^{n}\ln\left(1-\hat{\lambda}_i\right)
\end{aligned}$$

where $\hat{\lambda}_i$ is the i^{th} eigenvalue of $\hat{\Pi}$. The underlying idea is the following: if $\operatorname{rank}\Pi = r$, then $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_n$ should be close to zero and the trace test has a small value. In contrast, if $\operatorname{rank}\Pi > r$, the likelihood ratio should be large. Like the ADF tests, the likelihood ratio has not a standard chi-squared distribution and critical values must be calculated by Monte Carlo simulations. If we would like to test the existence of r versus $r+1$ cointegration relationships, Johansen considers the maximum eigenvalue test:

$$\begin{aligned}
\mathrm{LR}_{\max}(r) &= -2\ln\Lambda \\
&= -2\ln\frac{\sup_{\mathcal{H}_0}\ell(\theta)}{\sup_{\mathcal{H}_1}\ell(\theta)} \\
&= -T\ln\left(1-\hat{\lambda}_{r+1}\right)
\end{aligned}$$

Again, Johansen (1988) provides critical values calculated by Monte Carlo simulations.

Remark 130 *There are different ways to estimate the eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ of $\hat{\Pi}$ (Johansen, 1988, 1991). It is interesting to notice that they are equal to the square of partial correlations $\hat{\rho}_1^2, \ldots, \hat{\rho}_n^2$ between Δy_t and y_{t-1} conditionally to Δy_{t-i} for $i = 1, \ldots, p-1$ such that $\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq \cdots \geq \hat{\rho}_n^2$. Indeed, if we consider the two linear regressions:*

$$\Delta y_t = a_0 + \sum_{i=1}^{p-1} A_i \Delta y_{t-i} + \varepsilon_{0,t}$$

and:

$$y_{t-1} = b_0 + \sum_{i=1}^{p-1} B_i \Delta y_{t-i} + \varepsilon_{1,t}$$

we deduce from the Frish-Waugh theorem that Equation (10.40) can be written as:

$$\varepsilon_{0,t} = \Pi\varepsilon_{1,t} + \varepsilon_t$$

Let $S_{0,0} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{0,t}\hat{\varepsilon}_{0,t}^\top$, $S_{0,1} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{1,t}\hat{\varepsilon}_{1,t}^\top$, and $S_{1,1} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{1,t}\hat{\varepsilon}_{1,t}^\top$ be the sample covariance matrices of $\operatorname{var}(\varepsilon_{0,t})$, $\operatorname{cov}(\varepsilon_{0,t}, \varepsilon_{1,t})$ and $\operatorname{var}(\varepsilon_{1,t})$. Johansen (1988) shows that the eigenvalues are the solutions of the matrix equation $\left|\lambda S_{1,1} - S_{0,1}^\top S_{0,0}^{-1} S_{0,1}\right| = \mathbf{0}$.

10.2.4 GARCH and stochastic volatility models

One of the main assumptions when estimating the time series model $y_t = x_t^\top \beta + \varepsilon_t$ is that the residuals have a constant variance. However, this hypothesis is not verified when we consider economic and financial time series. For example, Figure 10.15 represents the monthly returns R_t of the S&P 500 index between 1950 and 2017, and the square R_t^2 of these returns. It is obvious that the volatility is not constant and homogenous during this period.

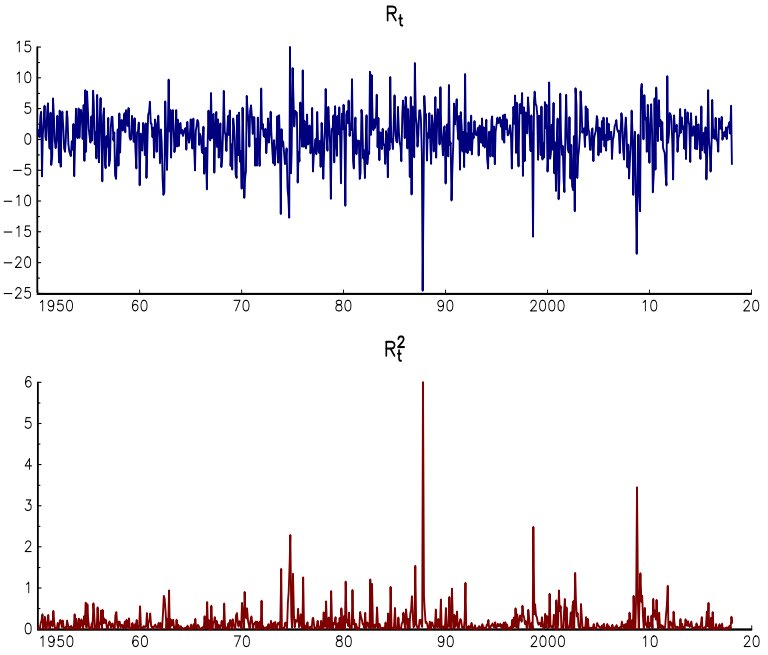


FIGURE 10.15: Monthly returns of the S&P 500 index (in %)

10.2.4.1 GARCH models

Definition We consider the linear regression model:

$$y_t = x_t^\top \beta + \varepsilon_t$$

where $\varepsilon_t = \sigma_t e_t$ and e_t is a centered standardized random variable $\mathcal{N}(0,1)$, which is independent from the past values ε_{t-i}. Moreover, we have:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 \tag{10.43}$$

where $\alpha_i \geq 0$ for all $i \geq 0$. We notice that the process ε_t is not autocorrelated and the conditional variance $\text{var}_{t-1}(\varepsilon_t) = \mathbb{E}\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right]$ is equal to σ_t^2. Therefore, this variance depends on the square of the past values ε_{t-i}, and is time-varying: an important shock will increase the current conditional variance, and the probability to have high magnitude shocks in the future. This process reproduces the stylized facts that are often observed in financial time series, and has been introduced by Engle (1982) under the name autoregressive conditional heteroscedasticity (or ARCH) model.

A natural extension of the ARCH(q) model is to consider that the conditional variance also depends on its past values:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \gamma_1\sigma_{t-1}^2 + \gamma_2\sigma_{t-2}^2 + \cdots + \gamma_p\sigma_{t-p}^2 + \\ &\quad \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 \end{aligned} \tag{10.44}$$

where the polynomial $(1 - \gamma_1 L - \cdots - \gamma_p L^p)$ has its roots outside the unit circle. To ensure that $\sigma_t^2 > 0$, wa can impose that $\alpha_i \geq 0$ and $\gamma_i \geq 0$. This model has been first formulated by Bollerslev (1986), and is known as a GARCH(p,q) process. The process y_t is stationary if the following condition holds:

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \gamma_i < 1$$

We can then show that the unconditional mean $\mathbb{E}[\varepsilon_t]$ is equal to zero, whereas the unconditional variance has the expression:

$$\mathbb{E}\left[\varepsilon_t^2\right] = \frac{\alpha_0}{1 - (\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \gamma_i)}$$

If we set $\eta_t = \varepsilon_t^2 - \sigma_t^2$, we get:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \gamma_i)\,\varepsilon_{t-i}^2 + \eta_t - \sum_{i=1}^{p} \gamma_i \eta_{t-i} \tag{10.45}$$

where $m = \max(p, q)$, $\gamma_i = 0$ for $i > p$ and $\alpha_i = 0$ for $i > q$. Since we have:

$$\mathbb{E}[\eta_t \mid \mathcal{F}_{t-1}] = \mathbb{E}\left[\varepsilon_t^2 - \sigma_t^2 \mid \mathcal{F}_{t-1}\right] = 0$$

and for $s \neq t$:

$$\begin{aligned}\mathbb{E}[\eta_s \eta_t] &= \mathbb{E}\left[\left(\varepsilon_s^2 - \sigma_s^2\right)\left(\varepsilon_t^2 - \sigma_t^2\right)\right] \\ &= \mathbb{E}\left[\left(\sigma_s^2 e_s^2 - \sigma_s^2\right)\left(\sigma_t^2 e_t^2 - \sigma_t^2\right)\right] \\ &= \mathbb{E}\left[\sigma_s^2\sigma_t^2\left(e_s^2 - 1\right)\left(e_t^2 - 1\right)\right] \\ &= 0\end{aligned}$$

we deduce that a GARCH(p,q) process for ε_t is equivalent to an ARMA(m,q) process for ε_t^2, where η_t is the innovation. Using the formulation (10.45), we retrieve the formula of the unconditional variance since we have:

$$\mathbb{E}\left[\varepsilon_t^2\right] = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \gamma_i)\,\mathbb{E}\left[\varepsilon_{t-i}^2\right] + \mathbb{E}[\eta_t] - \sum_{i=1}^{p} \gamma_i \mathbb{E}[\eta_{t-i}]$$

and $\mathbb{E}\left[\varepsilon_{t-i}^2\right] = \mathbb{E}\left[\varepsilon_t^2\right]$.

If $\sum_{i=1}^{m}(\alpha_i + \gamma_i) = \sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{p}\gamma_i = 1$, then the process ε_t^2 has a unit root and we obtain an integrated GARCH (or IGARCH) process (Engle and Bollerslev, 1986). For instance, the IGARCH(1,1) process is equal to:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \gamma_1\sigma_{t-1}^2 + \alpha_1\varepsilon_{t-1}^2 \\ &= \alpha_0 + \left(1 + \alpha_1\left(e_{t-1}^2 - 1\right)\right)\sigma_{t-1}^2\end{aligned}$$

If we assume that $\alpha_0 = 0$, we obtain:

$$\sigma_t^2 = \sigma_0^2 \prod_{i=0}^{t-1} \left(1 + \alpha_1\left(e_i^2 - 1\right)\right)$$

and:

$$\ln \sigma_t^2 = \ln \sigma_0^2 + \sum_{i=0}^{t-1} \ln\left(1 + \alpha_1\left(e_i^2 - 1\right)\right)$$

We deduce that if $\mathbb{E}\left[\ln\left(1 + \alpha_1\left(e_i^2 - 1\right)\right)\right] > 0$, then σ_t^2 tends to $+\infty$. If $\mathbb{E}\left[\ln\left(1 + \alpha_1\left(e_i^2 - 1\right)\right)\right] < 0$, σ_t^2 goes to 0. While shocks are persistent in unit root processes (e.g. the random walk), we see here that a shock in an IGARCH process may be persistent or not. Therefore, the concept of volatility persistence is not obvious and depends on the parameter values. Another important property of GARCH models is that they are heavy-tailed processes, meaning that the kurtosis (and even statistical moments) are greater than that of the normal distribution.

Remark 131 *GARCH models have been extended to multivariate processes, non-normal distributions (EGARCH), etc. In practice, these models are not very useful, because they are not tractable and difficult to calibrate. Only ARCH(p) and GARCH(p,q) models are used by professionals with low orders of p and q.*

Estimation If we make the approximation $y_t \sim \mathcal{N}\left(x_t^\top \beta, \sigma_t^2\right)$, the log-likelihood of the t^{th} observation is:

$$\boldsymbol{\ell}_t(\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_t^2 - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}$$

where $\varepsilon_t = y_t - x_t^\top \beta$ is the residual. The vector of parameters is $\theta = (\beta, \alpha, \gamma)$ where $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)$ and $\gamma = (\gamma_1, \ldots, \gamma_p)$. We then define the estimator by maximizing the log-likelihood function:

$$\hat{\theta}_{\text{QML}} = \arg\max \sum_{t=1}^{T} \boldsymbol{\ell}_t(\theta)$$

This estimator is called the quasi-maximum likelihood estimator, because we have assumed that ε_t is Gaussian. Of course, this is not true because σ_t^2 depends on past values of ε_t. However, we can show that this approach is consistent and defines a '*good*' estimator.

Application to the S&P 500 index We consider the monthly returns R_t of the S&P 500 index that have been plotted on page 664. To understand the dependence structure, we have reported in Figure 10.16 the autocorrelation function[30] (ACF), the partial autocorrelation function[31] (PACF) and the 95% significance test[32] of R_t and R_t^2. We deduce that the hypothesis that R_t is not autocorrelated is accepted, but the hypothesis that R_t^2 is not autocorrelated is rejected. This suggests that R_t is an heteroscedastic process. Since the

[30] Let y_t be a discrete-time series process. We reiterate that the autocorrelation function for lag k is defined by $\rho_y(k) = \dfrac{\gamma_y(k)}{\gamma_y(0)}$ where $\gamma_y(k) = \operatorname{cov}(y_t, y_{t-k})$.

[31] Partial autocorrelation is the autocorrelation between y_t and y_{t-k} after removing any linear dependence on $y_{t-1}, \ldots, y_{t-k+1}$. It is denoted by $\phi_y(k,k)$ and corresponds to the coefficient $\phi_{k,k}$ of the linear regression:

$$y_t - \bar{y} = \sum_{i=1}^{k} \phi_{k,i}\left(y_{t-i} - \bar{y}\right) + u_t$$

[32] The standard deviation of $\rho_y(k)$ and $\phi_y(k,k)$ is approximatively equal to $1/\sqrt{T}$ where T is the number of observations in the sample.

Ljung-Box test[33] is not rejected for large lag values – $Q_{R_t^2}(100) = 126.99$ and the p-value is 3.5%, we can assume that R_t is a GARCH process:

$$\begin{cases} R_t = c + \varepsilon_t \\ \varepsilon_t \sim \text{GARCH}(p, q) \end{cases}$$

Using $p = q = 1$, the quasi-maximum likelihood estimation gives:

$$\begin{cases} R_t = 69 \cdot 10^{-4} + \varepsilon_t \\ \sigma_t^2 = 8.3 \cdot 10^{-5} + 0.838 \cdot \sigma_{t-1}^2 + 0.120 \cdot \varepsilon_{t-1}^2 \end{cases}$$

All the parameters are statistically significant at the 99.9% confidence level. The annualized value of σ_t^2 is given in Figure 10.17. We notice that it varies between 9% and 33.3%, whereas the mean is equal to 14.15%. This value is close to the long-run volatility calculated with the full period, which is equal to 14.30%. In Figure 10.17, we have also reported the standardized residuals $e_t = \sigma_t^{-1}\varepsilon_t$. The ACF and PACF values of e_t^2 show that e_t^2 is not autocorrelated, and validate the choice of the GARCH(1,1) model.

Remark 132 *We have not discussed the choice of the lags p and q of the GARCH model. However, the estimation of higher order GARCH models does not improve the results of the GARCH(1,1) model. For example, if we consider a GARCH(2,2) model, we obtain the following results:*

$$\begin{cases} R_t = 66 \cdot 10^{-4} + \varepsilon_t \\ \sigma_t^2 = 8.5 \cdot 10^{-5} + 0.812 \cdot \sigma_{t-1}^2 + 0.073 \cdot \varepsilon_{t-1}^2 + 0.077 \cdot \varepsilon_{t-2}^2 \end{cases}$$

The estimate $\hat{\gamma}_2$ is equal to 0, and the p-values of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are equal to 16.3% and 3.4%. Therefore, this model is less convincing than the GARCH(1,1) model, where all the parameters are statistically significant at the 99.9% confidence level.

10.2.4.2 Stochastic volatility models

The Kalman filter approach A stochastic volatility model can be viewed as a GARCH model, where an innovation process is introduced in the equation of the conditional variance σ_t^2:

$$\gamma(L)\,\sigma_t^2 = \alpha(L)\,\varepsilon_t + \eta_t \tag{10.46}$$

where ε_t is a process with zero mean and unit variance, and η_t is the innovation process with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \sigma_\eta^2$. The parameter σ_η is known as the volatility of the volatility or vovol[34]. We have $\gamma(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p$ and $\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_q L^q$. In order to ensure the positivity of the conditional variance, we may prefer to use an EGARCH parametrization $h_t = \ln \sigma_t^2$:

$$y_t = x_t^\top \beta + \epsilon_t$$

where:

$$\begin{cases} \epsilon_t = \exp\left(\frac{1}{2}h_t\right) \cdot \varepsilon_t \\ h_t = \alpha_0 + \sum_{i=1}^p \gamma_i h_{t-i} + \sum_{i=1}^q \alpha_i \varepsilon_{t-i} + \eta_t \end{cases}$$

[33] The Ljung-Box test is a statistical test of randomness based on a number s of lags:

$$Q_y(s) = T(T+2)\sum_{k=1}^{s} \frac{\hat{\rho}_y^2(k)}{T-k}$$

Under the null hypothesis that the data are independently distributed, we have $Q_y(s) \sim \chi^2(s)$.

[34] See page 570 for its definition.

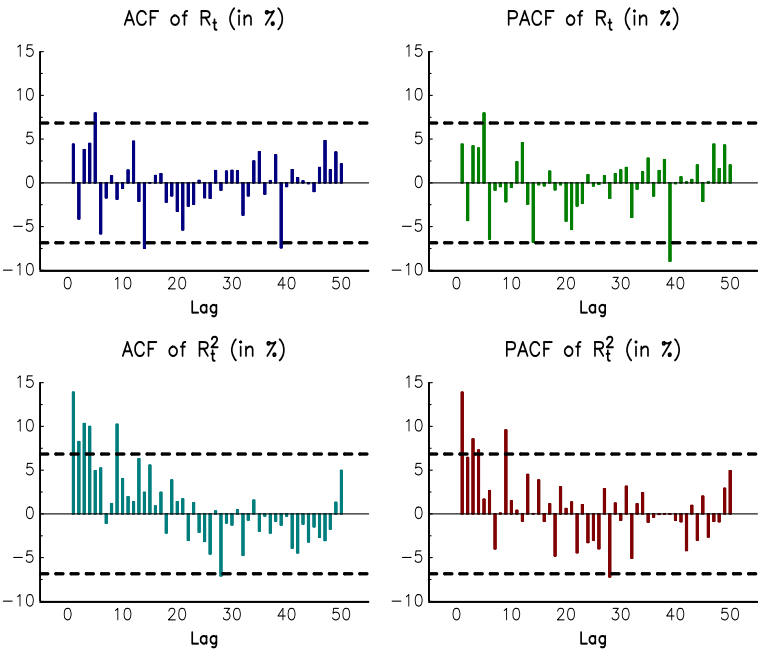


FIGURE 10.16: ACF and PACF of R_t and R_t^2

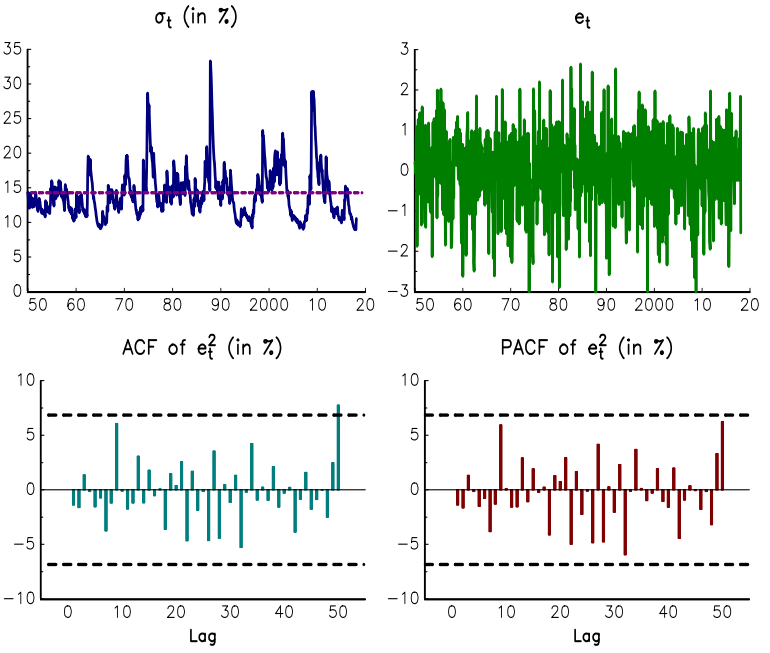


FIGURE 10.17: Diagnostic checking of the GARCH(1,1) model

The processes ε_t and η_t are not necessarily normal, and we can use heavy-tail probability distributions. However, the estimation of the model by the method of maximum likelihood is complex because the log-likelihood is a mixture of conditional log-variances h_t.

If we consider the canonical stochastic volatility model ($p = 1$ and $q = 0$), we have:

$$\begin{cases} y_t = \exp\left(\frac{1}{2}h_t\right) \cdot \varepsilon_t \\ h_t = \alpha_0 + \gamma_1 h_{t-1} + \eta_t \end{cases} \tag{10.47}$$

Since h_t is an AR(1) process, we deduce that:

$$\mathbb{E}[h_t] = \frac{\alpha_0}{1-\gamma_1}$$

and:

$$\text{var}(h_t) = \frac{\sigma_\eta^2}{1-\gamma_1^2}$$

Harvey *et al.* (1994) propose to define the measurement variable as $\ln y_t^2$ instead of y_t. We obtain the state space model representation:

$$\begin{cases} \ln y_t^2 = c + h_t + \xi_t \\ h_t = \alpha_0 + \gamma_1 h_{t-1} + \eta_t \end{cases} \tag{10.48}$$

where h_t is the state variable, $c = \mathbb{E}\left[\ln \varepsilon_t^2\right]$ and[35] $\xi_t = \ln \varepsilon_t^2 - \mathbb{E}\left[\ln \varepsilon_t^2\right]$. If we approximate ξ_t as a Gaussian random variable, we can estimate h_t by using the Kalman filter with the initialization $h_0 \sim \mathcal{N}\left(\frac{\alpha_0}{1-\gamma_1}, \frac{\sigma_\eta^2}{1-\gamma_1^2}\right)$.

An alternative model of Process (10.47) is:

$$\begin{cases} y_t = \exp\left(\frac{1}{2}h_t\right) \cdot \varepsilon_t \\ h_t = \alpha_0 + \gamma_1 (h_{t-1} - \alpha_0) + \eta_t \end{cases}$$

or equivalently:

$$\begin{cases} y_t = \exp\left(\frac{1}{2}\alpha_0\right) \cdot \exp\left(\frac{1}{2}h_t\right) \cdot \varepsilon_t \\ h_t = \gamma_1 h_{t-1} + \eta_t \end{cases} \tag{10.49}$$

We deduce the following state space model representation:

$$\begin{cases} \ln y_t^2 = c + h_t + \xi_t \\ h_t = \gamma_1 h_{t-1} + \eta_t \end{cases} \tag{10.50}$$

where $c = \alpha_0 + \mathbb{E}\left[\ln \varepsilon_t^2\right]$ and $\xi_t = \ln \varepsilon_t^2 - \mathbb{E}\left[\ln \varepsilon_t^2\right]$. Again, we can estimate h_t by using the Kalman filter with the initialization $h_0 \sim \mathcal{N}\left(0, \frac{\sigma_\eta^2}{1-\gamma_1^2}\right)$.

Using the monthly returns of the S&P 500 index, we estimate the model (10.50) by maximum likelihood:

$$\begin{cases} \ln\left(R_t - \bar{R}\right)^2 = -7.86 + h_t + \xi_t \\ h_t = 0.93 \cdot h_{t-1} + \eta_t \end{cases}$$

where $\sigma_\eta = 21\%$. All the coefficients are significant at the 99% confidence level. In Figure 10.18, we report the annualized volatility estimated by Kalman filter and smoother. We can compare these values with those obtained with the GARCH(1,1) model and the 12-month historical volatilities. We notice that the stochastic volatility model produces "*less noisy*" volatilities, because large shocks may be due to an increase of h_t, but also to a shock on ε_t.

[35] Since $\varepsilon_t \sim \mathcal{N}(0,1)$, Abramowitz and Stegun (1970) showed that $\ln \varepsilon_t^2$ has a log-chi-squared distribution with mean $\psi(1) - \ln 2$ and variance $\pi^2/2$, where $\psi(x)$ is the digamma function.

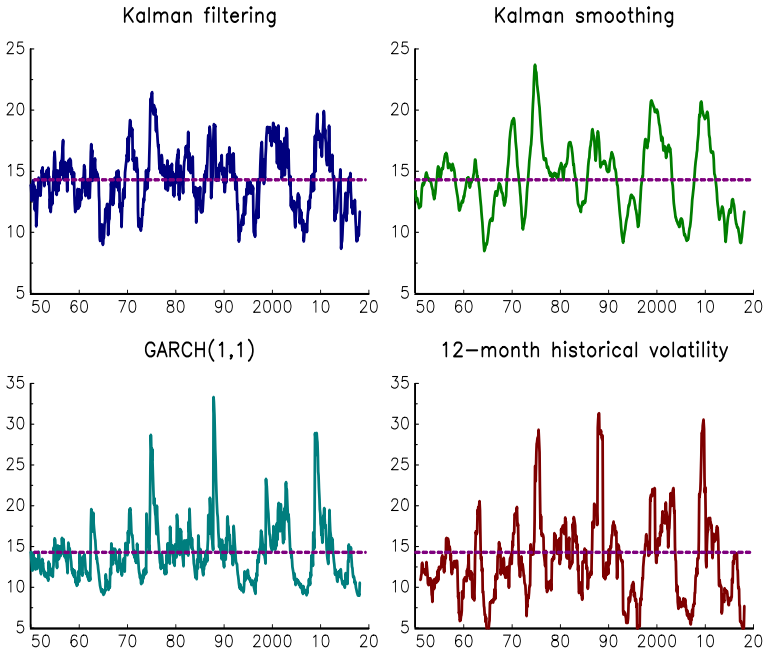


FIGURE 10.18: Estimation of the stochastic volatility model

Remark 133 *The canonical stochastic volatility model has been extended in many directions, such as asymmetric or fat-tailed distributions (Harvey and Shephard, 1996; Broto and Ruiz, 2004).*

The MCMC approach Kim *et al.* (1998) suggest to use Markov Chain Monte Carlo (MCMC) approach for estimating the previous model. This method is more flexible than the Kalman filter, but it also more time-consuming and complex. In Figure 10.19, we report the estimates of the stochastic volatility for 4 algorithms: griddy Gibbs sampler, Random Walk Metropolis algorithm, Metropolis-Hastings method, and Metropolis-Hastings algorithm within griddy Gibbs. We notice that these MCMC estimates are close to the KF estimates.

10.2.5 Spectral analysis

Until now, we have analyzed stochastic processes in the time domain. In this section, we focus on the frequency domain or the spectral analysis. We do not present the time-frequency (or wavelet) analysis since this approach has not been successful to solve financial problems.

10.2.5.1 Fourier analysis

Let $(y_t, t \in \mathbb{Z})$ be a centered stationary process. The (discrete) Fourier transform of y_t is equal to:

$$y(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} y_t e^{-i\lambda t}$$

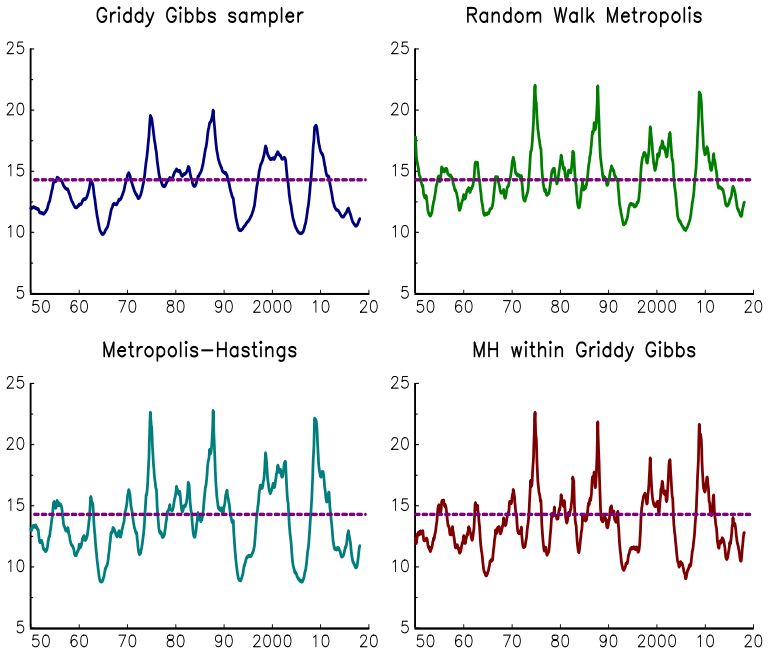


FIGURE 10.19: MCMC estimates of the stochastic volatility model

where $\lambda \in [0, 2\pi]$. We define the inverse Fourier transform as follows:

$$y_t = \int_{-\pi}^{\pi} y(\lambda)\, e^{i\lambda t}\, \mathrm{d}\lambda$$

The idea of Fourier analysis is to approximate the process y_t by a weighted finite sum of sine and cosine functions:

$$y_t \approx \sum_{j=0}^{n} \alpha_j \cos(\lambda_j t) + \sum_{j=0}^{n} \beta_j \sin(\lambda_j t)$$

where α_j and β_j are the Fourier coefficients. Under some technical assumptions, we can show that[36]:

$$\begin{aligned} y_t &= \lim_{n\to\infty} \left(\sum_{j=0}^{n} \alpha_j \cos(\lambda_j t) + \sum_{j=0}^{n} \beta_j \sin(\lambda_j t) \right) \\ &= \int_0^{\pi} \cos(\lambda t)\ \mathrm{d}A(\lambda) + \int_0^{\pi} \sin(\lambda t)\ \mathrm{d}B(\lambda) \end{aligned}$$

As noticed by Pollock (1999), the Fourier analysis is made under the assumptions that $A(\lambda)$ and $B(\lambda)$ are two independent stochastic processes with zero mean, and non-overlapping increments of each process are uncorrelated[37]. Moreover, we have:

$$\mathrm{var}(\mathrm{d}A(\lambda)) = \mathrm{var}(\mathrm{d}B(\lambda)) = 2 \cdot \mathrm{d}F_y(\lambda)$$

[36] We have the correspondence $\mathrm{d}A(\lambda_j) = \alpha_j$ and $\mathrm{d}B(\lambda_j) = \beta_j$.

[37] This means that $\mathbb{E}[\mathrm{d}A(\lambda)] = \mathbb{E}[\mathrm{d}B(\lambda)] = 0$, $\mathbb{E}[\mathrm{d}A(\lambda) \cdot \mathrm{d}B(\lambda')] = 0$ for all (λ, λ') and $\mathbb{E}[\mathrm{d}A(\lambda) \cdot \mathrm{d}A(\lambda')] = \mathbb{E}[\mathrm{d}B(\lambda) \cdot \mathrm{d}B(\lambda')] = 0$ if $\lambda \neq \lambda'$.

$F_y(\lambda)$ is called the spectral distribution function, and its derivative $f_y(\lambda)$ is the spectral density function. Pollock (1999) shows that:

$$\begin{aligned} y_t &= \int_0^\pi e^{i\lambda t}\,\mathrm{d}Z(\lambda) + \int_0^\pi e^{-i\lambda t}\,\mathrm{d}Z^\star(\lambda) \\ &= \int_{-\pi}^\pi e^{i\lambda t}\,\mathrm{d}Z(\lambda) \end{aligned}$$

where:

$$\mathrm{d}Z(\lambda) = \frac{\mathrm{d}A(\lambda) - i\mathrm{d}B(\lambda)}{2}$$

and:

$$\mathrm{d}Z^\star(\lambda) = \frac{\mathrm{d}A(\lambda) + i\mathrm{d}B(\lambda)}{2}$$

The decomposition $y_t = \int_{-\pi}^\pi e^{i\lambda t}\,\mathrm{d}Z(\lambda)$ is the spectral representation of y_t. We notice that the processes $Z(\lambda)$ and $Z^\star(\lambda)$ are not independent:

$$\begin{aligned} \mathbb{E}\left[\mathrm{d}Z(\lambda)\,\mathrm{d}Z^\star(\lambda)\right] &= \frac{1}{2}\operatorname{var}(\mathrm{d}A(\lambda)) \\ &= f_y(\lambda)\,\mathrm{d}\lambda \end{aligned}$$

but we verify that $\mathbb{E}\left[\mathrm{d}Z(\lambda)\,\mathrm{d}Z^\star(\lambda')\right] = 0$ if $\lambda \neq \lambda'$.

10.2.5.2 Definition of the spectral density function

Let $(y_t, t \in \mathbb{Z})$ be a centered stationary process. We denote by $\gamma_y(k)$ the autocovariance function. We can show that there is a function $f_y(\lambda)$ such that:

$$\gamma_y(k) = \int_{-\pi}^\pi f_y(\lambda)\, e^{i\lambda k}\,\mathrm{d}\lambda \tag{10.51}$$

The function $f_y(\lambda)$ is called the spectral density of the process y_t. We can demonstrate that the two following conditions are equivalent:

1. y_t has the spectral density $f_y(\lambda)$.
2. There is a white noise $(\varepsilon_t, t \in \mathbb{Z})$ and a sequence $(\psi_s, s \in \mathbb{Z})$ satisfying $\sum_{s=-\infty}^\infty \psi_s^2 < \infty$ such that:

$$y_t = \sum_{s=-\infty}^\infty \psi_s \varepsilon_{t-s} \tag{10.52}$$

In this case, the spectral density function $f_y(\lambda)$ is defined by:

$$\begin{aligned} f_y(\lambda) &= \frac{1}{2\pi}\sum_{k=-\infty}^\infty \gamma_y(k)\, e^{-i\lambda k} \qquad (10.53) \\ &= \frac{\gamma_y(0)}{2\pi} + \frac{1}{\pi}\sum_{k=1}^\infty \gamma_y(k)\cos(\lambda k) \\ &= \frac{1}{2\pi}\sum_{k=-\infty}^\infty \gamma_y(k)\cos(\lambda k) \end{aligned}$$

Therefore, the spectral density function contains the same information that the autocovariance function. We can use both $f_y(\lambda)$ or $\gamma_y(k)$ to characterize a stationary stochastic process. The only difference is that the autocovariance function is a representation in the time domain whereas the spectral density function is a representation in the frequency domain. This result is not surprising if we refer to the Fourier analysis of time series. Indeed, the spectral density function is also the variance of the processes $Z(\lambda)$ and $Z^{\star}(\lambda)$.

Remark 134 *We verify that:*

$$\begin{aligned}
\int_{-\pi}^{\pi} f_y(\lambda) e^{i\lambda k} \mathrm{d}\lambda &= \int_{-\pi}^{\pi} \left(\frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_y(h) e^{-i\lambda h} \right) e^{i\lambda k} \mathrm{d}\lambda \\
&= \sum_{h=-\infty}^{\infty} \frac{\gamma_y(h)}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda(k-h)} \mathrm{d}\lambda \\
&= \gamma_y(k) + \sum_{h \neq k} \frac{\gamma_y(h)}{2\pi} \left[\frac{e^{i\lambda(k-h)}}{i(k-h)} \right]_{-\pi}^{\pi} \\
&= \gamma_y(k) + \sum_{h \neq k} \frac{\gamma_y(h)}{2\pi} \left(\frac{e^{i\pi(k-h)} - e^{-i\pi(k-h)}}{i(k-h)} \right) \\
&= \gamma_y(k) + \sum_{h \neq k} \gamma_y(h) \frac{\sin(\pi(k-h))}{\pi(k-h)} \\
&= \gamma_y(k)
\end{aligned}$$

because $\sin(\pi(k-h)) = 0$.

10.2.5.3 Frequency domain localization

The information contained in the autocovariance function is encoded differently in the spectral density. Consider the white noise process $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. We have $\gamma_\varepsilon(0) = \sigma^2$ and $\gamma_\varepsilon(k) = 0$ for $k \neq 0$. We deduce that:

$$\begin{aligned}
f_\varepsilon(\lambda) &= \frac{1}{2\pi} \gamma_\varepsilon(0) \cos(\lambda 0) \\
&= \frac{\sigma^2}{2\pi}
\end{aligned}$$

The spectral density of the white noise process is then a constant. It is the worst localized signal in the frequency domain. Consider now the process η_t such that $f_\eta(\lambda_c) = c$ and $f_\eta(\lambda) = 0$ for $\lambda \neq \lambda_c$. It is the best localized signal in the frequency domain. Let us analyze the cycle signal:

$$y_t = 2\sin\left(\frac{2\pi}{p} t\right)$$

In Figures 10.20 and 10.21, we represent this cycle for different periods p and the corresponding autocorrelation function $\rho_y(k) = \gamma_y(k)/\gamma_y(0)$. We calculate the spectral density as follows:

$$f_y(\lambda) \approx \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{10000} \gamma_y(k) \cos(\lambda k)$$

In Figure 10.22, we notice that the function $f_y(\lambda)$ can be approximated by $f_\eta(\lambda)$ where the frequency λ_c is equal to $2\pi/p$. This is then the inverse of the period p (normalized by

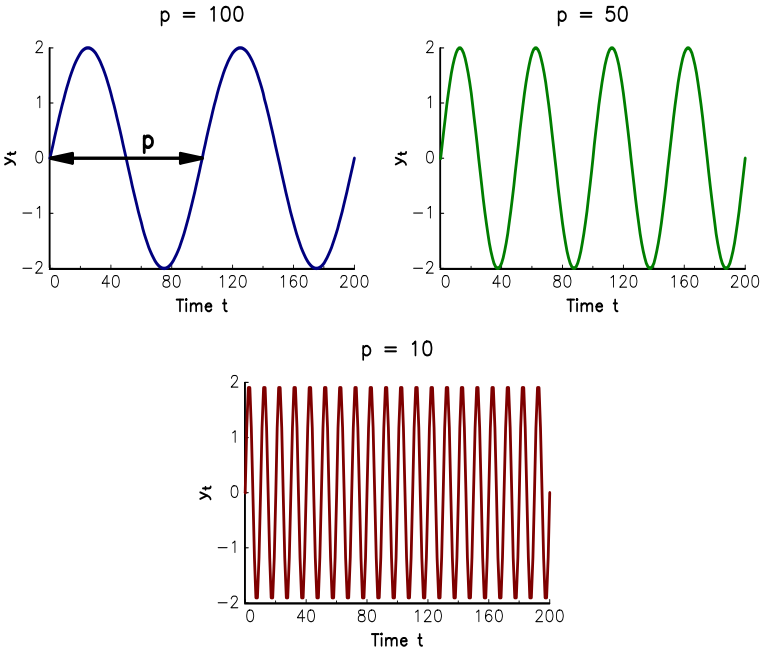


FIGURE 10.20: Time representation of the process x_t

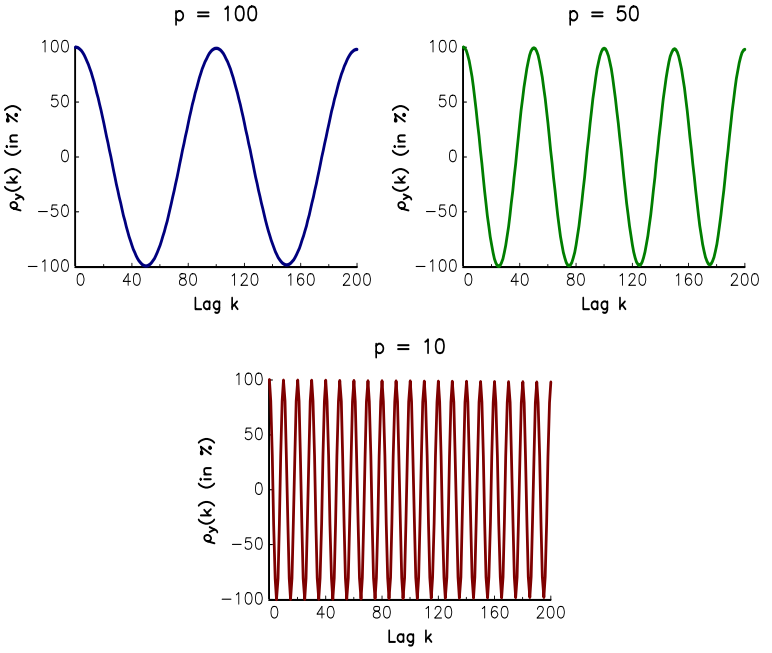


FIGURE 10.21: Autocorrelation representation of the process x_t

2π). We recall that $1/p$ is also the number of cycle per time unit. Suppose that we calculate the spectral density of an economic cycle using monthly data, for example the business cycle that has a 7-year period. Then the spectral density must have a high value c at the frequency λ_c:

$$\lambda_c = \frac{2\pi}{7 \times 12} = 0.0748$$

The localization of low frequency phenomena will be at λ close to 0 whereas the localization of high frequency phenomena will be at higher frequencies λ (close to π). λ is also called the harmonic frequency. Low harmonic frequencies corresponds to long-term components while high harmonic frequencies are more focused on short-term components.

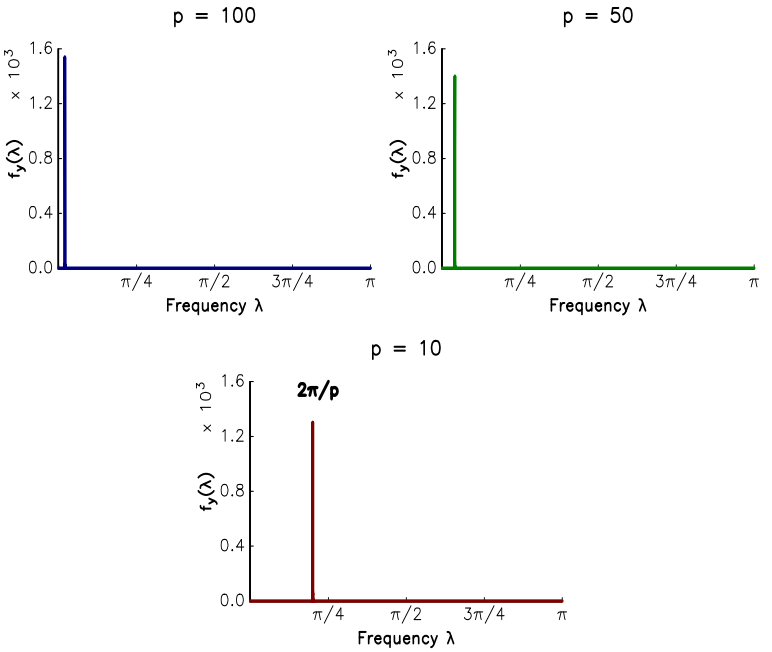


FIGURE 10.22: Spectral representation of the process x_t

10.2.5.4 Main properties

Independent processes Let x_t and y_t be two centered and independent stationary processes. If we consider the process $z_t = x_t + y_t$, we have:

$$\begin{aligned}
\gamma_{x+y}(k) &= \mathbb{E}[z_t z_{t-k}] \\
&= \mathbb{E}[x_t x_{t-k}] + \mathbb{E}[y_t y_{t-k}] + \mathbb{E}[x_t y_{t-k}] + \mathbb{E}[y_t x_{t-k}] \\
&= \gamma_x(k) + \gamma_y(k)
\end{aligned}$$

It follows that:

$$\begin{aligned}
f_z(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_z(k) \cos(\lambda k) \\
&= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_x(k) \cos(\lambda k) + \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_y(k) \cos(\lambda k) \\
&= f_x(\lambda) + f_y(\lambda)
\end{aligned}$$

The spectral density function of the sum of independent processes is then the sum of their spectral density functions:

$$f_{x+y}(\lambda) = f_x(\lambda) + f_y(\lambda)$$

Linear filtering Equation (10.51) shows that the autocovariance function $\gamma_y(k)$ is the inverse Fourier transform of the spectral density function $f_y(\lambda)$, whereas Equation (10.53) means that the spectral density function $f_y(\lambda)$ is the Fourier transform of the autocovariance function $\gamma_y(k)$. However, we may wonder what the implication of Equation (10.52) is. We recognize the Wold decomposition $y_t = \psi(L)\varepsilon_t$ where $\psi(L) = \sum_{s=-\infty}^{\infty} \psi_s L^s$. It follows that the autocovariance function of y_t is equal to:

$$\begin{aligned}
\gamma_y(k) &= \mathbb{E}[y_t y_{t-k}] \\
&= \mathbb{E}\left[\left(\sum_{r=-\infty}^{\infty} \psi_r \varepsilon_{t-r}\right)\left(\sum_{s=-\infty}^{\infty} \psi_s \varepsilon_{t-s-k}\right)\right] \\
&= \sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \psi_r \psi_s \gamma_\varepsilon(s+k-r)
\end{aligned}$$

We deduce that the spectral density function is:

$$\begin{aligned}
f_y(\lambda) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_y(k)\, e^{-i\lambda k} \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left(\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \psi_r \psi_s \gamma_\varepsilon(s+k-r)\right) e^{-i\lambda k}
\end{aligned}$$

We introduce the index $h = s + k - r$:

$$\begin{aligned}
f_y(\lambda) &= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\left(\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \psi_r \psi_s \gamma_\varepsilon(h)\right) e^{-i\lambda(h-s+r)} \\
&= \sum_{r=-\infty}^{\infty} \psi_r e^{-i\lambda r} \sum_{s=-\infty}^{\infty} \psi_s e^{i\lambda s}\left(\frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \gamma_\varepsilon(h)\, e^{-i\lambda h}\right) \\
&= \psi\left(e^{-i\lambda}\right)\psi\left(e^{i\lambda}\right) f_\varepsilon(\lambda) \\
&= \psi\left(e^{-i\lambda}\right)\psi^*\left(e^{-i\lambda}\right) f_\varepsilon(\lambda) \\
&= \left|\psi\left(e^{-i\lambda}\right)\right|^2 f_\varepsilon(\lambda) \\
&= \frac{\sigma^2}{2\pi}\left|\psi\left(e^{-i\lambda}\right)\right|^2
\end{aligned}$$

This result is important because we obtain an analytical expression of $f_y(\lambda)$ that does not need to use the Fourier transform of the autocovariance function $\gamma_y(k)$. Moreover, we can generalize this calculus based on the Wold decomposition to any linear filter $\varphi(L)$:

$$y_t = \varphi(L)\, x_t$$

where x_t is a stationary centered process. By using the same approach, we obtain:

$$f_y(\lambda) = \left|\varphi\left(e^{-i\lambda}\right)\right|^2 f_x(\lambda)$$

Therefore, Fourier analysis transforms convolutions into multiplications. We can write the complex transfer function $z = \varphi\left(e^{-i\lambda}\right)$ in the polar form $\varphi\left(e^{-i\lambda}\right) = r(\lambda) e^{i\theta(\lambda)}$ where $r(\lambda) = |z|$ and $\theta(\lambda) = \text{atan2}(\text{Im}(z), \text{Re}(z))$. We deduce that:

$$\begin{aligned} y_t &= \int_{-\pi}^{\pi} e^{i\lambda t} \, \mathrm{d}Z_y(\lambda) \\ &= \int_{-\pi}^{\pi} r(\lambda) e^{i(\lambda t - \theta(\lambda))} \, \mathrm{d}Z_x(\lambda) \end{aligned}$$

$r(\lambda)$ and $\theta(\lambda)$ are called the power-shift and phase-shift of the filter since they impact respectively the amplitude and the period of the cyclical components.

Spectral density of some useful processes Let us first consider the AR(1) process $y_t = \phi y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$. We recall that the autocovariance function is equal to:

$$\gamma_y(k) = \frac{\sigma^2 \phi^{|k|}}{1 - \phi^2}$$

The Fourier transform of $\gamma_y(k)$ gives:

$$\begin{aligned} f_y(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_y(k) e^{-i\lambda k} \\ &= \frac{1}{2\pi} \left(\frac{\sigma^2}{1-\phi^2}\right) \sum_{k=-\infty}^{\infty} \phi^{|k|} e^{-i\lambda k} \\ &= \frac{1}{2\pi} \left(\frac{\sigma^2}{1-\phi^2}\right) \left(1 + \sum_{k=1}^{\infty} \phi^k e^{i\lambda k} + \sum_{k=1}^{\infty} \phi^k e^{-i\lambda k}\right) \\ &= \frac{1}{2\pi} \left(\frac{\sigma^2}{1-\phi^2}\right) \left(1 + \frac{\phi e^{i\lambda}}{1 - \phi e^{i\lambda}} + \frac{\phi e^{-i\lambda}}{1 - \phi e^{-i\lambda}}\right) \\ &= \frac{1}{2\pi} \left(\frac{\sigma^2}{1-\phi^2}\right) \left(\frac{1 - \phi e^{i\lambda} \phi e^{-i\lambda}}{(1 - \phi e^{i\lambda})(1 - \phi e^{-i\lambda})}\right) \\ &= \frac{\sigma^2}{2\pi} \frac{1}{(1 - 2\phi \cos\lambda + \phi^2)} \end{aligned}$$

To calculate the spectral density function, we can also use the result obtained for linear filters. We have:

$$f_y(\lambda) = \left|\phi\left(e^{-i\lambda}\right)\right|^2 f_\varepsilon(\lambda)$$

where $\phi(L) = 1 - \phi L$. We deduce that:

$$\begin{aligned} \left|\phi\left(e^{-i\lambda}\right)\right|^2 &= \left|1 - \phi e^{-i\lambda}\right|^2 \\ &= \left|1 - \phi(\cos\lambda - i\sin\lambda)\right|^2 \\ &= (1 - \phi\cos\lambda)^2 + \phi^2 \sin^2\lambda \\ &= 1 - 2\phi\cos\lambda + \phi^2 \end{aligned}$$

We obtain the same expression of the spectral density function:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{(1 - 2\phi\cos\lambda + \phi^2)}$$

More generally, the spectral density function of the ARMA model $\Phi(L) y_t = \Theta(L) \varepsilon_t$ is given by:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|\Theta\left(e^{-i\lambda}\right)\right|^2}{\left|\Phi\left(e^{-i\lambda}\right)\right|^2}$$

In the case of the MA(1) process $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, we have:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi}\left(1 - 2\theta\cos\lambda + \theta^2\right)$$

For an ARMA(1,1) process $y_t = \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$, we combine the AR(1) and MA(1) filters and we obtain:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left(1 - 2\theta\cos\lambda + \theta^2\right)}{\left(1 - 2\phi\cos\lambda + \phi^2\right)}$$

We now consider the process $z_t = x_t + y_t$, where $x_t = \phi_1 x_{t-1} + u_t$ is an AR(1) process and $y_t = v_t - \theta_1 v_{t-1}$ is an MA(1) process that is independent from x. We have:

$$\begin{aligned} f_z(\lambda) &= f_x(\lambda) + f_y(\lambda) \\ &= \frac{\sigma_u^2}{2\pi} \frac{1}{\left(1 - 2\phi_1\cos\lambda + \phi_1^2\right)} + \frac{\sigma_v^2}{2\pi}\left(1 - 2\theta_1\cos\lambda + \theta_1^2\right) \end{aligned}$$

In Figure 10.23, we represent the spectral density function of different process[38]:

- a white noise process $y_t = \varepsilon_t$;
- an AR(p) process $y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$;
- a MA(q) process $y_t = \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$;
- an ARMA(p,q) process $y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$; ARMA #1 corresponds to the set of parameters $\phi_1 = 0.75$, $\phi_2 = -0.5$, $\theta_1 = 0.75$, $\theta_2 = -0.5$ and $\theta_3 = 0.25$ whereas ARMA #2 corresponds to $\phi_1 = 0.5$, $\phi_2 = 0.15$, $\theta_1 = 0.75$, $\theta_2 = -0.1$ and $\theta_3 = 0.15$.

We notice that some processes are well-located in the frequency domain, meaning that they are more '*cyclical*'. On the contrary, MA processes are more '*flat*'.

We now introduce the notion of stationary form, which is an important concept in spectral analysis. Consider the following model:

$$\begin{cases} z_t &= x_t + y_t \\ x_t &= x_{t-1} + u_t - \theta_1 u_{t-1} \\ y_t &= v_t \end{cases}$$

where $u_t \sim \mathcal{N}\left(0, \sigma_u^2\right)$ and $v_t \sim \mathcal{N}\left(0, \sigma_v^2\right)$. It is obvious that z_t is not stationary, implying that there is no spectral density function associated to the process z_t. However, we notice that $(1 - L) z_t$ is stationary because we have:

$$z_t - z_{t-1} = (1 - \theta_1 L) u_t + (1 - L) v_t$$

$\mathcal{S}(z_t) = z_t - z_{t-1}$ is called the '*stationary form*' of z_t and we have:

$$f_{\mathcal{S}(z)}(\lambda) = \left(1 - 2\theta_1\cos\lambda + \theta_1^2\right)\frac{\sigma_u^2}{2\pi} + 2\left(1 - \cos\lambda\right)\frac{\sigma_v^2}{2\pi}$$

[38] The standard deviation σ of the noise ε_t is set to 20%.

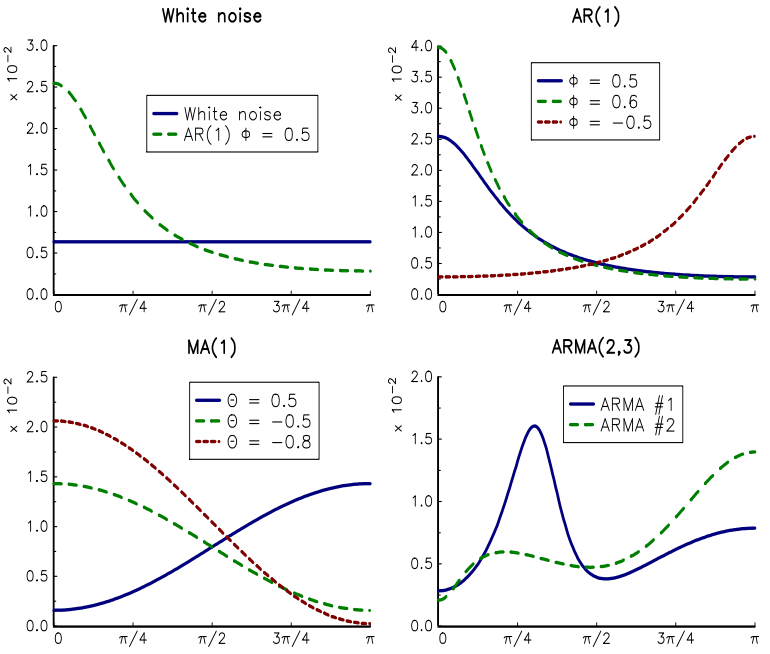


FIGURE 10.23: Spectral density function of ARMA processes

In the general case, if x_t has a stationary form, there is a lag polynomial $\varphi(L)$ such that $\mathcal{S}(x_t) = \varphi(L)\, x_t$.

On page 652, we have already studied structural time series models (STSM) with unobservable components (Harvey, 1990). They generally have a state space representation. Since they are non stationary, we have to find the corresponding stationary form. Let us consider the '*local level*' model (LL):

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \eta_t \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ and $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$. The stationary form of y_t is:

$$\begin{aligned} \mathcal{S}(y_t) &= y_t - y_{t-1} \\ &= \eta_t + (1-L)\,\varepsilon_t \end{aligned}$$

We deduce that:

$$f_{\mathcal{S}(y)}(\lambda) = \frac{\sigma_\eta^2 + 2(1-\cos\lambda)\,\sigma_\varepsilon^2}{2\pi}$$

The '*local linear trend*' model (LLT) is given by:

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ and $\zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$. The stationary form of y_t is:

$$\begin{aligned} \mathcal{S}(y_t) &= (1-L)^2\, y_t \\ &= \zeta_{t-1} + (1-L)\,\eta_t + (1-L)^2\,\varepsilon_t \end{aligned}$$

whereas the spectral density function is:

$$f_{\mathcal{S}(y)}(\lambda) = \frac{\sigma_\zeta^2 + 2(1-\cos\lambda)\sigma_\eta^2 + 4(1-\cos\lambda)^2\sigma_\varepsilon^2}{2\pi}$$

The '*basic structural model*' (BSM) has the following expression:

$$\begin{cases} y_t = \mu_t + \gamma_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \\ \sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$, $\zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$ and $\omega_t \sim \mathcal{N}\left(0, \sigma_\omega^2\right)$. The stationary form of y_t is:

$$\begin{aligned} \mathcal{S}(y_t) &= (1-L)(1-L^s)\, y_t \\ &= (1-L)(1-L^s)\,\varepsilon_t + (1-L^s)\,\eta_t + \left(\sum_{i=1}^{s} L^s\right)\zeta_{t-i} + \\ &\quad \left(1-2L+L^2\right)\omega_t \end{aligned}$$

It follows that the spectral density function is equal to:

$$\begin{aligned} f_{\mathcal{S}(y)}(\lambda) &= g_{(1-L)(1-L^s)}(\lambda)\frac{\sigma_\varepsilon^2}{2\pi} + g_{(1-L^s)}(\lambda)\frac{\sigma_\eta^2}{2\pi} + \\ &\quad g_{\left(\sum_{i=1}^{s} L^s\right)}(\lambda)\frac{\sigma_\zeta^2}{2\pi} + g_{(1-2L+L^2)}(\lambda)\frac{\sigma_\omega^2}{2\pi} \end{aligned}$$

where:

$$\begin{aligned} g_{(1-L^s)}(\lambda) &= 2(1-\cos s\lambda) \\ g_{(1-2L+L^2)}(\lambda) &= 6 - 8\cos\lambda + 2\cos 2\lambda \\ g_{\left(\sum_{i=1}^{s} L^s\right)}(\lambda) &= s + 2\sum_{j=1}^{s-1}(s-j)\cos j\lambda \\ g_{(1-L)(1-L^s)}(\lambda) &= 4(1-\cos s\lambda)(1-\cos\lambda) \end{aligned}$$

We now consider a variant of the basic structural model:

$$y_t = \mu_t + \beta_t + \gamma_t + \varepsilon_t$$

where μ_t is the long-run component, β_t is the mean-reverting component and γ_t is the seasonal component:

$$\begin{cases} \mu_t = \mu_{t-1} + \eta_t \\ \beta_t = \phi\beta_{t-1} + \zeta_t \\ \sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \end{cases}$$

where η_t, ζ_t and ω_t are independent white noise processes with variance σ_η^2, σ_ζ^2 and σ_ω^2. μ_t is then a random walk, β_t is an AR(1) process and γ_t is a stochastic seasonal process because we have $\gamma_t = \gamma_{t-s} + \omega_t - \omega_{t-1}$. If $\sigma_\omega^2 = 0$, the seasonal component is deterministic ($\gamma_t = \gamma_{t-s}$). As for the basic structural model, the stationary form is:

$$\begin{aligned} \mathcal{S}(y_t) &= (1-L)(1-L^s)\, y_t \\ &= (1-L)(1-L^s)\,\varepsilon_t + (1-L^s)\,\eta_t + \\ &\quad \left(\frac{1-L-L^s+L^{s+1}}{1-\phi L}\right)\zeta_t + \left(1-2L+L^2\right)\omega_t \end{aligned}$$

We deduce that:

$$\begin{aligned} f_{\mathcal{S}(y)}(\lambda) &= g_{(1-L)(1-L^s)}(\lambda) \cdot \frac{\sigma_\varepsilon^2}{2\pi} + g_{(1-L^s)}(\lambda) \cdot \frac{\sigma_\eta^2}{2\pi} + \\ & g_{\left(\frac{1-L-L^s+L^{s+1}}{1-\phi L}\right)}(\lambda) \cdot \frac{\sigma_\zeta^2}{2\pi} + g_{(1-2L+L^2)}(\lambda) \cdot \frac{\sigma_\omega^2}{2\pi} \end{aligned}$$

where:

$$g_{\left(\frac{1-L-L^s+L^{s+1}}{1-\phi L}\right)}(\lambda) = \frac{g_{(1-L-L^s+L^{s+1})}(\lambda)}{g_{(1-\phi L)}(\lambda)}$$

and[39]:

$$g_{(1-L-L^s+L^{s+1})}(\lambda) = 4 - 4\cos\lambda - 4\cos s\lambda + 2\cos(s-1)\lambda + 2\cos(s+1)\lambda$$

The '*cycle model*' (CM) is defined by the following state space model:

$$\left\{ \begin{array}{l} y_t = \psi_t \\ \begin{pmatrix} \psi_t \\ \psi_t^\star \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^\star \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^\star \end{pmatrix} \end{array} \right.$$

where $\kappa_t \sim \mathcal{N}(0, \sigma^2)$ and $\kappa_t^\star \sim \mathcal{N}(0, \sigma^2)$. Harvey (1990) showed that:

$$y_t = \left(\frac{1-\rho\cos\lambda_c L}{1-2\rho\cos\lambda_c L+\rho^2 L^2}\right)\kappa_t + \left(\frac{\rho\sin\lambda_c L}{1-2\rho\cos\lambda_c L+\rho^2 L^2}\right)\kappa_t^\star$$

and:

$$f_y(\lambda) = \left(\frac{1+\rho^2-2\rho\cos\lambda_c\cos\lambda}{1+\rho^4+4\rho^2\cos^2\lambda_c - 4\rho(1+\rho^2)\cos\lambda_c\cos\lambda + 2\rho^2\cos 2\lambda}\right)\frac{\sigma^2}{2\pi}$$

We have represented the spectral density function of the previous structural time series models in Figures 10.24 and 10.25. The set of parameters are the following:

- local level model: $\sigma_\varepsilon = 0.20$, $\sigma_\eta = 0.10$ for Model #1, $\sigma_\eta = 0.20$ for Model #2 and $\sigma_\eta = 0.30$ for Model #3;
- local linear trend model: $\sigma_\varepsilon = 0.20$, $\sigma_\zeta = 0.10$, $\sigma_\eta = 0.10$ for Model #4, $\sigma_\eta = 0.20$ for Model #5 and $\sigma_\eta = 0.30$ for Model #6;
- basic structural model: $\sigma_\varepsilon = 0.10$, $\sigma_\eta = 0.10$, $\sigma_\zeta = 0.10$, $\sigma_\omega = 0.10$, $s = 4$ for Model #7 and $s = 12$ for Model #9; for Model #8, we have $\sigma_\varepsilon = 0.20$, $\sigma_\eta = 0.30$, $\sigma_\zeta = 0.10$, $\sigma_\omega = 0.10$ and $s = 4$ whereas for Model #10 we have $\sigma_\varepsilon = 0.10$, $\sigma_\eta = 0.10$, $\sigma_\zeta = 0.10$, $\sigma_\omega = 0.20$ and $s = 12$;
- cycle model: $\sigma = 0.10$.

In the case of the cycle model, we verify that we obtain the spectral density function of a pure deterministic cycle with $\lambda^\star = \lambda_c$ when $\rho \to 1$. When ρ is small, the process is not well localized in the frequency domain (see Figure 10.25).

[39] We recall that $g_{(1-\phi L)}(\lambda) = 1 - 2\phi\cos\lambda + \phi^2$.

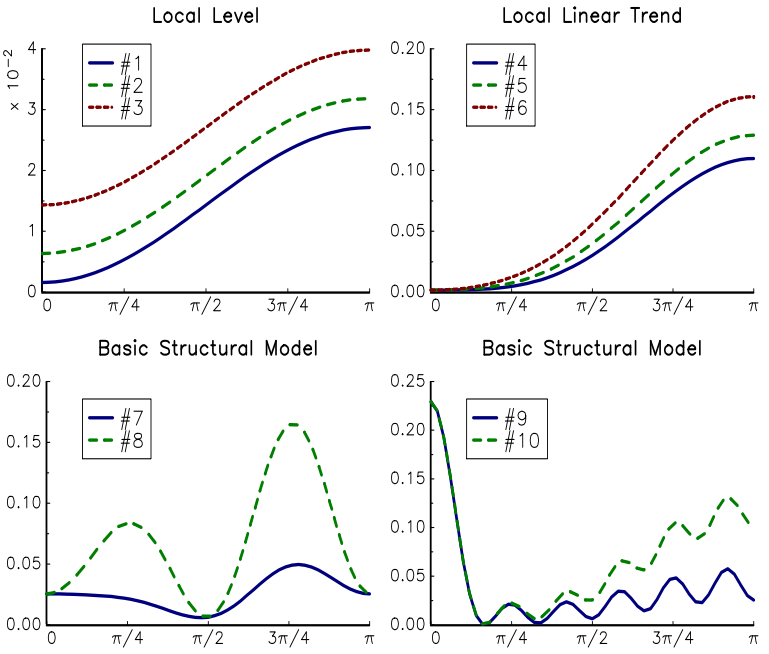


FIGURE 10.24: Spectral density function of LL, LLT and BSM

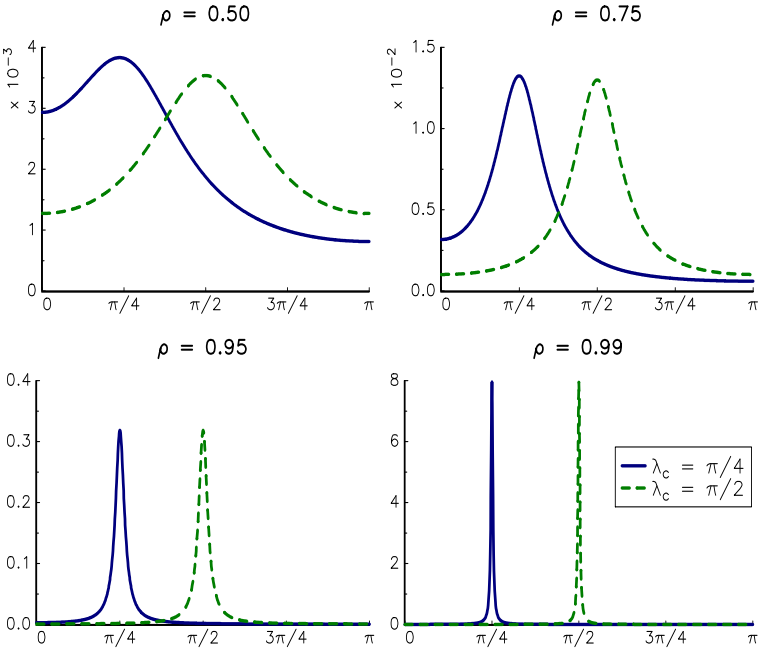


FIGURE 10.25: Spectral density function of the stochastic cycle model

10.2.5.5 Statistical estimation in the frequency domain

The periodogram The periodogram is an estimate of the spectral density function. For that, we define the discrete Fourier transform (DFT) of the time series $\{y_t, t = 1, \ldots, n\}$ as follows[40]:

$$d_y(\lambda_j) = \sum_{t=1}^{n} y_t e^{-i\lambda_j t}$$

where $\lambda_j = 2\pi(j-1)/n$ and $j \in \{1, \ldots, n\}$. The periodogram $I_y(\lambda_j)$ for the frequency λ_j is then equal to:

$$I_y(\lambda_j) = \frac{|d_y(\lambda_j)|^2}{2\pi n} = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{-i\lambda_j t}\right|^2 \tag{10.54}$$

Under some conditions[41], we can show that:

$$\lim_{n\to\infty} \mathbb{E}[I_y(\lambda)] = f_y(\lambda)$$

It follows that $I_y(\lambda)$ is a natural estimator $\hat{f}_y(\lambda)$ of the spectral density function[42]. Indeed, we have:

$$\begin{aligned} d_y(\lambda_j) &= \sum_{t=1}^{n} y_t e^{-i\lambda_j t} \\ &= \sum_{t=1}^{n} y_t \cos(\lambda_j t) - i\sum_{t=1}^{n} y_t \sin(\lambda_j t) \end{aligned}$$

and:

$$\begin{aligned} |d_y(\lambda_j)|^2 &= \left(\sum_{t=1}^{n} y_t \cos(\lambda_j t)\right)^2 + \left(\sum_{t=1}^{n} y_t \sin(\lambda_j t)\right)^2 \\ &= \sum_{s=1}^{n}\sum_{t=1}^{n} y_s y_t \cos(\lambda_j s)\cos(\lambda_j t) + \\ &\quad \sum_{s=1}^{n}\sum_{t=1}^{n} y_s y_t \sin(\lambda_j s)\sin(\lambda_j t) \end{aligned}$$

Since we have $\cos(a-b) = \cos a \cos b + \sin a \sin b$, it follows that:

$$|d_y(\lambda_j)|^2 = \sum_{s=1}^{n}\sum_{t=1}^{n} y_s y_t \cos(\lambda_j(t-s))$$

[40] We define the Fourier transform for some particular frequencies λ_j because we generally use the FFT algorithm to compute it (see Remark 135 on page 684). However, we can also define the Fourier transform for all $\lambda \in [0, 2\pi]$.

[41] In particular, we reiterate that the process must be stationary and centered.

[42] In many textbooks, the normalization factor 2π in the periodogram formula is omitted implying that $\hat{f}_y(\lambda) = (2\pi)^{-1} I_y(\lambda)$. We prefer to adopt the convention to include the scaling factor.

We recall that the empirical covariance function is $\hat{\gamma}_y(k) = n^{-1}\sum_{t=1}^{n} y_{t-k}y_t$. Finally, we obtain:

$$\begin{aligned} I_y(\lambda_j) &= \frac{1}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n} y_s y_t \cos(\lambda_j(t-s)) \\ &= \frac{1}{2\pi n}\sum_{k=-(n-1)}^{n-1}\sum_{t=1}^{n} y_{t-k}y_t \cos(\lambda_j k) \\ &= \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1} \hat{\gamma}_y(k)\cos(\lambda_j k) \end{aligned}$$

This formula is similar to Equation (10.53) where the theoretical autocovariance function is replaced by the empirical autocovariance function and the sum of the infinite series is replaced by the truncated sum. Therefore, the periodogram of the time series y_t contains the same information than the empirical autocovariance function. This is why we can retrieve it using the inverse discrete Fourier transform (DFT):

$$\hat{\gamma}(k) = \frac{1}{n}\sum_{j=1}^{n} I(\lambda_j)\, e^{i\lambda_j k}$$

Remark 135 *In practice, we use the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) to compute $I(\lambda_j)$ and $\hat{\gamma}(k)$. These algorithms assume that the length n of the time series is a power of 2 and take the advantage of many symmetries of cosine and sine functions*[43].

More generally, the asymptotic probability distribution of the periodogram is a chi-squared distribution under some assumptions[44]:

$$\lim_{n\to\infty} 2\frac{I_y(\lambda)}{f_y(\lambda)} \sim \chi^2_2$$

We retrieve the previous result:

$$\lim_{n\to\infty} \mathbb{E}[I_y(\lambda)] = \frac{f_y(\lambda)}{2}\cdot \mathbb{E}\left[\chi^2_2\right] = f_y(\lambda)$$

One of the drawbacks of the estimator (10.54) is its high variance:

$$\lim_{n\to\infty} \operatorname{var}(I_y(\lambda)) = \frac{f_y^2(\lambda)}{4}\cdot \operatorname{var}\left(\chi^2_2\right) = f_y^2(\lambda)$$

In particular, the variance does not go to zero when n tends to ∞. This is why we use in practice the smoothed periodogram defined by:

$$I_y^\star(\lambda_j) = \sum_{s=-m}^{s=m} w_m(s)\cdot I_y(\lambda_j)$$

where $w_m(s)$ is a smoothed window function and m is the bandwidth. The function $w_m(s)$ is equal to $W(s/m)$ where $W(u)$ is a normalized function, which is also called the spectral window function (see Table 10.11).

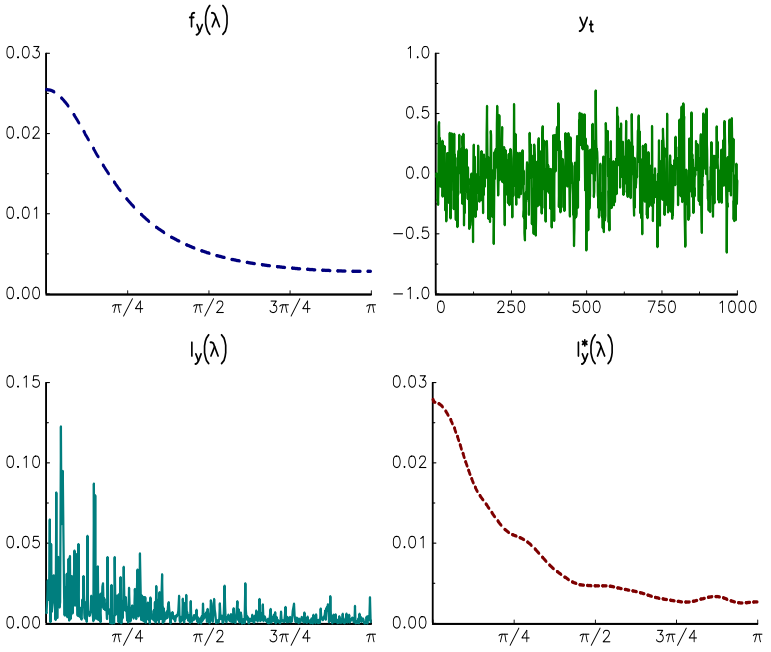


FIGURE 10.26: Estimation of the spectral density function

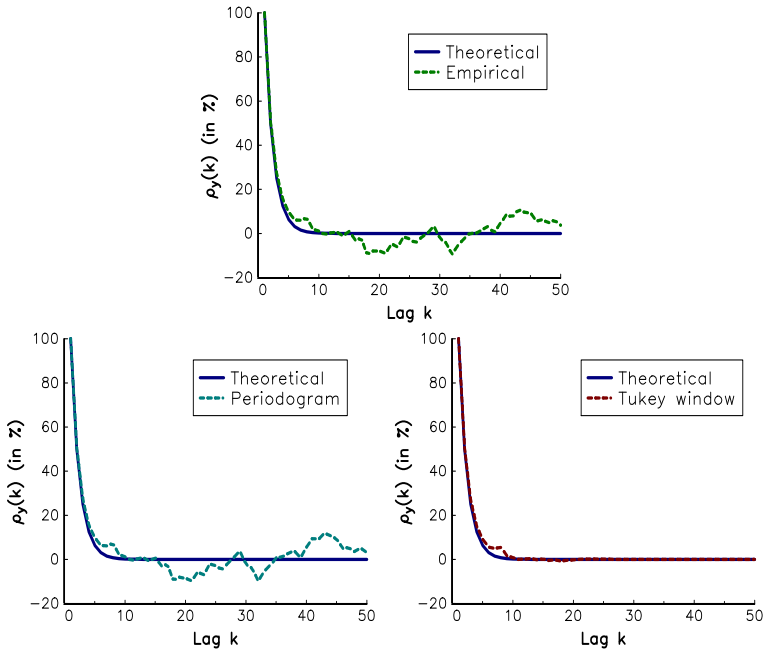


FIGURE 10.27: Estimation of the autocorrelation function

TABLE 10.11: Spectral window functions

Name	$W(u)$
Bartlett	$1-\lvert u\rvert$
Parzen	$\left(1-6\lvert u\rvert^2+6\lvert u\rvert^3\right)\cdot\mathbb{1}\left\{\lvert u\rvert<\frac{1}{2}\right\}+$ $2\left(1-\lvert u\rvert\right)^3\cdot\mathbb{1}\left\{\lvert u\rvert\geq\frac{1}{2}\right\}$
Tukey	$1-2a+2a\cos(\pi u)$
Rectangular	1
Daniell	$(\pi u)^{-1}\sin(\pi u)$
Priestley	$3\left((\pi u)^{-3}\sin(\pi u)-(\pi u)^{-1}\cos(\pi u)\right)$

In Figure 10.26, we consider an AR(1) process y_t with $\phi=0.5$ and $\sigma=20\%$. We represent its spectral density function $f_y(\lambda)$, and then simulate the process with 1 000 observations. We calculate the periodogram, but we notice that $I_y(\lambda)$ is noisy. This is why we estimate the smoothed periodogram $I^\star(\lambda_j)$ with the Tukey window, whose parameters are $a=0.25$ and $m=50$. We obtain a less noisy estimator. In order to illustrate the impact of smoothing, we estimate the autocorrelation function. We first apply the inverse discrete Fourier transform to $I(\lambda_j)$ and $I^\star(\lambda_j)$ in order to obtain the autocovariance functions $\hat{\gamma}_y(k)=n^{-1}\sum_{j=1}^n I_y(\lambda_j)\,e^{i\lambda_j k}$ and $\hat{\gamma}_y^\star(k)=n^{-1}\sum_{j=1}^n I_y^\star(\lambda_j)\,e^{i\lambda_j k}$ and then normalize:

$$\hat{\rho}_y(k)=\frac{\hat{\gamma}_y(k)}{\hat{\gamma}(1)}\quad\text{or}\quad\hat{\rho}_y^\star(k)=\frac{\hat{\gamma}_y^\star(k)}{\hat{\gamma}_y^\star(1)}$$

In Figure 10.27, we compare these functions with the theoretical autocorrelation function $\rho_y(k)=\phi^k$. We also report the empirical autocovariance function calculated directly by the means of convolution. We verify that it is exactly equal to the inverse discrete Fourier transform of the periodogram. However, we notice that the function $\hat{\rho}_y(k)$ does not converge to zero when the lag k is large. This is not the case with the smoothed periodogram, that is less biased in finite samples.

The Whittle estimator Whittle (1953) proposes an original method to estimate the parameters θ of stationary Gaussian models in the frequency domain. Let us consider the process y_t and we note $Y=(y_1,\ldots,y_n)$ the vector of joint observations. Since y_t is centered, we have $Y\sim\mathcal{N}(0,\Sigma)$ where Σ is the Toeplitz covariance matrix. The log-likelihood function is then:

$$\ell(\theta)=-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma|-\frac{1}{2}Y^\top\Sigma^{-1}Y \tag{10.55}$$

Gray (2006) shows that the eigendecomposition $U\Lambda U^*$ of Σ can be approximated by $V\Lambda V^*$ where V is a circulant matrix. In this case, the eigenvalue $\Lambda_{j,j}$ is equal to $2\pi f_y(\lambda_j)$, whereas the eigenvector V_j is related to the Fourier coefficients:

$$V_{t,j}\propto n^{-1/2}e^{-i\lambda_j t}$$

Since V is an unitary matrix, we have:

$$\ln|\Sigma|\simeq\ln|V\Lambda V^*|=\ln\left(|V|\,|\Lambda|\,|V^*|\right)=\ln|\Lambda|=n\ln 2\pi+\sum_{j=1}^n\ln f_y(\lambda_j)$$

[43] If it is not the case, y_t is padded with trailing zeros to length 2^m where m is the nearest integer greater than or equal to $\ln n/\ln 2$.

[44] For example, this result is valid if y_t is a Gaussian process or if the autocovariance function decreases rapidly. If $\lambda=0$, the chi-squared distribution has one degree of freedom (see Exercise 10.3.14 on page 712).

and:

$$\begin{aligned} Y^\top \Sigma^{-1} Y &\simeq Y^\top \left(V \Lambda V^*\right)^{-1} Y \\ &= Z^\top \Lambda^{-1} Z \\ &= \sum_{j=1}^{n} \frac{|z_j|^2}{2\pi f_y(\lambda_j)} \end{aligned}$$

where $Z = (z_1, \ldots, z_n) = VY$. We notice that:

$$\begin{aligned} z_j^2 &= \left| \sum_{t=1}^{n} n^{-1/2} y_t e^{-i\lambda_j t} \right|^2 \\ &= 2\pi I_y(\lambda_j) \end{aligned}$$

We deduce that the log-likelihood function can be written as:

$$\ell(\theta) \simeq -n \ln 2\pi - \frac{1}{2} \sum_{j=1}^{n} \ln f_y(\lambda_j) - \frac{1}{2} \sum_{j=1}^{n} \frac{I_y(\lambda_j)}{f_y(\lambda_j)} \tag{10.56}$$

There is a fundamental difference between time domain maximum likelihood (TDML) and frequency domain maximum likelihood (FDML). Indeed, in the time domain, we can define the log-likelihood for a given observation date t. In the frequency domain, defining the log-likelihood for a given frequency λ_j does not make sense.

In practice, we may observe a significant difference between the values given by Equations (10.55) and (10.56). Nevertheless, TDML and FDML estimators are generally very close. For example, we consider the AR(1) process $y_t = \phi y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$, $\phi = 0.5$ and $\sigma = 20\%$. In Figure 10.28, we report the probability density function of the two estimators $\hat{\phi}_{\text{TDML}}$ and $\hat{\phi}_{\text{FDML}}$ when the sample is equal to 300. We verify that they are very close.

10.2.5.6 Extension to multidimensional processes

The previous results can be generalized to the multivariate case. Let us now consider the m-dimensional time series $y_t = (y_{t,1}, \ldots, y_{t,m})$. Since y_t is centered, the autocovariance matrix is defined as $\Gamma_y(k) = \mathbb{E}\left[y y_{t-k}^\top\right]$. We notice that $\Gamma_y(k)$ is not necessarily a symmetric matrix[45], but we have $\Gamma_y(k)^\top = \Gamma_y(-k)$. The $m \times m$ spectral matrix $f_y(\lambda)$ is then defined by:

$$f_y(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Gamma_y(k) e^{-i\lambda k} \tag{10.57}$$

It follows that $f_y(\lambda)$ is an Hermitian matrix. Moreover, the diagonal elements are real, but the off-diagonal elements are complex. Similarly, the multivariate periodogram is equal to:

$$I_y(\lambda_j) = \frac{d_y(\lambda_j)\, d_y(\lambda_j)^*}{2\pi n} \tag{10.58}$$

where $\lambda_j = 2\pi(j-1)/n$, $j = \{1, \ldots, n\}$ and $d_y(\lambda_j)$ is the multidimensional discrete Fourier transform[46] of the time series y_t:

$$d_y(\lambda_j) = \sum_{t=1}^{n} y_t e^{-i\lambda_j k}$$

[45] Because we generally have $\mathbb{E}[x_t y_{t-k}] \neq \mathbb{E}[y_t x_{t-k}]$ for two unidimensional time series x_t and y_t.

[46] $d_y(\lambda_j)$ is a vector of dimension m.

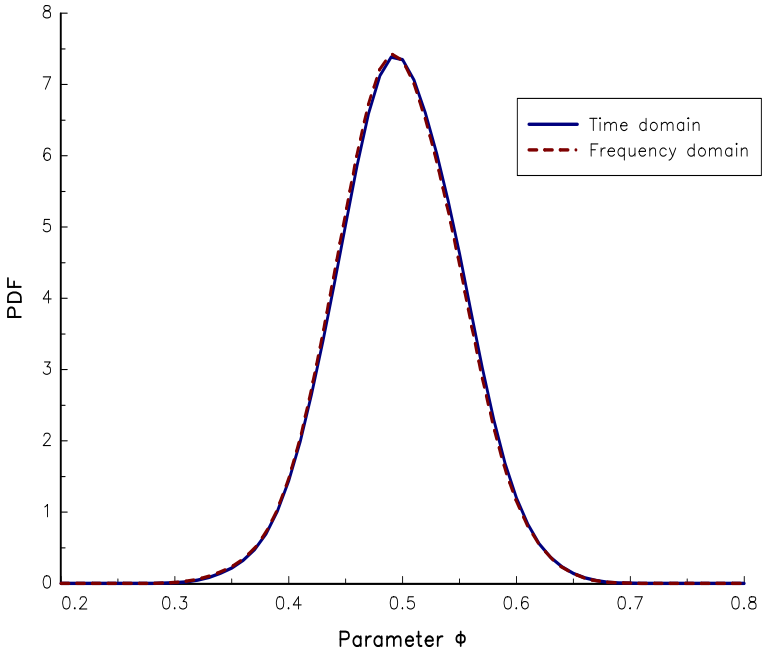


FIGURE 10.28: PDF of TDML and FDML estimators

The main properties obtained in the case $m = 1$ are also valid in the case $m > 1$. For instance, we have:

1. the spectral matrix of the multidimensional white noise process $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$ is equal to $f_\eta(\lambda) = (2\pi)^{-1} \Sigma$;
2. $\lim_{n\to\infty} \mathbb{E}[I_y(\lambda)] = f_y(\lambda)$;
3. $\lim_{n\to\infty} \operatorname{var}(I_y(\lambda)) = f_y(\lambda) \odot f_y(\lambda)^*$;
4. if x_t and y_t are two independent multidimensional stochastic processes, then $f_{x+y}(\lambda) = f_x(\lambda) + f_y(\lambda)$;
5. if $z_t = Ay_t$ and A is a real matrix, then $f_z(\lambda) = Af_y(\lambda) A^\top$;
6. the spectral density function of the linear filter $y_t = \Psi(L) x_t$ is given by:

$$f_y(\lambda) = \Psi\left(e^{-i\lambda}\right) f_x(\lambda) \Psi\left(e^{-i\lambda}\right)^*$$

The parameters θ of a stationary centered Gaussian model can also be estimated using the FDML method with the following Whittle log-likelihood function:

$$\ell(\theta) \propto -\frac{1}{2}\sum_{j=0}^{n} \ln|f_y(\lambda_j)| - \frac{1}{2}\sum_{j=1}^{n} \operatorname{trace}\left(f_y(\lambda_j)^{-1} I_z(\lambda_j)\right) \tag{10.59}$$

These models can generally be written as a stationary state space model:

$$\begin{cases} y_t = Z\alpha_t + \epsilon_t \\ \alpha_t = T\alpha_{t-1} + R\eta_t \end{cases}$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, H)$ and $\eta_t \sim \mathcal{N}(\mathbf{0}, Q)$. It follows that:

$$y_t = Z(I - TL)^{-1} R\eta_t + \epsilon_t$$

where L is the lag operator and I is the identity matrix, whose dimension is equal to the size of the state vector α_t. Using the above properties, we deduce that the spectral density function of y_t is:

$$\begin{aligned} f_y(\lambda) &= Z\left(I - Te^{-i\lambda}\right)^{-1} R f_\eta(\lambda) R^\top \left(\left(I - Te^{-i\lambda}\right)^{-1}\right)^* Z^\top + f_\epsilon(\lambda) \\ &= \frac{Z\left(I - Te^{-i\lambda}\right)^{-1} RQR^\top \left(\left(I - Te^{-i\lambda}\right)^{-1}\right)^* Z^\top + H}{2\pi} \end{aligned} \tag{10.60}$$

For instance, Roncalli (1996) extensively used the Whittle method to estimate VARMA, SSM and complex Gaussian models.

We now consider the special case $m = 2$ and we note $z_t = (x_t, y_t)$ the bivariate process. The autocovariance matrix $\Gamma_z(k)$ becomes:

$$\Gamma_z(k) = \begin{pmatrix} \gamma_x(k) & \gamma_{x,y}(k) \\ \gamma_{y,x}(k) & \gamma_y(k) \end{pmatrix}$$

where $\gamma_{x,y}(k)$ is the autocovariance function between x_t and y_t. For the spectral matrix $f_z(\lambda)$, we have:

$$f_z(\lambda) = \begin{pmatrix} f_x(\lambda) & f_{x,y}(\lambda) \\ f_{y,x}(\lambda) & f_y(\lambda) \end{pmatrix}$$

where $f_x(\lambda)$ and $f_y(\lambda)$ are the spectral density functions of the stochastic processes x_t and y_t, and $f_{x,y}(\lambda)$ is the cross spectrum:

$$f_{x,y}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{x,y}(k) e^{-i\lambda k}$$

Similarly, the bivariate periodogram of the process z_t takes the following form:

$$I_z(\lambda_j) = \begin{pmatrix} I_x(\lambda_j) & I_{x,y}(\lambda_j) \\ I_{y,x}(\lambda_j \lambda) & I_y(\lambda_j \lambda) \end{pmatrix}$$

where $I_x(\lambda_j)$ and $I_y(\lambda_j)$ are the periodograms of the stochastic processes x_t and y_t, and $I_{x,y}(\lambda_j)$ is the cross periodogram:

$$I_{x,y}(\lambda_j) = \frac{d_x(\lambda_j) d_y(\lambda_j)^*}{2\pi n}$$

Let us consider the bivariate process $z_t = (x_t, y_t)$, which has the following SSM form:

$$\begin{cases} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} + \epsilon_t \\ \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} = \begin{pmatrix} 0.9 & 0 & 0 \\ 0.5 & -0.1 & 0 \\ 0 & 0 & 0.3 \end{pmatrix} \begin{pmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \alpha_{3,t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \eta_t \end{cases}$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \text{diag}(0.5, 0.2))$ and $\eta_t \sim \mathcal{N}(\mathbf{0}, \text{diag}(1.0, 1.0))$. We also have:

$$\begin{cases} x_t = \alpha_{1,t} + \epsilon_{1,t} \\ y_t = \alpha_{2,t} + \epsilon_{2,t} \end{cases}$$

and:

$$\begin{cases} \alpha_{1,t} = 0.9 \cdot \alpha_{1,t-1} + \eta_{1,t} \\ \alpha_{2,t} = 0.5 \cdot \alpha_{1,t-1} - 0.1 \cdot \alpha_{2,t-1} + \eta_{2,t} \\ \alpha_{3,t} = 0.3 \cdot \alpha_{3,t-1} - \eta_{2,t} \end{cases}$$

Therefore, the bivariate process (x_t, y_t) can be viewed as a noisy restricted VAR(1) model where the residuals of the second VAR component follow a MA(1) process[47]. Below, we give the values of $f_z(\lambda)$ calculated with Equation (10.60):

λ	$f_x(\lambda)$	$f_y(\lambda)$	$f_{x,y}(\lambda)$	$f_{y,x}(\lambda)$
0	15.995	3.452	7.234	7.234
0.5	0.770	0.312	$0.285 + 0.140i$	$0.285 - 0.140i$
$pi/4$	0.376	0.234	$0.104 + 0.091i$	$0.104 - 0.091i$
$pi/2$	0.168	0.211	$0.004 + 0.044i$	$0.004 - 0.044i$
pi	0.124	0.242	-0.024	-0.024

We verify that $f_x(\lambda)$ and $f_y(\lambda)$ are real numbers, $f_{x,y}(\lambda)$ and $f_{y,x}(\lambda)$ are complex numbers, and $f_{y,x}(\lambda)$ is the complex conjugate of $f_{x,y}(\lambda)$.

In order to interpret the cross spectrum, we can write $f_{x,y}(\lambda)$ in the complex form:

$$f_{x,y}(\lambda) = \text{cs}_{x,y}(\lambda) + i\,\text{qs}_{x,y}(\lambda)$$

where $\text{cs}_{x,y}(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_{x,y}(k) \cos(\lambda k)$ is the cospectrum and $\text{qs}_{x,y}(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_{x,y}(k) \sin(-\lambda k)$ is the quadrature spectrum. The cospectrum is the simultaneous covariance between x_t and y_t at frequency λ, whereas the quadrature spectrum is the lagged covariance[48] by the phase $\pi/2$. Alternatively, we can write $f_{x,y}(\lambda)$ in the polar form:

$$f_{x,y}(\lambda) = r_{x,y}(\lambda)\, e^{i\theta_{x,y}(\lambda)}$$

where $r_{x,y}(\lambda)$ is the gain and $\theta_{x,y}(\lambda)$ is the phase spectrum (Engle, 1976). We have:

$$\begin{aligned} r_{x,y}^2(\lambda) &= |f_{x,y}(\lambda)|^2 \\ &= \text{cs}_{x,y}^2(\lambda) + \text{qs}_{x,y}^2(\lambda) \end{aligned}$$

It follows that the squared gain function $r_{x,y}^2(\lambda)$ is a dependence measure between x_t and y_t. On the contrary, $\theta_{x,y}(\lambda)$ determine the lead-lag relationship between x_t and y_t.

We consider the linear filtering model:

$$\begin{aligned} y_t &= \sum_{k=-\infty}^{\infty} \varphi_k L^k x_t + \varepsilon_t \\ &= \varphi(L)\, x_t + \varepsilon_t \end{aligned}$$

[47] We have $\eta_{2,t} = \alpha_{3,t} - 0.3\alpha_{3,t-1}$.

[48] Because we have:

$$\text{qs}_{x,y}(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_{x,y}(k) \cos\left(\frac{\pi}{2} + \lambda k\right)$$

where x_t and ε_t are two independent stochastic processes and $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. We have $\gamma_{y,x}(k) = \sum_{s=-\infty}^{\infty} \varphi_s \gamma_x(k-s)$ and:

$$\begin{aligned} f_{y,x}(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \varphi_s \gamma_x(k-s) e^{-i\lambda k} \\ &= \sum_{s=-\infty}^{\infty} \varphi_s e^{-i\lambda s} \left(\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_x(k-s) e^{-i\lambda(k-s)} \right) \\ &= \sum_{s=-\infty}^{\infty} \varphi_s e^{-i\lambda s} f_x(\lambda) \\ &= \varphi\left(e^{-i\lambda}\right) f_x(\lambda) \end{aligned}$$

We have seen that $f_y(\lambda) = \left|\varphi\left(e^{-i\lambda}\right)\right|^2 f_x(\lambda) + f_\varepsilon(\lambda)$. It follows that:

$$\begin{aligned} f_y(\lambda) &= \left|\varphi\left(e^{-i\lambda}\right)\right|^2 f_x(\lambda) + f_\varepsilon(\lambda) \\ &= \left|\frac{f_{y,x}(\lambda)}{f_x(\lambda)}\right|^2 f_x(\lambda) + f_\varepsilon(\lambda) \\ &= \frac{r_{y,x}^2(\lambda)}{f_x(\lambda)} + f_\varepsilon(\lambda) \\ &= f_{y|x}(\lambda) + f_\varepsilon(\lambda) \end{aligned} \tag{10.61}$$

We have decompose the spectral density function of y_t into two terms: the first term $f_{y|x}(\lambda)$ can be seen as the conditional expectation of $f_y(\lambda)$ with respect to $f_x(\lambda)$ while the second term $f_\varepsilon(\lambda)$ is the component due to the noise process. Equation (10.61) is close to the Gaussian conditional expectation formula or the linear regression of y_t on x_t. The fraction of the variance of $f_y(\lambda)$ explained by the linear filter – or the coefficient of determination R^2 – is equal to:

$$\begin{aligned} R^2 &= \frac{f_{y|x}(\lambda)}{f_y(\lambda)} \\ &= \frac{|f_{y,x}(\lambda)|^2}{f_y(\lambda) f_x(\lambda)} \\ &= c_{y,x}^2(\lambda) \end{aligned}$$

$c_{y,x}(\lambda)$ is called the coherency function. If the two processes x_t and y_t are uncorrelated, then $f_{y,x}(\lambda) = 0$ and $c_{y,x}^2(\lambda) = 0$. If $\sigma_\varepsilon = 0$, then $f_\varepsilon(\lambda) = 0$, $|f_{y,x}(\lambda)|^2 = f_y(\lambda) f_x(\lambda)$ and $c_{y,x}^2(\lambda) = 1$. We deduce that $0 \leq R^2 \leq 1$ and $0 \leq c_{y,x}^2(\lambda) \leq 1$. On page 610, we have seen that the coefficient of determination R^2 in the time domain associated to the linear regression $y_t = \beta_0 + \beta x_t + u_t$ is equal to the square of the cross-correlation $\rho_{y,x}^2$ between x_t and y_t. By analogy, the coherence function $c_{y,x}(\lambda)$ can be viewed as the cross-correlation between x_t and y_t in the frequency domain (Engle, 1976). Moreover, we have:

$$\begin{aligned} c_{y,x}(\lambda) &= \frac{f_{x,y}(\lambda)}{\sqrt{f_y(\lambda) f_x(\lambda)}} \\ &= \frac{r_{x,y}(\lambda)}{\sqrt{f_y(\lambda) f_x(\lambda)}} e^{i\theta_{x,y}(\lambda)} \\ &= \tilde{r}_{x,y}(\lambda) e^{i\theta_{x,y}(\lambda)} \end{aligned}$$

Since $c_{y,x}(\lambda)$ is a complex function, it is less easy to manipulate than a correlation function. This is why $\tilde{r}_{x,y}(\lambda)$ may be preferred to define the cross-correlation between x_t and y_t in the frequency domain. For example, we consider the bivariate process:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.5 & \varphi_{x,y} \\ \varphi_{y,x} & 0.5 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

where ε_t and η_t are two uncorrelated white noise processes with same variance. In Figure 10.29, we represent $c_{y,x}(\lambda)$ in polar coordinates.

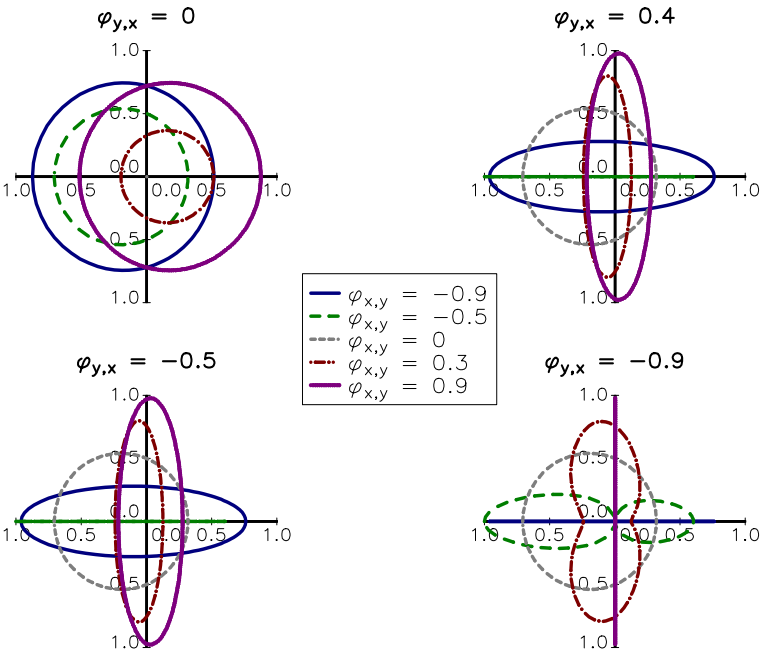


FIGURE 10.29: Coherency function $c_{y,x}(\lambda)$

10.2.5.7 Some applications

White noise testing There are many spectral procedures for testing the white noise hypothesis. For instance, we have shown that the asymptotic distribution of $2f_y(\lambda)^{-1} I_y(\lambda)$ is a chi-squared distribution χ^2_2. The hypothesis $\mathcal{H}_0 : y_t \sim \mathcal{N}(0, \sigma^2)$ implies that $4\pi\sigma^{-2} I_y(\lambda) \sim \chi^2_2$. Therefore, it suffices to estimate the empirical volatility and use goodness of fit tests like the Kolmogorov-Smirnov statistic or a QQ plot (Pawitan and O'Sullivan, 1994) for testing the null hypothesis:

$$\mathcal{H}_0 : \frac{4\pi}{\hat{\sigma}^2} I_y(\lambda) \sim \chi^2_2$$

Another idea is to verify that the periodogram does not contain a value significantly larger than the other values. Since the cumulative distribution function of χ^2_2 is $\mathbf{F}(x) = 1 - e^{-x^2/2}$, we deduce that:

$$\Pr\left\{ \frac{4\pi}{\hat{\sigma}^2} I_y(\lambda) \leq x \right\} = \mathbf{F}(x) = 1 - e^{-x^2/2}$$

Let I_y^+ be the maximum periodogram ordinate:

$$I_y^+ = \sup\{ I_y(\lambda_j) : j = 1, \ldots, q \}$$

where $q = \lfloor n/2 \rfloor$ is the largest integer less than $n/2$. We note:

$$\xi = \frac{4\pi}{\hat{\sigma}^2} I_y^+$$

If $j \neq j'$, then $I_y(\lambda_j)$ and $I_y(\lambda_{j'})$ are independent and we have:

$$\Pr\{\xi \leq x\} = \mathbf{F}(x)^q = \left(1 - e^{-x^2/2}\right)^q$$

Rejecting the null hypothesis at the confidence level α is equivalent to verify that ξ is too large or $\mathbf{F}(\xi)^q \geq \alpha$. Therefore, we deduce that $\mathcal{H}_0$ is rejected if the following condition is satisfied:

$$\xi > \sqrt{\ln \frac{1}{\left(1 - \alpha^{1/q}\right)^2}}$$

Remark 136 *We can also test the presence of a unit root by considering the null hypothesis* $\mathcal{H}_0 : \mathcal{S}(x_t) = y_t - y_{t-1} \sim \mathcal{N}(0, \sigma^2)$.

Cycle identification Cycle testing can be viewed as the contrary of white noise testing. Indeed, if the process y_t contains a cycle at frequency λ_c, we must observe a peak in the periodogram. Fisher (1929) defined the g statistic:

$$g = \frac{I_y(\lambda_c)}{\sum_{j=1}^q I_y(\lambda_j)}$$

where $q = \lfloor n/2 \rfloor$ is the largest integer less than $n/2$. Under the hypothesis that the process y_t contains a cycle at frequency λ_c, Fisher showed that the distribution function of g satisfies[49]:

$$\Pr\{g \geq x\} = 1 - \sum_{j=1}^{q} (-1)^j \binom{q}{j} (1 - jx)_+^{q-1}$$

If this probability is lower than the level α, then we accept the presence of a cycle. In practice, the frequency λ_c is unknown and is estimated by the largest periodogram value:

$$\lambda_c = \{\lambda : \sup I_y(\lambda_j)\}$$

We consider the famous example of Canadian lynx data set, which collects the annual record of the number of the Canadian lynx '*trapped*' in the Mackenzie River district of the North-West Canada for the period 1821–1934 (Tong, 1990). In the first panel in Figure 10.30, we represent the time series $y_t = \log x_t$ where x_t is the number of lynx. We also report the periodogram $I_y(\lambda)$ in the second panel and we observe a peak at the frequency $\lambda_c = 0.6614$. The Fisher test is equal to $g = 0.59674$ whereas the p-value is close to zero. We deduce that the cycle period is equal to $p = 2\pi/\lambda_c = 9.5$ years. In the third and fourth panels, we have represented the cycle c_t and the residuals $\varepsilon_t = y_t - c_t$. Finally, we obtain the following model:

$$\begin{aligned} y_t &= c_t + \varepsilon_t \\ &= 2.904 + 0.607 \cdot \sin\left(\frac{2\pi}{9.5} \cdot t - 1.138\right) + \varepsilon_t \end{aligned}$$

[49] When the number of observations n is large, Priestley (1981) showed that the probability distribution of the statistic $g^\star = 2ng$ is $\Pr\{g^\star \geq x\} = ne^{-x/2}$.

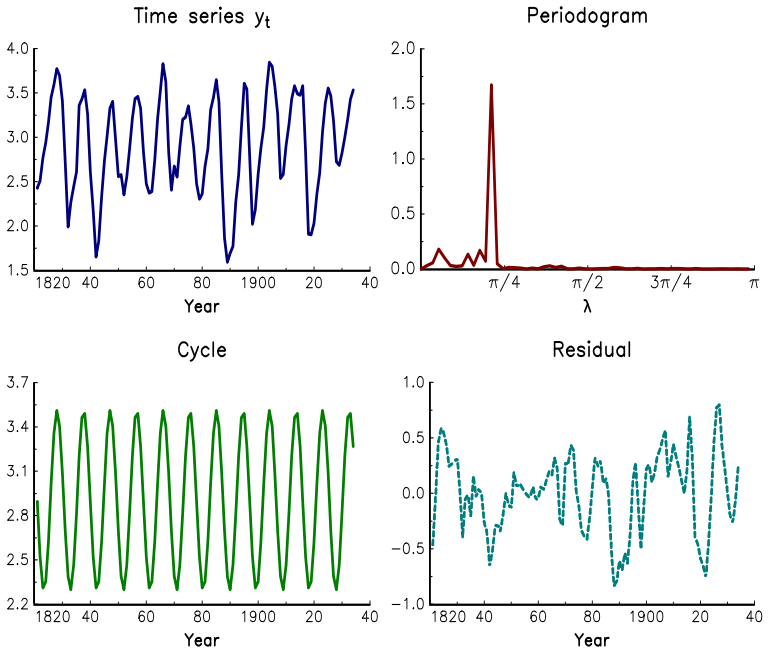


FIGURE 10.30: Detection of the cycle in the Canadian lynx data set

Long memory time series and fractional processes The fractional white noise (FWN) process is defined as:

$$(1-L)^d y_t = \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$ and[50] $|d| < 1/2$. Granger and Joyeux (1980) and Hosking (1981) showed that y_t has an infinite moving average process:

$$y_t = (1-L)^{-d} \varepsilon_t = \sum_{k=0}^{\infty} \theta_k \varepsilon_{t-k}$$

where[51]:

$$\theta_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$$

We can also write y_t as an infinite auto-regressive process:

$$\sum_{k=0}^{\infty} \phi_k y_{t-k} = \varepsilon_t$$

where:

$$\phi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

In Figure 10.31, we report the function ϕ_k for several values of d. If $d < 0$ (respectively

[50] This condition implies that y_t is stationary.

[51] If $\alpha > 0$, the gamma function $\Gamma(\alpha)$ is equal to $\int_0^\infty t^{\alpha-1} e^{-t}\, \mathrm{d}t$. We also have $\Gamma(0) = \infty$ and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. This last property is used to calculate $\Gamma(\alpha)$ when α is negative. For instance, $\Gamma(-0.5) = -2\Gamma(0.5) \simeq 7.0898$.

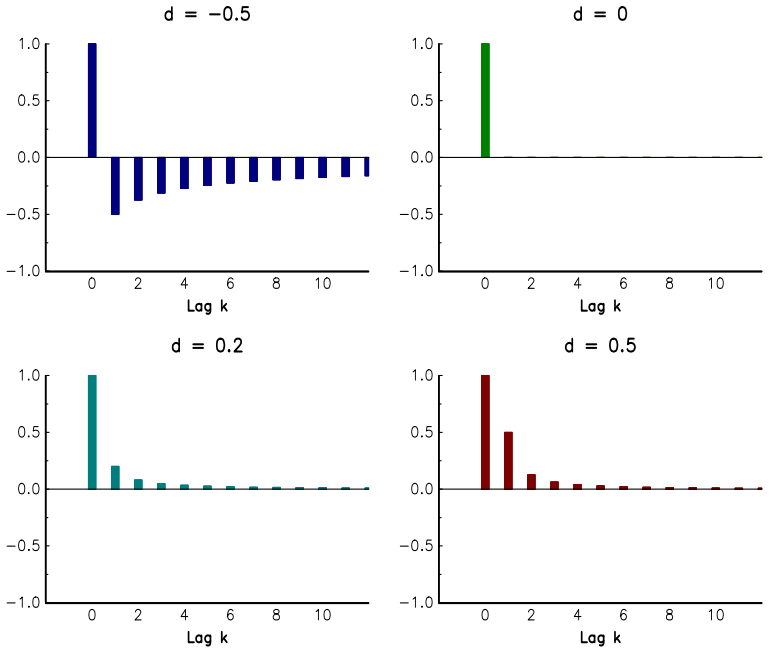


FIGURE 10.31: AR representation of the fractional process

$d > 0$), then ϕ_k is negative (respectively positive) for $k \geq 1$. Hosking (1981) showed that[52]:

$$\gamma_y(k) = \sigma^2 \frac{\Gamma(1-2d)\,\Gamma(k+d)}{\Gamma(d)\,\Gamma(1-d)\,\Gamma(k-d+1)}$$

and:

$$\rho_y(k) = \frac{\Gamma(1-d)\,\Gamma(k+d)}{\Gamma(d)\,\Gamma(k-d+1)}$$

We verify that $\rho_y(k) < 0$ if $d < 0$ and $\rho_y(k)$ if $d > 0$. In the case of the AR(1) process $y_t = \phi y_{t-1} + \varepsilon_t$, we recall that the autocorrelation function is $\rho_y(k) = \phi^k$. When $d < 0$, there is a big difference between FWN and AR(1) processes, because $\rho_y(k)$ is an oscillating function in the case of the AR(1) process. When $d > 0$, the tappering pattern is more pronounced for the AR(1) than for the FWN process. In order to understand this difference, we give here the asymptotic behavior of the coefficients (Baillie, 1996):

$$\begin{aligned}
\lim_{k\to\infty} \theta_k &= \frac{1}{\Gamma(d)} k^{d-1} \\
\lim_{k\to\infty} \phi_k &= \frac{1}{\Gamma(-d)} k^{-(d+1)} \\
\lim_{k\to\infty} \gamma_y(k) &= \frac{\Gamma(1-2d)}{\Gamma(d)\,\Gamma(1-d)} k^{2d-1} \\
\lim_{k\to\infty} \rho_y(k) &= \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}
\end{aligned}$$

[52]The variance of y_t is infinite when the fractional differencing parameter d is equal to 1/2.

We deduce that these coefficients decline at a slower rate than for the AR(1) process[53]. For instant, if $d = 1/2$, then $\rho_y(k) = 1$. In Table 10.12, we calculate $\rho_y(k)$ for different values of d. When we compare these results with the ones obtained for an AR(1) process with parameter ϕ, we observe the hyperbolic decay for the FWN process. For instance, for the FWN process with $d = 0.45$, $\rho_y(10)$ and $\rho_y(10000)$ are respectively equal to 99.15% and 97.79%. For the AR(1) process with $\phi = 0.999$, we obtain $\rho_y(10) = 99\%$ but[54] $\rho_y(10000) = 0\%$. We also obtain the following paradox: the autocorrelation function $\rho_y(10000)$ is higher for the FWN process with $d = 0.20$ than for the AR(1) process with $\phi = 0.999$!

TABLE 10.12: Autocorrelation function $\rho_y(k)$ (in %) of FWN and AR(1) processes

k	FWN process			AR(1) process		
d / ϕ	0.2	0.45	0.499	0.50	0.90	0.999
0	100.00	100.00	100.00	100.00	100.00	100.00
1	25.00	81.82	99.60	50.00	90.00	99.90
5	9.65	69.90	99.29	3.13	59.05	99.50
10	6.37	65.22	99.15	0.10	34.87	99.00
100	1.60	51.81	98.69	0.00	0.00	90.48
500	0.61	44.11	98.38	0.00	0.00	60.64
1000	0.40	41.15	98.24	0.00	0.00	36.77
5000	0.15	35.04	97.93	0.00	0.00	0.67
10000	0.10	32.69	97.79	0.00	0.00	0.00

Since $y_t = (1-L)^{-d}\varepsilon_t$ is a stationary process, we deduce that its spectral density function is:

$$\begin{aligned} f_y(\lambda) &= \left|(1-L)^{-d}\right|^2 f_\varepsilon(\lambda) \\ &= \left|1-e^{-i\lambda}\right|^{-2d}\frac{\sigma^2}{2\pi} \\ &= \frac{\sigma^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d} \end{aligned} \tag{10.62}$$

In Figure 10.32, we represent the function $f_y(\lambda)$ for different values of d, and we compare $f_y(\lambda)$ with the spectral density function obtained for the AR(1) process. We notice that the high frequency components dominate when d is negative, whereas the spectral density function is concentrated at low frequencies when d is positive. In particular, we have $f_y(0) = +\infty$ when $d > 0$.

Granger and Joyeux (1980) extended the FWN process to a more larger class of stationary processes called autoregressive fractionally integrated moving average (or ARFIMA) models. The ARFIMA(p,d,q) is defined by:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$. This model can be seen as an extension of ARMA models by introducing a fractional unit root. This is why y_t is said to be fractionally integrated and we note $y_t \sim I(d)$ where $|d| < 1/2$. The properties of the FWN process can be generalized

[53]Fractional white noise processes are also called hyperbolic decay time series, whereas autoregressive processes are called geometric decay time series.

[54]In fact, the non-rounding autocorrelation is equal to $\rho_y(10000) = 0.0045\%$.

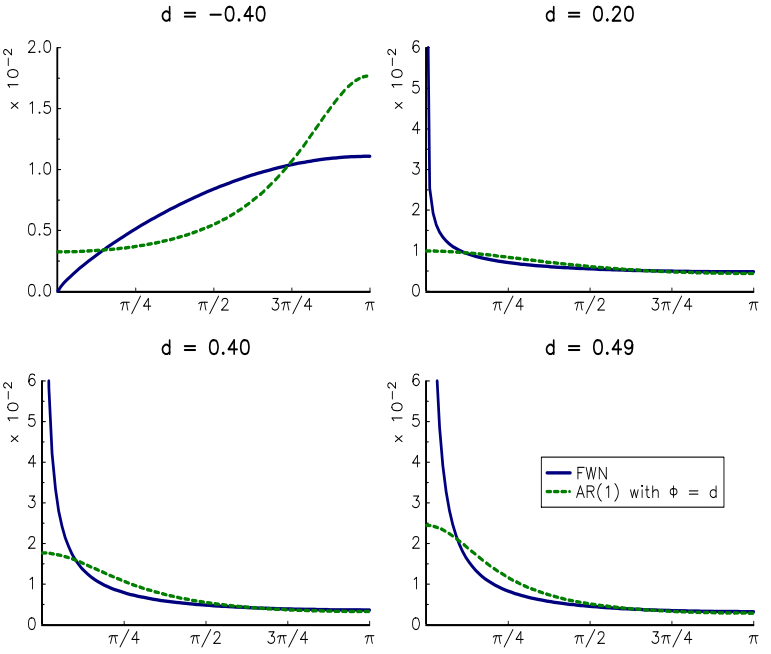


FIGURE 10.32: Spectral density function of the FWN process

to the ARFIMA process. For instance, the stationary ARFIMA(p,d,q) process possesses long memory if $d > 0$ and short memory if $d < 0$. The AR(∞) and MA(∞) representation are $\Phi^{\star}(L) y_t = \varepsilon_t$ and $y_t = \Theta^{\star}(L) \varepsilon_t$ where $\Phi^{\star}(L) = \Phi(L)(1-L)^d \Theta(L)^{-1}$ and $\Theta^{\star}(L) = \Phi(L)^{-1}(1-L)^{-d}\Theta(L)$. The coefficients of $\Phi^{\star}(L)$ and $\Theta^{\star}(L)$ can be calculated numerically using truncated convolutions. Sowell (1992) also provides exact formulations of the autocovariance function. Concerning the spectral density function, we obtain:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d} \frac{\left|\Theta\left(e^{-i\lambda}\right)\right|^2}{\left|\Phi\left(e^{-i\lambda}\right)\right|^2}$$

If we consider the ARFIMA(1,d,1) model defined by $(1-\phi L)(1-L)^d y_t = (1-\theta L)\varepsilon_t$, we obtain:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d} \frac{\left(1-2\theta\cos\lambda+\theta^2\right)}{\left(1-2\phi\cos\lambda+\phi^2\right)}$$

We consider the ARMA processes represented in Figure 10.23 by adding a long memory component. Results are given in Figure 10.33. We notice how the function $\left(2\sin\frac{\lambda}{2}\right)^{-2d}$ impacts the white noise process. When $d < 0$ (respectively $d > 0$), the spectral density function becomes an increasing (respectively decreasing) function. This function is equal to 1 for $\lambda^{\star} = 2\arcsin 1/2 = 1.047$ radians. When $d < 0$, the short memory part reduces low frequency components and magnifies high frequency components. We observe the contrary when $d > 0$. This is coherent with the previous analysis since the ARFIMA process is persistent when $d > 0$ while it is mean-reverting when $d < 0$.

There are different approaches for estimating the ARFIMA(p,d,q) model. Sowell (1992) derives the exact ML estimator by considering the joint distribution of the sample $Y =$

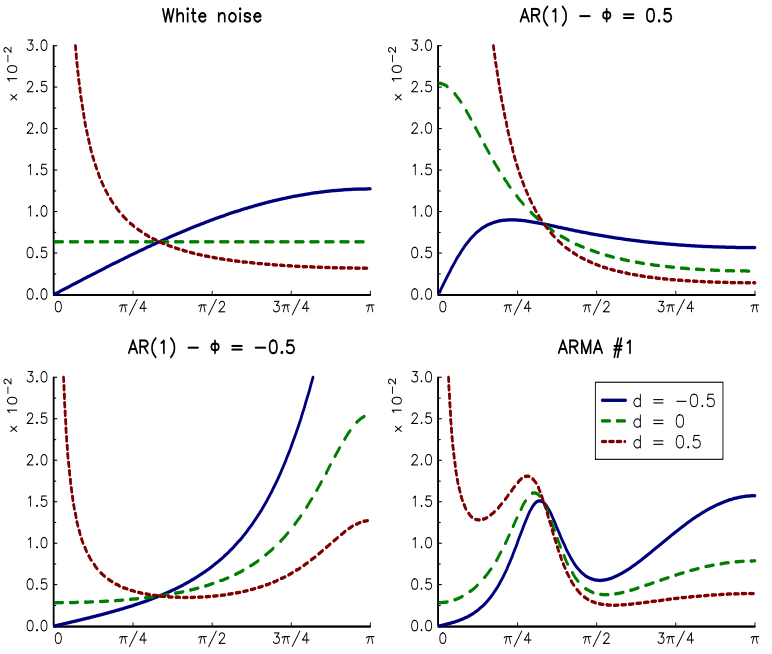


FIGURE 10.33: Spectral density function of the ARFIMA process

$(y_1, \ldots, y_n)$. We recall that the log-likelihood function is:

$$\ell(\theta) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}Y^\top \Sigma Y$$

where Σ is the covariance matrix of Y:

$$\Sigma = \begin{pmatrix} \gamma_y(0) & \gamma_y(1) & & \gamma_y(n-1) \\ \gamma_y(1) & \gamma_y(0) & & \gamma_y(n-2) \\ & & \ddots & \\ \gamma_y(n-1) & \gamma_y(n-2) & & \gamma_y(0) \end{pmatrix}$$

However, this approach may be time-consuming because it requires the inverse of the $n \times n$ matrix Σ, where the elements $\Sigma_{i,j} = \gamma_y(|i-j|)$ are themselves complicated to calculate[55]. This is why it is better to estimate the vector θ of parameters by considering the FDML approach. The whittle log-likelihood is:

$$\begin{aligned} \ell(\theta) \quad \propto \quad & -\frac{n}{2}\ln\sigma^2 + \frac{d}{2}\sum_{j=1}^{n}\ln\left(4\sin^2\frac{\lambda}{2}\right) - \sum_{j=1}^{n}\ln\frac{\left|\Theta\left(e^{-i\lambda}\right)\right|}{\left|\Phi\left(e^{-i\lambda}\right)\right|} - \\ & \frac{1}{2n}\sum_{j=1}^{n}\frac{\left|\sum_{t=1}^{n} y_t e^{-i\lambda_j t}\right|^2 \left|\Phi\left(e^{-i\lambda}\right)\right|^2}{\sigma^2\left(2\sin\frac{\lambda}{2}\right)^{-2d}\left|\Theta\left(e^{-i\lambda}\right)\right|^2} \end{aligned}$$

Another famous approach is the semiparametric estimation proposed by Geweke and Porter-Hudak (1983). When $d > 0$, we have seen that the low frequency part dominates the spectrum:

$$f_y(\lambda) \simeq \frac{\sigma^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d}$$

[55] Sowell (1992) shows that they are function of the hypergeometric function.

when $\lambda \to 0$. We deduce that:

$$\ln f_y(\lambda) \simeq \left(\ln \sigma^2 - \ln 2\pi\right) - d \ln \left(4 \sin^2 \frac{\lambda}{2}\right)$$

Geweke and Porter-Hudak (1983) estimate the parameters d and σ by considering the following linear regression:

$$\ln I_y(\lambda_j) = c - d \ln \left(4 \sin^2 \frac{\lambda_j}{2}\right) + u_j$$

where[56] $\lambda \leq \lambda_{\min}$.

The fractional white noise process is related to the Hurst exponent H, which can be characterized in several ways:

1. let $c > 0$ be a scalar; the probability distribution of y_t at the date ct is equal to the probability distribution of y_t at the time t multiplied by c^H:

   $$y_{ct} \overset{\mathcal{D}}{\sim} c^H y_t$$

   we say that the process y_t is H self-similar or has the self-similarity property;

2. if we consider the asymptotic distribution of $\rho_y(k)$, we verify that:

   $$\rho_y(k) \sim |k|^{2(H-1)}$$

3. for the spectral generating function, we obtain[57]:

   $$f_y(\lambda) \sim |\lambda|^{-(2H+1)}$$

4. the Hurst exponent is related to the fractional differencing parameter since we have:

   $$H = d + \frac{1}{2}$$

From the previous properties, we notice that $H \in [0, 1]$.

For estimating H, we generally use the R/S (or rescaled range) statistic. Let S_t and R_t be the standard deviation and the range of the sample $\{y_1, \ldots, y_t\}$. We have:

$$R_t = \max_{k \leq t} \sum_{j=1}^{k} (y_j - \bar{y}) - \min_{k \leq t} \sum_{j=1}^{k} (y_j - \bar{y})$$

where $\bar{y} = n^{-1} \sum_{t=1}^{n} y_t$. Lo (1991) defines the rescaled range statistic called Q_t as:

$$Q_t = \frac{R_t}{S_t}$$

We can show that $Q_t \sim ct^H$ when $t \to \infty$. We can then estimate the Hurst exponent by performing the linear regression:

$$\log Q_t = a + H \cdot \log t + u_t$$

[56] Generally, $\lambda_{\min}$ is set to $2\pi\sqrt{n}$.

[57] Since we have $c^{-H} y_{ct} \overset{\mathcal{D}}{\sim} y_t$, the spectrum satisfies the equation $c^{-(2H+1)} f_y\left(c^{-1}\lambda\right) = f_y(\lambda)$.

Lo (1991) proposes to test the null hypothesis $\mathcal{H}_0 : H = 0.5$ by considering the statistic $V_t = Q_t/\sqrt{t}$. Under the null hypothesis, the 95% confidence interval is $[0.809, 1.862]$.

We consider the daily return of the S&P 500 index and the daily variation of the VIX index from January 2007 to December 2017. In Figure 10.34, we report the estimated relationship $\log Q_t = \hat{a} + \hat{H} \log t$ for the two time series. We obtained $\hat{H} = 0.56$ for the S&P 500 index and $\hat{H} = 0.41$ for the VIX index. If we use the statistic V_t, we do not refuse the null hypothesis $\mathcal{H}_0$ at the 95% confidence level. However, we see that V_t reaches the lower bound in the case of the VIX index, whereas it can be higher than the upper bound in the case of the S&P 500 index. The S&P 500 could then exhibit long-range dependence, whereas the VIX index may be more mean-reverting. However, the whittle estimate $\hat{d}$ is respectively equal to -0.09 and -0.16 and is significant at the 99% confidence level. This confirms that the VIX index has short memory, but it contradicts that the S&P 500 index has long memory.

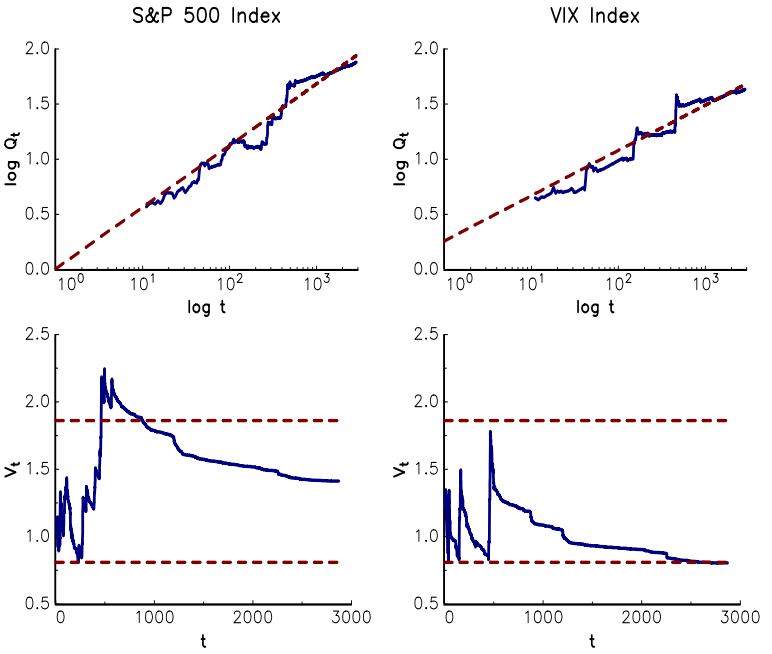


FIGURE 10.34: R/S analysis and estimation of the Hurst exponent

Signal decomposition On page 693, we do not explain how the cyclical component c_t is calculated. We reiterate that the Fourier transform of the logarithm of the number of lynx y_t is given by $d_y(\lambda_j) = \sum_{t=1}^{n} y_t e^{-i\lambda_j t}$. To recover the signal, we use the inverse Fourier transform $y_t = n^{-1} \sum_{j=1}^{n} d_y(\lambda_j) e^{i\lambda_j t}$. If we define $d_y^c(\lambda)$ as follows:

$$d_y^c(\lambda) = \begin{cases} d(\lambda_c) & \text{if } \lambda = \lambda_c \text{ or } \lambda = 2\pi - \lambda_c \\ 0 & \text{otherwise} \end{cases}$$

the cyclical component is equal to:

$$c_t = \bar{y} + \frac{1}{n} \sum_{j=1}^{n} d_y^c(\lambda_j) e^{i\lambda_j t}$$

Estimating the cyclical component is equivalent to apply the inverse Fourier transform to the Fourier coefficient corresponding to the cycle frequency λ_c.

The previous method can be generalized to the partition $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$:

$$\Lambda = \bigcup_{k=1}^{m} \Lambda_k$$

where $\Lambda_k \bigcap \Lambda_{k'} = \emptyset$. We define the component y_t^k as the inverse Fourier transform of the coefficients $d_y^k(\lambda)$:

$$d_y^k(\lambda) = \begin{cases} d(\lambda) & \text{if } \lambda \in \Lambda_k \\ 0 & \text{otherwise} \end{cases}$$

It follows that we have decomposed the original signal y_t into m signals[58]:

$$y_t = \sum_{k=1}^{m} y_t^k$$

This method allows to extract given frequency components. It is related to Parseval's theorem, which states that the sum of squares of a time series is equal to the sum of squares of its Fourier transform:

$$\sum_{t=1}^{n} |y_t|^2 = \frac{1}{n} \sum_{j=1}^{n} |d_y(\lambda_j)|^2$$

We consider the time series $y_t = c_t + u_t$, which is the sum of a long-term cyclical component and a residual component. The long-term component is the sum of three cycles c_t^k, whose periods are larger than 5 years. The residual component is the sum of a white noise process and 5 short-term cycles, whose periods are lower than 1 year. We represent y_t and the components c_t^k, c_t and u_t in Figure 10.35 for the 2 500 dates[59]. Then, we consider the signal reconstruction based on the m most significant frequencies:

$$y_t^m = \frac{1}{n} \sum_{j=1}^{n} d_y^m(\lambda_j) e^{i\lambda_j t}$$

where $d_y^m(\lambda_j)$ is equal to zero for the 2 500 frequencies except for the m frequencies with the highest values of $|d_y(\lambda_j)|$. Figure 10.36 shows the reconstructed signal y_t^m for different values of m. We notice that we may describe the dynamics of y_t with very few Fourier coefficients. Using the Parseval's theorem, we define the energy ratio as follows:

$$\mathcal{ER}_m = \frac{\sum_{t=1}^{n} |y_t^m|^2}{\sum_{t=1}^{n} |y_t|^2} = \frac{\sum_{k=1}^{m} |\delta_{k:n}|^2}{\sum_{j=1}^{n} |d_y(\lambda_j)|^2}$$

where $\delta_{k:n}$ is the k^{th} reverse order statistic of $d_y(\lambda_j)$. Results in Table 10.13 shows that one Fourier frequency explains 35% of the total variance of y_t, two Fourier frequencies explain 70% of the total variance of y_t, etc. With 50 Fourier frequencies, that is 2% of all the Fourier frequencies, we explain more than 96% of the total variance of y_t.

[58] This approach is also called subband coding or subband decomposition.
[59] We assume that each year is composed of 250 trading days.

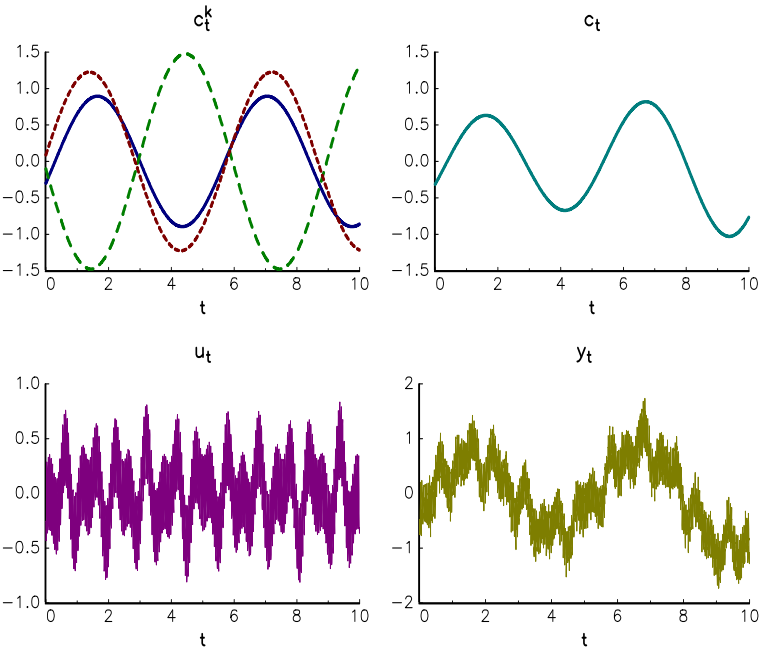


FIGURE 10.35: Spectral decomposition of the signal y_t

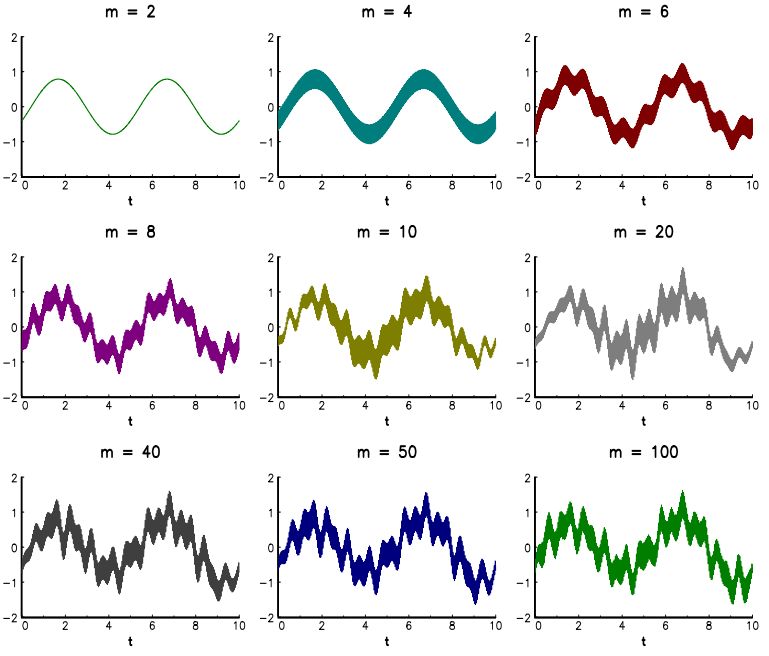


FIGURE 10.36: Reconstructed signal y_t^m

TABLE 10.13: Value of the energy ratio $\mathcal{ER}_m$ (in %)

m	1	2	3	4	5	10	25	50	75	100	2 500
$\mathcal{ER}_m$	34.8	69.5	73.8	78.1	79.9	87.4	94.4	96.5	97.3	97.5	100.0

Filtering theory Previously, we have implicitly used the concept of filtering in order to extract the cycle component. More generally, the filtering technique consists in removing some undesired frequencies. From a given signal x_t, we build another signal y_t such that:

$$y_t = \varphi(L)\, x_t$$

x_t is also called the input process, whereas y_t is the output process. Since $\varphi(L)$ is independent from time t, $\varphi(L)$ is a time-invariant filter. We distinguish four families of such filters:

1. low-pass filters reduce components of high frequencies in order to get the dynamics due to low frequencies; a famous example is the moving average:

$$y_t = \frac{1}{m}\sum_{k=0}^{m-1} x_{t-k}$$

2. high-pass filters reduce components of low frequencies in order to get the dynamics due to high frequencies; the linear difference $y_t = x_t - x_{t-1}$ is an example of high-pass filters;

3. band-pass filters combine a low-pass filter with a high-pass filter; they can be used to study the dynamics due to medium frequencies;

4. band-stop filters are the opposite of band-pass filters; therefore, they remove medium frequencies.

We also make the distinction between causal and non-causal filters. In the case of a causal filter, $\varphi(L)$ can be written as $\sum_{k=0}^{m-1} \varphi_k L^k$ meaning that y_t does not depend on the future values of x_t. If $\varphi(L)$ is linear, it is said linear time-invariant or LTI filter.

We recall that the spectral density function of y_t is:

$$f_y(\lambda) = \left|\varphi\left(e^{-i\lambda}\right)\right|^2 f_x(\lambda) \tag{10.63}$$

The function $\varphi\left(e^{-i\lambda}\right)$ is known as the frequency response or transfer function. Equation (10.63) can be seen as the frequency-domain version of the time-domain equality $\text{var}(Y) = \text{var}(aX+b) = a^2\,\text{var}(X)$. Let us consider the unit signal:

$$x_t = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$

We can then calculate the impulse response $y_t = \varphi(L)\, x_t$. If $y_t = 0$ for any date $t \geq t^\star$, the impulse response is finite, and the filter is known as a finite impulse response (or FIR) filter. An example is the moving average filter:

$$\varphi(L) = \frac{L^0 + L^1}{2} \Rightarrow \begin{cases} y_0 = y_1 = \dfrac{1}{2} \\ y_t = 0 \text{ for } t > 1 \end{cases}$$

Otherwise, we have an infinite impulse response (or IIR) filter. For instance, this is the case of the AR(1) filter:

$$\varphi(L) = (1 - \phi L)^{-1} \Rightarrow y_t = \phi^t > 0$$

The output signal y_t is the discrete convolution of the input signal x_t and the vector φ_t of the filter coefficients:

$$y_t = \varphi_t * x_t$$

Using the Fourier transform, we deduce that:

$$d_y(\lambda) = d_\varphi(\lambda)\, d_x(\lambda)$$

We can then obtain the output signal y_t by multiplying the discrete Fourier transforms of φ_t and x_t, and taking the inverse Fourier transform of the product. For extracting a cycle, we use the following transfer function:

$$d_\varphi(\lambda) = \begin{cases} 1 & \text{if } \lambda = \lambda_c \\ 0 & \text{otherwise} \end{cases}$$

For a low-pass filter, we can consider a transfer function such that:

$$d_\varphi(\lambda) = \begin{cases} 1 & \text{if } \lambda \in [0, \lambda^\star[\\ 0 & \text{if } \lambda \in [\lambda^\star, \pi] \end{cases}$$

Of course, we can specify more complicated filters. For example, the Hodrick-Prescott and Baxter-King filters that are used for estimating the business cycle are respectively a high-pass filter and a band-pass filter.

Remark 137 *Time domain analysis consists in localizing common and residual patterns associated to a signal in the time scale. Frequency domain analysis (or spectral analysis) does the same job by considering a frequency scale. The wavelet analysis combines the two approaches to study the signal in the time-frequency domain.*

10.3 Exercises

10.3.1 Probability distribution of the t-statistic in the case of the linear regression model

We consider the linear regression model:

$$y_i = x_i^\top \beta + u_i$$

where y_i is a scalar, x_i is a $K \times 1$ vector and u_i is a random variable. By considering a sample of n observations, the matrix form of the linear regression model is:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

We assume the standard assumptions: $\mathcal{H}_1 : \mathbf{U} \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$, $\mathcal{H}_2 : \mathbf{X}$ is a $n \times K$ matrix of known values, $\mathcal{H}_3 : \text{rank}(\mathbf{X}) = K$ and $\mathcal{H}_4 : \lim n^{-1}\left(\mathbf{X}^\top\mathbf{X}\right) = Q$. We note $\hat{\beta}$ the OLS estimator and $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$ the hat matrix.

1. Show that the matrices $\mathbf{H}$ and $\mathbf{L} = I_n - \mathbf{H}$ are symmetric and idempotent.
2. Show that $\mathbf{LX} = 0$ and $\mathbf{X}^\top \mathbf{L} = 0$. Deduce that $\hat{\mathbf{U}} = \mathbf{LU}$.
3. Calculate trace $(\mathbf{L})$ and rank $(\mathbf{L})$.
4. We note RSS (β) the residual sum of squares. Show that $\hat{\sigma}^2 = (n-K)^{-1} \operatorname{RSS}\left(\hat{\beta}\right)$ is an unbiased estimator of σ^2.
5. By considering the normalized random vector $\mathbf{V} = (\sigma I_n)^{-1} \mathbf{U}$, find the probability distribution of $\hat{\sigma}^2$.
6. Show that $\hat{\beta}$ and $\hat{\mathbf{U}}$ are independent.
7. Find the probability distribution of $t\left(\hat{\beta}_j\right)$:

$$t\left(\hat{\beta}_j\right) = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\left(\hat{\beta}_j\right)}$$

10.3.2 Linear regression without a constant

We consider the linear model $y_i = x_i^\top \beta + \varepsilon_i$ where $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$ and $\mathbb{E}\left[\varepsilon_i \varepsilon_j\right] = 0$.

1. Write this model in the matrix form: $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$. Compute the sum of squared residuals $\varepsilon^\top \varepsilon$. Deduce that the least squares estimator $\hat{\beta} = \arg\min \varepsilon^\top \varepsilon$ is the solution of a quadratic programming problem.
2. We assume that the linear model does not contain an intercept.
   (a) Show that the residuals are not centered.
   (b) Write the quadratic programming problem associated to the least squares estimator if we impose that the residuals are centered.
   (c) Transform the previous optimization problem with explicit constraints into an optimization problem with implicit constraints. Deduce the analytical solution.

10.3.3 Linear regression with linear constraints

1. We consider the linear regression:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

   (a) Let RSS $(\beta) = \mathbf{U}^\top \mathbf{U}$ denote the residual sum of squares. Calculate RSS (β) with respect to $\mathbf{Y}$, $\mathbf{X}$ and β.
   (b) Deduce the OLS estimator:

$$\hat{\beta} = \arg\min \operatorname{RSS}(\beta)$$

   (c) We assume that $\mathbf{U} \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$. Show that $\hat{\beta}$ is an unbiased estimator. Deduce the variance of $\hat{\beta}$.

2. We now introduce a system $\mathcal{C}$ of constraints defined by:

$$\mathcal{C} = \begin{cases} A\beta = B \\ C\beta \geq D \end{cases}$$

TABLE 10.14: Numerical example

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 28.88 & 23.15 & 20.75 & 21.60 & 22.97 \\ 23.15 & 35.01 & 24.73 & 25.14 & 24.73 \\ 20.75 & 24.73 & 28.23 & 22.42 & 21.63 \\ 21.60 & 25.14 & 22.42 & 32.22 & 24.17 \\ 22.97 & 24.73 & 21.63 & 24.17 & 33.10 \end{pmatrix} \quad \mathbf{X}^\top\mathbf{Y} = \begin{pmatrix} 100.53 \\ 136.62 \\ 128.20 \\ 146.07 \\ 117.01 \end{pmatrix}$$

(a) Show that the constrained estimator $\tilde{\beta}$ is the solution of a quadratic programming problem.

(b) We consider the numerical example given in Table 10.14.

i. Calculate $\tilde{\beta}$ when $\sum_{i=1}^5 \beta_i = 1$.

ii. Calculate $\tilde{\beta}$ when $\beta_1 = \beta_2 = \beta_5$.

iii. Calculate $\tilde{\beta}$ when $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4 \geq \beta_5$.

iv. We assume that $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4 \leq \beta_5$ and $\sum_{i=1}^5 \beta_i = 1$. Verify that:

$$\tilde{\beta} = \begin{pmatrix} -2.6276 \\ 0.9069 \\ 0.9069 \\ 0.9069 \\ 0.9069 \end{pmatrix}$$

Deduce the values of the Lagrange coefficients for the inequality constraints given that the Lagrange coefficient of the equality constraint is equal to -192.36304.

3. We assume that $A\beta = B$.

(a) Write the Lagrange function and deduce the constrained OLS estimator $\tilde{\beta}$.

(b) Show that we can write the explicit constraints $A\beta = B$ into implicit constraints $\beta = C\gamma + D$. Write the constrained residual sum of squares $\text{RSS}(\gamma)$. Deduce the expressions of the estimator $\hat{\gamma}$ and the constrained estimator $\tilde{\beta}$.

(c) Verify the coherence between the two estimators.

(d) Verify that we obtain the same estimates in the case $\beta_1 = \beta_2$ and $\beta_1 = \beta_5 + 1$.

10.3.4 Maximum likelihood estimation of the Poisson distribution

We consider a sample $\mathbf{Y} = \{y_1, \ldots, y_n\}$ generated by the Poisson distribution $\mathcal{P}(\lambda)$.

1. Find the MLE of λ.

2. Calculate the information matrix $\mathcal{I}(\lambda)$. Deduce the variance of $\hat{\lambda}$. Compare this expression with the direct computation based on the Hessian matrix.

10.3.5 Maximum likelihood estimation of the exponential distribution

We consider a sample $\mathbf{Y} = \{y_1, \ldots, y_n\}$ generated by the Exponential distribution $\mathcal{E}(\lambda)$.

1. Find the MLE of λ.
2. Calculate the information matrix $\mathcal{I}(\lambda)$. Deduce the variance of $\hat{\lambda}$. Compare this expression with the direct computation based on the Hessian matrix.

10.3.6 Relationship between the linear regression and the maximum likelihood method

We consider the standard linear regression model:

$$y_i = x_i^\top \beta + u_i$$

where $u_i \sim \mathcal{N}\left(0, \sigma^2\right)$.

1. Write the log-likelihood function $\boldsymbol{\ell}(\theta)$ associated to the sample $\mathbf{Y} = (y_1, \ldots, y_n)$.
2. Find the ML estimator $\hat{\theta}$. What is the relationship between $\hat{\theta}_{\text{ML}}$ and $\hat{\theta}_{\text{OLS}}$?
3. Compute var $\left(\hat{\theta}_{\text{ML}}\right)$.

10.3.7 The Gaussian mixture model

We consider the mixture Y of two independent Gaussian random variables Y_1 and Y_2. The distribution function of Y is:

$$f(y) = \pi_1 f_1(y) + \pi_2 f_2(y)$$

where $Y_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$, $Y_2 \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$ and $\pi_1 + \pi_2 = 1$.

1. Show that:
$$\mathbb{E}\left[Y^k\right] = \pi_1 \mathbb{E}\left[Y_1^k\right] + \pi_2 \mathbb{E}\left[Y_2^k\right]$$
2. Deduce $\mathbb{E}[Y]$ and var (Y).
3. Find the expression of the skewness coefficient $\gamma_1(Y)$.

10.3.8 Parameter estimation of diffusion processes

We consider a sample $\mathbf{X} = \{x_0, x_1, \ldots, x_T\}$ of the diffusion process $X(t)$, which is observed at irregular times $t = \{t_0, t_1, \ldots, t_T\}$. Therefore, we have $x_i = X(t_i)$.

1. Find the expression of the log-likelihood function associated to the geometric Brownian motion:
$$\mathrm{d}X(t) = \mu X(t)\, \mathrm{d}t + \sigma X(t)\, \mathrm{d}W(t)$$
2. Same question with the Ornstein-Uhlenbeck process:
$$\mathrm{d}X(t) = a\left(b - X(t)\right) \mathrm{d}t + \sigma\, \mathrm{d}W(t)$$

3. We consider the general SDE:

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t)$$

Calculate the log-likelihood function associated to the Euler-Maruyama scheme. Apply this result to the Cox-Ingersoll-Ross process:

$$\mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\,\mathrm{d}W(t)$$

Compare this approach with the quasi-maximum likelihood (QML) estimation by assuming conditional normality of innovation processes.

4. Find the orthogonality conditions of the GMM approach for the GBM, OU and CIR processes.

5. Same question if we consider the CKLS process (Chan *et al.*, 1992):

$$\mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma|X(t)|^{\gamma}\,\mathrm{d}W(t)$$

10.3.9 The Tobit model

1. Give the first two moments of the truncated random variable $X \mid X \geq c$ where $X \sim \mathcal{N}(\mu, \sigma^2)$.

2. Calculate the first two moments of the censored random variable $\tilde{Y} = \max(X, c)$.

3. Illustrate the difference between truncation and censoring when the parameters are $\mu = 2$, $\sigma = 3$ and $c = 1$.

We consider the Tobit model defined as follows:

$$\begin{cases} y_i = \max(0, y_i^\star) \\ y_i^\star = x_i^\top \beta + u_i \end{cases}$$

where $u_i \sim \mathcal{N}(0, \sigma^2)$. We have a linear model between the latent variable $y_i^\star$ and K explanatory variables x_i. However, we do not directly observe the data $\{x_i, y_i^\star\}$. Indeed, we observe the data $\{x_i, y_i\}$, implying that the dependent variable y_i is censored.

4. Write the log-likelihood function $\boldsymbol{\ell}(\theta)$ where $\theta = (\beta, \sigma^2)$.

5. Find the first-order conditions of the ML estimator $\hat{\theta}$.

6. Calculate the Hessian matrix $H(\theta)$ associated to the log-likelihood function $\boldsymbol{\ell}(\theta)$.

7. Show that the information matrix has the following representation (Amemiya, 1973):

$$\mathcal{I}(\theta) = \begin{pmatrix} \sum_{i=1}^n a_i x_i x_i^\top & \sum_{i=1}^n b_i x_i \\ \sum_{i=1}^n b_i x_i & \sum_{i=1}^n c_i \end{pmatrix}$$

where a_i, b_i and c_i are three scalars to define.

8. Show that the OLS estimator based on non-censored data is biased.

9. Compute the conditional expectations $\mathbb{E}[y_i \mid y_i > 0]$ and $\mathbb{E}[y_i \mid y_i \leq 0]$, and the unconditional expectation $\mathbb{E}[y_i]$. Propose an OLS estimator $\tilde{\beta}$ and compare it with the ML estimator $\hat{\beta}$.

10. We consider the data given in Table 10.15. Using the method of maximum likelihood, estimate the Tobit model:

$$\begin{cases} y_i = \max\left(0, y_i^\star\right) \\ y_i^\star = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{2,1} + u_i \end{cases}$$

where $u_i \sim \mathcal{N}\left(0, \sigma^2\right)$. Calculate the OLS estimates based on the non-censored data. Verify that:

$$\hat{\beta}^{(\text{OLS})} - \hat{\sigma}^{(\text{ML})} \left(\mathbf{X}_1^\top \mathbf{X}_1\right)^{-1} \mathbf{X}_0^\top \mathbf{\Lambda}_0^{(\text{ML})} = \hat{\beta}^{(\text{ML})}$$

and:

$$\hat{\beta}^{(\text{OLS})} - \hat{\sigma}^{(\text{ML})} \left(\mathbf{X}_1^\top \mathbf{X}_1\right)^{-1} \mathbf{X}_1^\top \mathbf{\Lambda}_1^{(\text{ML})} \neq \hat{\beta}^{(\text{ML})}$$

Comment on these results.

11. How to calculate the predicted value $\breve{y}_i^\star$ given that we know if the observation is censured or not? Compare the numerical value of $\breve{y}_i^\star$ with the unconditional predicted value $\hat{y}_i^\star$.

TABLE 10.15: Data of the Tobit example

i	1	2	3	4	5	6	7	8	9	10
y_i	4.0	0.0	0.5	0.0	0.0	17.4	18.0	0.0	0.0	9.7
$x_{1,i}$	−4.3	−9.2	−2.8	−2.7	−8.4	2.0	5.3	−8.1	0.9	−7.8
$x_{2,i}$	−1.2	−5.5	1.8	−3.4	2.9	9.3	9.1	6.8	−6.3	4.3
i	11	12	13	14	15	16	17	18	19	20
y_i	9.7	1.8	6.5	26.1	0.0	5.0	21.6	6.2	9.9	1.4
$x_{1,i}$	0.5	6.7	−0.9	4.0	−8.6	2.3	7.1	7.3	9.3	−0.2
$x_{2,i}$	5.3	−8.6	2.1	9.2	−8.7	−7.9	9.2	−7.5	−4.4	1.7
i	21	22	23	24	25	26	27	28	29	30
y_i	5.0	0.0	0.0	18.1	0.0	7.7	0.0	0.0	0.0	4.0
$x_{1,i}$	−3.5	−1.5	−2.6	8.4	8.6	5.8	−7.9	0.9	−7.3	2.3
$x_{2,i}$	0.1	−4.5	−8.9	3.8	−8.5	8.3	1.8	−6.2	8.4	6.7

10.3.10 Derivation of Kalman filter equations

We consider the standard state space model described on page 647:

$$\begin{cases} y_t = Z_t \alpha_t + d_t + \epsilon_t \\ \alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \end{cases}$$

1. Show that the prior estimates are given by the following relationships:

$$\begin{aligned} \hat{\alpha}_{t|t-1} &= T_t \hat{\alpha}_{t-1|t-1} + c_t \\ P_{t|t-1} &= T_t P_{t-1|t-1} T_t^\top + R_t Q_t R_t^\top \end{aligned}$$

2. Deduce that the innovation $v_t = y_t - \mathbb{E}_{t-1}\left[y_t\right]$ is a centered Gaussian random vector, whose covariance matrix F_t is equal to:

$$F_t = Z_t P_{t|t-1} Z_t^\top + H_t$$

3. Show that the joint distribution of the random vector (α_t, v_t) conditionally to the filtration $\mathcal{F}_{t-1}$ is equal to:

$$\begin{pmatrix} \alpha_t \\ v_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{\alpha}_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & P_{t|t-1}Z_t^\top \\ Z_t P_{t|t-1} & F_t \end{pmatrix}\right)$$

4. Deduce that:

$$\hat{\alpha}_{t|t} = \mathbb{E}\left[\mathbb{E}_{t-1}\left[\alpha_t\right] \mid v_t = y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right]$$

and:

$$\begin{aligned} \hat{\alpha}_{t|t} &= \hat{\alpha}_{t|t-1} + P_{t|t-1}Z_t^\top F_t^{-1}\left(y_t - Z_t\hat{\alpha}_{t|t-1} - d_t\right) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}Z_t^\top F_t^{-1} Z_t P_{t|t-1} \end{aligned}$$

5. Summarize the equations of the Kalman filter. Deduce that there exists a matrix K_t such that:

$$\hat{\alpha}_{t+1|t} = T_{t+1}\hat{\alpha}_{t|t-1} + c_{t+1} + K_t v_t$$

Rewrite the state space model as an innovation process. What is the interpretation of K_t?

6. Show that the state space model can be written as:

$$\begin{cases} y_t = Z_t^\star \alpha_t^\star \\ \alpha_t^\star = T_t^\star \alpha_{t-1}^\star + R_t^\star \eta_t^\star \end{cases}$$

Define the state vector $\alpha_t^\star$, the matrices $Z_t^\star$, $T_t^\star$ and $R_t^\star$, and the random vector $\eta_t^\star$.

7. Deduce the Kalman filter when the white noise process ϵ_t and η_t are correlated:

$$\mathbb{E}\left[\epsilon_t \eta_t^\top\right] = C_t$$

10.3.11 Steady state of time-invariant state space model

We note $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

1. Calculate the steady state of the SSM associated to the AR(1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

2. Calculate the steady state of the SSM associated to the MA(1) process:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

3. Calculate the steady state of the SSM associated to the ARMA(1,1) process:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

4. Calculate the steady state of the SSM associated to the process:

$$\begin{aligned} y_t &= \mu + u_t \\ u_t &= \theta_1 u_{t-1} + \varepsilon_t \end{aligned}$$

10.3.12 Kalman information filter versus Kalman covariance filter

We assume that all square matrices are invertible and all non-square matrices have a Moore-Penrose pseudo-inverse. We note A, B, C and D four matrices of dimension $m \times m$, $n \times m$, $n \times n$ and $m \times m$.

1. Show that:
$$\left(I_m + AB^\top C^{-1}B\right)^{-1} A = \left(A^{-1} + B^\top C^{-1}B\right)^{-1}$$

2. Verify the following relationship:
$$\left(I_m + AB^\top C^{-1}B\right)^{-1} = I_m - AB^\top \left(C + BAB^\top\right)^{-1} B$$

3. Deduce that:
$$\left(I_m + AB^\top C^{-1}B\right)^{-1} AB^\top C^{-1} = AB^\top \left(C + BAB^\top\right)^{-1}$$

4. Calculate $\left(I_m + D^{-1}A\right)(A + D)^{-1}$.

We consider the following state space model:

$$\begin{cases} y_t = Z_t \alpha_t + \epsilon_t \\ \alpha_t = T_t \alpha_{t-1} + R_t \eta_t \end{cases}$$

It correspond to a special case of the model described on page 647 when there are no constants c_t and d_t in state and measurement equations.

5. Define the concept of information matrix.

6. We introduce the notations $\mathbb{I}_{t|t} = P_{t|t}^{-1}$, $\mathbb{I}_{t|t-1} = P_{t|t-1}^{-1}$, $\hat{\alpha}_{t|t}^\star = \mathbb{I}_{t|t}\hat{\alpha}_{t|t}$ and $\hat{\alpha}_{t|t-1}^\star = \mathbb{I}_{t|t-1}\hat{\alpha}_{t|t-1}$. How do you interpret the vectors $\hat{\alpha}_{t|t}^\star$ and $\hat{\alpha}_{t|t-1}^\star$?

7. By using the results of Questions 1-4, show that:
$$\mathbb{I}_{t|t} P_{t|t-1} = I_m + Z_t^\top H_t^{-1} Z_t P_{t|t-1}$$
Deduce that:
$$\mathbb{I}_{t|t} P_{t|t-1} Z_t^\top \left(Z_t P_{t|t-1} Z_t^\top + H_t\right)^{-1} = Z_t^\top H_t^{-1}$$

8. Verify that the recursive equations of the information filter are:
$$\begin{cases} \mathbb{I}_{t|t-1} = \left(T_t \mathbb{I}_{t-1|t-1}^{-1} T_t^\top + R_t Q_t R_t^\top\right)^{-1} \\ \hat{\alpha}_{t|t-1}^\star = \mathbb{I}_{t|t-1} T_t \mathbb{I}_{t-1|t-1}^{-1} \hat{\alpha}_{t-1|t-1}^\star \\ \mathbb{I}_{t|t} = \mathbb{I}_{t|t-1} + Z_t^\top H_t^{-1} Z_t \\ \hat{\alpha}_{t|t}^\star = \hat{\alpha}_{t|t-1}^\star + Z_t^\top H_t^{-1} y_t \end{cases}$$
What advantages would you see in using the Kalman information filter rather than the Kalman covariance filter?

9. We assume that the probability distribution of α_0 is diffuse. Give the log-likelihood function of the sample $\{y_1, \ldots, y_T\}$ when considering the Kalman information filter. How to take into account the diffuse assumption when considering the Kalman covariance filter?

10.3.13 Granger representation theorem

We assume that y_t is a VaR(p) process:

$$\Phi'(L)\, y_t = \mu_t + \varepsilon_t$$

where:

$$\Phi'(L) = I_n - \Phi_1' L - \ldots - \Phi_p' L^p$$

1. We consider the case $p = 1$. Show that:

$$\Delta y_t = \mu_t + (\Phi_1' - I_n)\, y_{t-1} + \varepsilon_t$$

2. We consider the case $p = 2$. Show that:

$$\Delta y_t = \mu_t + (\Phi_1' + \Phi_2' - I_n)\, y_{t-1} - \Phi_2' \Delta y_{t-1} + \varepsilon_t$$

3. Verify that the expression of Δy_t in the general case is equal to:

$$\Delta y_t = \mu_t + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \varepsilon_t$$

where $\Pi = -\Phi'(1) = \sum_{i=1}^{p} \Phi_i' - I_n$ and $\Phi_i = -\sum_{j=i+1}^{p} \Phi_j'$.

10.3.14 Probability distribution of the periodogram

Let y_t be a stationary centered process. We decompose the periodogram as the sum of two parts:

$$\begin{aligned} I_y(\lambda) &= \frac{1}{2\pi n} \left| \sum_{t=1}^{n} y_t e^{-i\lambda_j t} \right|^2 \\ &= \frac{a^2(\lambda_j) + b^2(\lambda_j)}{2\pi} \end{aligned}$$

where $a(\lambda_j) = n^{-1/2} \sum_{t=1}^{n} y_t \cos(\lambda_j t)$ and $b(\lambda_j) = n^{-1/2} \sum_{t=1}^{n} y_t \sin(\lambda_j t)$.

1. We assume that $y_t \sim \mathcal{N}(0, \sigma^2)$. Show that[60]:

$$\lim_{n\to\infty} a_n(\lambda_j) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2}\right)$$

and:

$$\lim_{n\to\infty} b_n(\lambda_j) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2}\right)$$

2. Verify that $a(\lambda_j)$ and $b(\lambda_j)$ are asymptotically independent. Deduce the probability distribution of $I_y(\lambda_j)$.

[60] We recall that:

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{t=1}^{n} \cos(\alpha t) \right) = 0$$

when $\alpha \neq 0$.

3. More generally, we assume that $\lim_{n\to\infty} 2f_y^{-1}(\lambda_j) I_y(\lambda_j) \sim \chi_2^2$ for any stationary centered process y_t. Calculate the first two moments of $I_y(\lambda_j)$ and the 95% confidence interval of $f_y(\lambda_j)$.

4. We consider Question 2 when $\lambda_j = 0$. What is the probability distribution of $I_y(0)$? Formulate an hypothesis about the probability distribution of $I_y(0)$ for all stationary centered process y_t. Show that $\lim_{n\to\infty} \mathbb{E}[I_y(0)] = f_y(0)$ and $\lim_{n\to\infty} \text{var}(I_y(0)) = 2f_y^2(0)$.

10.3.15 Spectral density function of structural time series models

We consider the following models:

(M1)
$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \eta_t \end{cases}$$

(M2)
$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \end{cases}$$

(M3)
$$\begin{cases} y_t = \mu_t + \beta_t + \gamma_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \eta_t \\ \beta_t = \phi\beta_{t-1} + \zeta_t \\ \sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \end{cases}$$

where ε_t, η_t, ζ_t and ω_t are independent white noise processes with variances σ_ε^2 , σ_η^2, σ_ζ^2 and σ_ω^2.

1. Write Models (M1) and (M2) in the state space form.

2. Find the stationary form of these two processes and calculate their spectral density function.

3. Illustrate graphically the difference between these spectral density functions.

4. We consider Model (M3). Give an interpretation of the components μ_t, β_t and γ_t.

5. Show that a stationary form of y_t is:
$$z_t = (1-L)(1-L^s) y_t$$

6. Give another stationary form of y_t.

7. Find the spectral density function of z_t.

10.3.16 Spectral density function of some processes

Calculate the spectral density function of the following processes:

1. y_t is a periodic random walk process:

$$y_t = y_{t-s} + \varepsilon_t$$

where $s > 1$ and $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

2. y_t is a fractional white noise process:

$$(1 - L)^d y_t = \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

3. z_t is the sum of an AR(1) process and a MA(1) process:

$$\begin{cases} z_t = x_t + y_t \\ x_t = \phi x_{t-1} + u_t \\ y_t = v_t - \theta v_{t-1} \end{cases}$$

where $u_t \sim \mathcal{N}\left(0, \sigma_u^2\right)$, $v_t \sim \mathcal{N}\left(0, \sigma_v^2\right)$ and $u_t \perp v_t$.

   (a) Simulate a sample z of 1 000 observations with the following parameter: $\phi = 0.75$, $\theta = 0.2$, $\sigma_u = 1$ and $\sigma_v = 0.5$. Draw the periodogram of z.

   (b) Estimate the parameters ϕ, σ_u, θ and σ_v by using the method of Whittle. Compare the estimated spectral density function with the periodogram of z and the theoretical spectral density function.

Chapter 11

Copulas and Dependence Modeling

One of the main challenges in risk management is the aggregation of individual risks. We can move the issue aside by assuming that the random variables modeling individual risks are independent or are only dependent by means of a common risk factor. The problem becomes much more involved when one wants to model fully dependent random variables. Again a classic solution is to assume that the vector of individual risks follows a multivariate normal distribution. However, all risks are not likely to be well described by a Gaussian random vector, and the normal distribution may fail to catch some features of the dependence between individual risks.

Copula functions are a statistical tool to solve the previous issue. A copula function is nothing else but the joint distribution of a vector of uniform random variables. Since it is always possible to map any random vector into a vector of uniform random variables, we are able to split the marginals and the dependence between the random variables. Therefore, a copula function represents the statistical dependence between random variables, and generalizes the concept of correlation when the random vector is not Gaussian.

11.1 Canonical representation of multivariate distributions

The concept of copula has been introduced by Sklar in 1959. During a long time, only a small number of people have used copula functions, more in the field of mathematics than this of statistics. The publication of Genest and MacKay (1986b) in the *American Statistician* marks a breakdown and opens areas of study in empirical modeling, statistics and econometrics. In what follows, we intensively use the materials developed in the books of Joe (1997) and Nelsen (2006).

11.1.1 Sklar's theorem

Nelsen (2006) defines a bi-dimensional copula (or a 2-copula) as a function $\mathbf{C}$ which satisfies the following properties:

1. $\operatorname{Dom}\mathbf{C} = [0,1] \times [0,1]$;
2. $\mathbf{C}(0,u) = \mathbf{C}(u,0) = 0$ and $\mathbf{C}(1,u) = \mathbf{C}(u,1) = u$ for all u in $[0,1]$;
3. $\mathbf{C}$ is 2-increasing:

$$\mathbf{C}(v_1,v_2) - \mathbf{C}(v_1,u_2) - \mathbf{C}(u_1,v_2) + \mathbf{C}(u_1,u_2) \geq 0$$

for all $(u_1,u_2) \in [0,1]^2$, $(v_1,v_2) \in [0,1]^2$ such that $0 \leq u_1 \leq v_1 \leq 1$ and $0 \leq u_2 \leq v_2 \leq 1$.

This definition means that $\mathbf{C}$ is a cumulative distribution function with uniform marginals:

$$\mathbf{C}(u_1, u_2) = \Pr\{U_1 \leq u_1, U_2 \leq u_2\}$$

where U_1 and U_2 are two uniform random variables.

Example 108 *Let us consider the function* $\mathbf{C}^{\perp}(u_1, u_2) = u_1 u_2$. *We have* $\mathbf{C}^{\perp}(0, u) = \mathbf{C}^{\perp}(u, 0) = 0$ *and* $\mathbf{C}^{\perp}(1, u) = \mathbf{C}^{\perp}(u, 1) = u$. *Since we have* $v_2 - u_2 \geq 0$ *and* $v_1 \geq u_1$, *it follows that* $v_1(v_2 - u_2) \geq u_1(v_2 - u_2)$ *and* $v_1 v_2 + u_1 u_2 - u_1 v_2 - v_1 u_2 \geq 0$. *We deduce that* $\mathbf{C}^{\perp}$ *is a copula function. It is called the product copula.*

Let $\mathbf{F}_1$ and $\mathbf{F}_2$ be any two univariate distributions. It is obvious that $\mathbf{F}(x_1, x_2) = \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2))$ is a probability distribution with marginals $\mathbf{F}_1$ and $\mathbf{F}_2$. Indeed, $u_i = \mathbf{F}_i(x_i)$ defines a uniform transformation ($u_i \in [0, 1]$). Moreover, we verify that $\mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(\infty)) = \mathbf{C}(\mathbf{F}_1(x_1), 1) = \mathbf{F}_1(x_1)$. Copulas are then a powerful tool to build a multivariate probability distribution when the marginals are given. Conversely, Sklar (1959) proves that any bivariate distribution $\mathbf{F}$ admits such a representation:

$$\mathbf{F}(x_1, x_2) = \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \tag{11.1}$$

and that the copula $\mathbf{C}$ is unique provided the marginals are continuous. This result is important, because we can associate to each bivariate distribution a copula function. We then obtain a canonical representation of a bivariate probability distribution: on one side, we have the marginals or the univariate directions $\mathbf{F}_1$ and $\mathbf{F}_2$; on the other side, we have the copula $\mathbf{C}$ that links these marginals and gives the dependence between the unidimensional directions.

Example 109 *The Gumbel logistic distribution is the function* $\mathbf{F}(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}$ *defined on* $\mathbb{R}^2$. *We notice that the marginals are* $\mathbf{F}_1(x_1) \equiv \mathbf{F}(x_1, \infty) = (1 + e^{-x_1})^{-1}$ *and* $\mathbf{F}_2(x_2) \equiv (1 + e^{-x_2})^{-1}$. *The quantile functions are then* $\mathbf{F}_1^{-1}(u_1) = \ln u_1 - \ln(1 - u_1)$ *and* $\mathbf{F}_2^{-1}(u_2) = \ln u_2 - \ln(1 - u_2)$. *We finally deduce that:*

$$\mathbf{C}(u_1, u_2) = \mathbf{F}\left(\mathbf{F}_1^{-1}(u_1), \mathbf{F}_2^{-1}(u_2)\right) = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}$$

is the Gumbel logistic copula.

11.1.2 Expression of the copula density

If the joint distribution function $\mathbf{F}(x_1, x_2)$ is absolutely continuous, we obtain:

$$\begin{aligned} f(x_1, x_2) &= \partial_{1,2} \mathbf{F}(x_1, x_2) \\ &= \partial_{1,2} \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \\ &= c(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \end{aligned} \tag{11.2}$$

where $f(x_1, x_2)$ is the joint probability density function, f_1 and f_2 are the marginal densities and c is the copula density:

$$c(u_1, u_2) = \partial_{1,2} \mathbf{C}(u_1, u_2)$$

We notice that the condition $\mathbf{C}(v_1, v_2) - \mathbf{C}(v_1, u_2) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, u_2) \geq 0$ is then equivalent to $\partial_{1,2} \mathbf{C}(u_1, u_2) \geq 0$ when the copula density exists.

Example 110 *In the case of the Gumbel logistic copula, we obtain* $c(u_1, u_2) = 2u_1 u_2 / (u_1 + u_2 - u_1 u_2)^3$. *We easily verify the 2-increasing property.*

From Equation (11.2), we deduce that:

$$c(u_1, u_2) = \frac{f\left(\mathbf{F}_1^{-1}(u_1), \mathbf{F}_2^{-1}(u_2)\right)}{f_1\left(\mathbf{F}_1^{-1}(u_1)\right) \cdot f_2\left(\mathbf{F}_2^{-1}(u_2)\right)} \tag{11.3}$$

We obtain a second canonical representation based on density functions. For some copulas, there is no explicit analytical formula. This is the case of the Normal copula, which is equal to $\mathbf{C}(u_1, u_2; \rho) = \Phi\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho\right)$. Using Equation (11.3), we can however characterize its density function:

$$\begin{aligned} c(u_1, u_2; \rho) &= \frac{2\pi\left(1-\rho^2\right)^{-1/2} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x_1^2 + x_2^2 - 2\rho x_1 x_2\right)\right)}{(2\pi)^{-1/2}\exp\left(-\frac{1}{2}x_1^2\right) \cdot (2\pi)^{-1/2}\exp\left(-\frac{1}{2}x_2^2\right)} \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{\left(x_1^2 + x_2^2 - 2\rho x_1 x_2\right)}{(1-\rho^2)} + \frac{1}{2}\left(x_1^2 + x_2^2\right)\right) \end{aligned}$$

where $x_1 = \mathbf{F}_1^{-1}(u_1)$ and $x_2 = \mathbf{F}_2^{-1}(u_2)$. It is then easy to generate bivariate non-normal distributions.

Example 111 *In Figure 11.1, we have built a bivariate probability distribution by considering that the marginals are an inverse Gaussian distribution and a beta distribution. The copula function corresponds to the Normal copula such that its Kendall's tau is equal to* 50%.

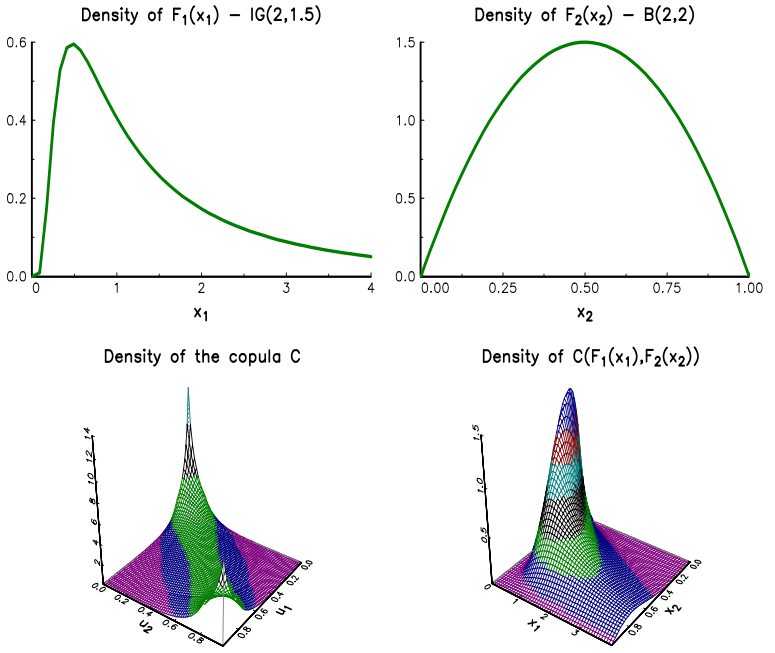


FIGURE 11.1: Example of a bivariate probability distribution with given marginals

11.1.3 Fréchet classes

The goal of Fréchet classes is to study the structure of the class of distributions with given marginals. These latter can be unidimensional, multidimensional or conditional. Let

us consider the bivariate distribution functions $\mathbf{F}_{12}$ and $\mathbf{F}_{23}$. The Fréchet class $\mathcal{F}(\mathbf{F}_{12}, \mathbf{F}_{23})$ is the set of trivariate probability distributions that are compatible with the two bivariate marginals $\mathbf{F}_{12}$ and $\mathbf{F}_{23}$. In this handbook, we restrict our focus on the Fréchet class $\mathcal{F}(\mathbf{F}_1, \ldots, \mathbf{F}_n)$ with univariate marginals.

11.1.3.1 The bivariate case

Let us first consider the bivariate case. The distribution function $\mathbf{F}$ belongs to the Fréchet class $(\mathbf{F}_1, \mathbf{F}_2)$ and we note $\mathbf{F} \in \mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ if and only if the marginals of $\mathbf{F}$ are $\mathbf{F}_1$ and $\mathbf{F}_2$, meaning that $\mathbf{F}(x_1, \infty) = \mathbf{F}_1(x_1)$ and $\mathbf{F}(\infty, x_2) = \mathbf{F}_2(x_2)$. Characterizing the Fréchet class $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ is then equivalent to find the set $\mathcal{C}$ of copula functions:

$$\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2) = \{\mathbf{F} : \mathbf{F}(x_1, x_2) = \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)), \mathbf{C} \in \mathcal{C}\}$$

Therefore this problem does not depend on the marginals $\mathbf{F}_1$ and $\mathbf{F}_2$.

We can show that the extremal distribution functions $\mathbf{F}^-$ and $\mathbf{F}^+$ of the Fréchet class $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ are:

$$\mathbf{F}^-(x_1, x_2) = \max(\mathbf{F}_1(x_1) + \mathbf{F}_2(x_2) - 1, 0)$$

and:

$$\mathbf{F}^+(x_1, x_2) = \min(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2))$$

$\mathbf{F}^-$ and $\mathbf{F}^+$ are called the Fréchet lower and upper bounds. We deduce that the corresponding copula functions are:

$$\mathbf{C}^-(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$$

and:

$$\mathbf{C}^+(u_1, u_2) = \min(u_1, u_2)$$

Example 112 *We consider the Fréchet class $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ where $\mathbf{F}_1 \sim \mathcal{N}(0,1)$ and $\mathbf{F}_2 \sim \mathcal{N}(0,1)$. We know that the bivariate normal distribution with correlation ρ belongs to $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$. Nevertheless, a lot of bivariate non-normal distributions are also in this Fréchet class. For instance, this is the case of this probability distribution:*

$$\mathbf{F}(x_1, x_2) = \frac{\Phi(x_1)\,\Phi(x_2)}{\Phi(x_1) + \Phi(x_2) - \Phi(x_1)\,\Phi(x_2)}$$

We can also show that[1]*:*

$$\mathbf{F}^-(x_1, x_2) := \Phi(x_1, x_2; -1) = \max(\Phi(x_1) + \Phi(x_2) - 1, 0)$$

and:

$$\mathbf{F}^+(x_1, x_2) := \Phi(x_1, x_2; +1) = \min(\Phi(x_1), \Phi(x_2))$$

Therefore, the bounds of the Fréchet class $\mathcal{F}(\mathcal{N}(0,1), \mathcal{N}(0,1))$ correspond to the bivariate normal distribution, whose correlation is respectively equal to -1 and $+1$.

[1] We recall that:

$$\Phi(x_1, x_2; \rho) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \phi(y_1, y_2; \rho)\, \mathrm{d}y_1\, \mathrm{d}y_2$$

11.1.3.2 The multivariate case

The extension of bivariate copulas to multivariate copulas is straightforward. Thus, the canonical decomposition of a multivariate distribution function is:

$$\mathbf{F}(x_1, \ldots, x_n) = \mathbf{C}(\mathbf{F}_1(x_1), \ldots, \mathbf{F}_n(x_n))$$

We note $\mathbf{C}_{\mathcal{E}}$ the sub-copula of $\mathbf{C}$ such that arguments that are not in the set $\mathcal{E}$ are equal to 1. For instance, with a dimension of 4, we have $\mathbf{C}_{12}(u, v) = \mathbf{C}(u, v, 1, 1)$ and $\mathbf{C}_{124}(u, v, w) = \mathbf{C}(u, v, 1, w)$. Let us consider the 2-copulas $\mathbf{C}_1$ and $\mathbf{C}_2$. It seems logical to build a copula of higher dimension with copulas of lower dimensions. In fact, the function $\mathbf{C}_1(u_1, \mathbf{C}_2(u_2, u_3))$ is not a copula in most cases (Quesada Molina and Rodríguez Lallena, 1994). For instance, we have:

$$\begin{aligned}
\mathbf{C}^-\left(u_1, \mathbf{C}^-(u_2, u_3)\right) &= \max(u_1 + \max(u_2 + u_3 - 1, 0) - 1, 0) \\
&= \max(u_1 + u_2 + u_3 - 2, 0) \\
&= \mathbf{C}^-(u_1, u_2, u_3)
\end{aligned}$$

However, the function $\mathbf{C}^-(u_1, u_2, u_3)$ is not a copula.

In the multivariate case, we define:

$$\mathbf{C}^-(u_1, \ldots, u_n) = \max\left(\sum_{i=1}^{n} u_i - n + 1, 0\right)$$

and:

$$\mathbf{C}^+(u_1, \ldots, u_n) = \min(u_1, \ldots, u_n)$$

As discussed above, we can show that $\mathbf{C}^+$ is a copula, but $\mathbf{C}^-$ does not belong to the set $\mathcal{C}$. Nevertheless, $\mathbf{C}^-$ is the best-possible bound, meaning that for all $(u_1, \ldots, u_n) \in [0, 1]^n$, there is a copula that coincide with $\mathbf{C}^-$ (Nelsen, 2006). This implies that $\mathcal{F}(\mathbf{F}_1, \ldots, \mathbf{F}_n)$ has a minimal distribution function if and only if $\max(\sum_{i=1}^{n} \mathbf{F}_i(x_i) - n + 1, 0)$ is a probability distribution (Dall'Aglio, 1972).

11.1.3.3 Concordance ordering

Using the result of the previous paragraph, we have:

$$\mathbf{C}^-(u_1, u_2) \leq \mathbf{C}(u_1, u_2) \leq \mathbf{C}^+(u_1, u_2)$$

for all $\mathbf{C} \in \mathcal{C}$. For a given value $\alpha \in [0, 1]$, the level curves of $\mathbf{C}$ are then in the triangle defined as follows:

$$\{(u_1, u_2) : \max(u_1 + u_2 - 1, 0) \leq \alpha, \min(u_1, u_2) \geq \alpha\}$$

An illustration is shown in Figure 11.2. In the multidimensional case, the region becomes a n-volume.

We now introduce a stochastic ordering on copulas. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two copula functions. We say that the copula $\mathbf{C}_1$ is smaller than the copula $\mathbf{C}_2$ and we note $\mathbf{C}_1 \prec \mathbf{C}_2$ if we verify that $\mathbf{C}_1(u_1, u_2) \leq \mathbf{C}_2(u_1, u_2)$ for all $(u_1, u_2) \in [0, 1]^2$. This stochastic ordering is called the concordance ordering and may be viewed as the first order of the stochastic dominance on probability distributions.

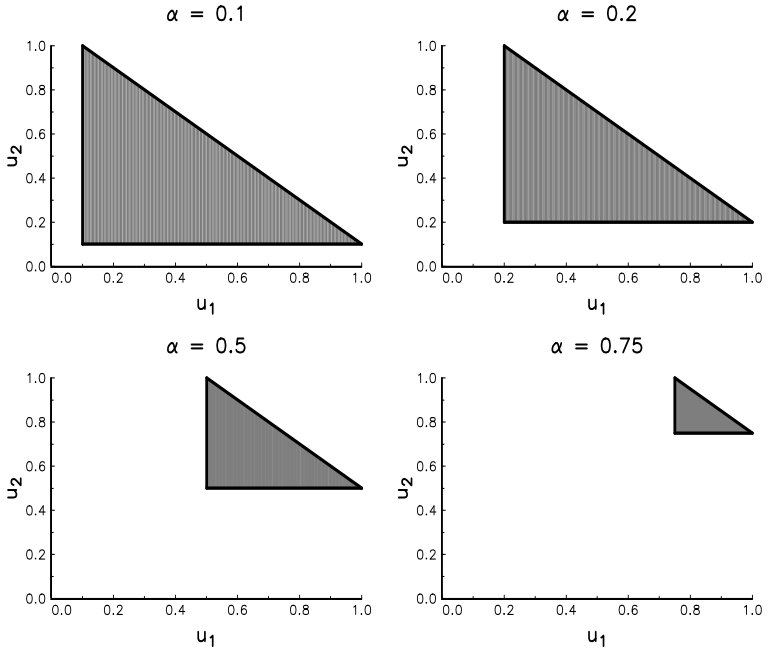


FIGURE 11.2: The triangle region of the contour lines $\mathbf{C}(u_1, u_2) = \alpha$

Example 113 *This ordering is partial because we cannot compare all copula functions. Let us consider the cubic copula defined by* $\mathbf{C}(u_1, u_2; \theta) = u_1 u_2 + \theta [u(u-1)(2u-1)] [v(v-1)(2v-1)]$ *where* $\theta \in [-1, 2]$. *If we compare it to the product copula* $\mathbf{C}^\perp$, *we have:*

$$\mathbf{C}\left(\frac{3}{4}, \frac{3}{4}; 1\right) = 0.5712 \geq \mathbf{C}^\perp\left(\frac{3}{4}, \frac{3}{4}\right) = 0.5625$$

but:

$$\mathbf{C}\left(\frac{3}{4}, \frac{1}{4}; 1\right) = 0.1787 \leq \mathbf{C}^\perp\left(\frac{3}{4}, \frac{1}{4}\right) = 0.1875$$

Using the Fréchet bounds, we always have $\mathbf{C}^- \prec \mathbf{C}^\perp \prec \mathbf{C}^+$. A copula $\mathbf{C}$ has a positive quadrant dependence (PQD) if it satisfies the inequality $\mathbf{C}^\perp \prec \mathbf{C} \prec \mathbf{C}^+$. In a similar way, $\mathbf{C}$ has a negative quadrant dependence (NQD) if it satisfies the inequality $\mathbf{C}^- \prec \mathbf{C} \prec \mathbf{C}^\perp$. As it is a partial ordering, there exist copula functions $\mathbf{C}$ such that $\mathbf{C} \not\succ \mathbf{C}^\perp$ and $\mathbf{C} \not\prec \mathbf{C}^\perp$. A copula function may then have a dependence structure that is neither positive or negative. This is the case of the cubic copula given in the previous example. In Figure 11.3, we report the cumulative distribution function (above panel) and its contour lines (right panel) of the three copula functions $\mathbf{C}^-$, $\mathbf{C}^\perp$ and $\mathbf{C}^+$, which plays an important role to understand the dependance between unidimensional risks.

Let $\mathbf{C}_\theta(u_1, u_2) = \mathbf{C}(u_1, u_2; \theta)$ be a family of copula functions that depends on the parameter θ. The copula family $\{\mathbf{C}_\theta\}$ is totally ordered if, for all $\theta_2 \geq \theta_1$, $\mathbf{C}_{\theta_2} \succ \mathbf{C}_{\theta_1}$ (positively ordered) or $\mathbf{C}_{\theta_2} \prec \mathbf{C}_{\theta_1}$ (negatively ordered). For instance, the Frank copula defined by:

$$\mathbf{C}(u_1, u_2; \theta) = -\frac{1}{\theta} \ln\left(1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right)$$

where $\theta \in \mathbb{R}$ is a positively ordered family (see Figure 11.4).

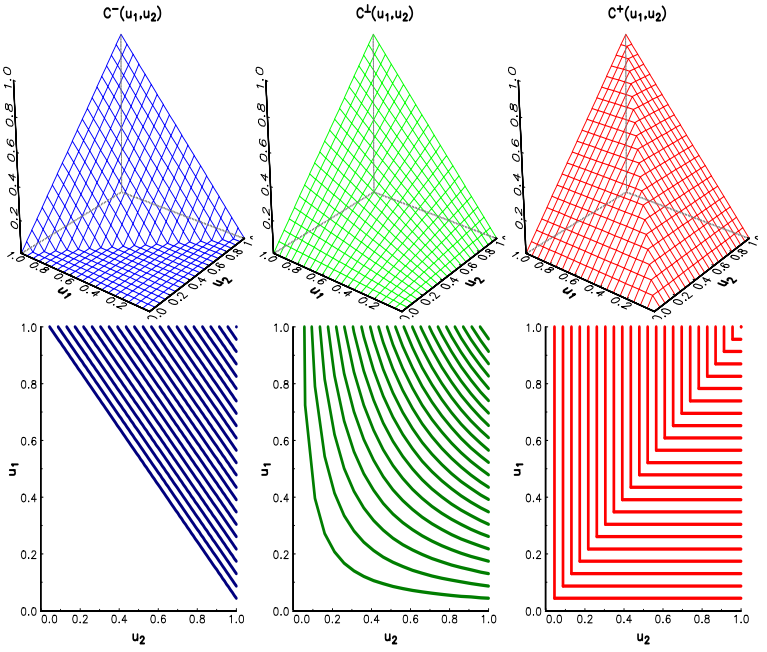


FIGURE 11.3: The three copula functions $\mathbf{C}^{-}$, $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$

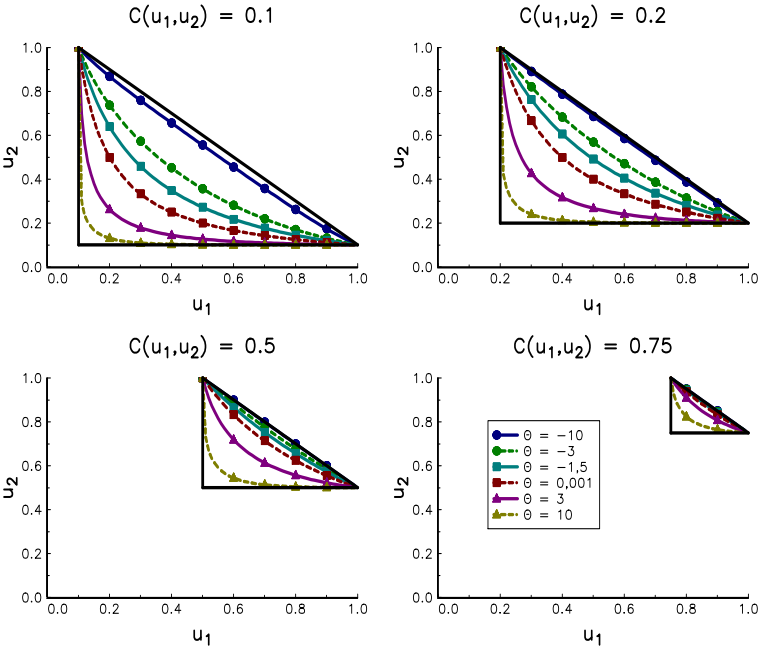


FIGURE 11.4: Concordance ordering of the Frank copula

Example 114 *Let us consider the copula function* $\mathbf{C}_\theta = \theta \cdot \mathbf{C}^- + (1-\theta) \cdot \mathbf{C}^+$ *where* $0 \leq \theta \leq 1$. *This copula is a convex sum of the extremal copulas* $\mathbf{C}^-$ *and* $\mathbf{C}^+$. *When* $\theta_2 \geq \theta_1$, *we have:*

$$\begin{aligned}
\mathbf{C}_{\theta_2}(u_1, u_2) &= \theta_2 \cdot \mathbf{C}^-(u_1, u_2) + (1-\theta_2) \cdot \mathbf{C}^+(u_1, u_2) \\
&= \mathbf{C}_{\theta_1}(u_1, u_2) - (\theta_2 - \theta_1) \cdot \left(\mathbf{C}^+(u_1, u_2) - \mathbf{C}^-(u_1, u_2)\right) \\
&\leq \mathbf{C}_{\theta_1}(u_1, u_2)
\end{aligned}$$

We deduce that $\mathbf{C}_{\theta_2} \prec \mathbf{C}_{\theta_1}$. *This copula family is negatively ordered.*

11.2 Copula functions and random vectors

Let $X = (X_1, X_2)$ be a random vector with distribution $\mathbf{F}$. We define the copula of (X_1, X_2) by the copula of $\mathbf{F}$:

$$\mathbf{F}(x_1, x_2) = \mathbf{C}\langle X_1, X_2 \rangle (\mathbf{F}_1(x_1), \mathbf{F}_2(x_2))$$

In what follows, we give the main results on the dependence of the random vector X found in Deheuvels (1978), Schweizer and Wolff (1981), and Nelsen (2006).

11.2.1 Countermonotonicity, comonotonicity and scale invariance property

We give here a probabilistic interpretation of the three copula functions $\mathbf{C}^-$, $\mathbf{C}^\perp$ and $\mathbf{C}^+$:

- X_1 and X_2 are countermonotonic – or $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^-$ – if there exists a random variable X such that $X_1 = f_1(X)$ and $X_2 = f_2(X)$ where f_1 and f_2 are respectively decreasing and increasing functions[2];
- X_1 and X_2 are independent if the dependence function is the product copula $\mathbf{C}^\perp$;
- X_1 are X_2 are comonotonic – or $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$ – if there exists a random variable X such that $X_1 = f_1(X)$ and $X_2 = f_2(X)$ where f_1 and f_2 are both increasing functions[3].

Let us consider a uniform random vector (U_1, U_2). We have $U_2 = 1 - U_1$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^-$ and $U_2 = U_1$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$. In the case of a standardized Gaussian random vector, we obtain $X_2 = -X_1$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^-$ and $X_2 = X_1$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$. If the marginals are log-normal, it follows that $X_2 = X_1^{-1}$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^-$ and $X_2 = X_1$ when $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$. For these three examples, we verify that X_2 is a decreasing (resp. increasing) function of X_1 if the copula function $\mathbf{C}\langle X_1, X_2 \rangle$ is $\mathbf{C}^-$ (resp. $\mathbf{C}^+$). The concepts of counter- and comonotonicity concepts generalize the cases where the linear correlation of a Gaussian vector is equal to -1 or $+1$. Indeed, $\mathbf{C}^-$ and $\mathbf{C}^+$ define respectively perfect negative and positive dependence.

[2] We also have $X_2 = f(X_1)$ where $f = f_2 \circ f_1^{-1}$ is a decreasing function.
[3] In this case, $X_2 = f(X_1)$ where $f = f_2 \circ f_1^{-1}$ is an increasing function.

We now give one of the most important theorems on copulas. Let (X_1, X_2) be a random vectors, whose copula is $\mathbf{C}\langle X_1, X_2\rangle$. If h_1 and h_2 are two increasing functions on $\operatorname{Im} X_1$ and $\operatorname{Im} X_2$, then we have:

$$\mathbf{C}\langle h_1(X_1), h_2(X_2)\rangle = \mathbf{C}\langle X_1, X_2\rangle$$

This means that copula functions are invariant under strictly increasing transformations of the random variables. To prove this theorem, we note $\mathbf{F}$ and $\mathbf{G}$ the probability distributions of the random vectors (X_1, X_2) and $(Y_1, Y_2) = (h_1(X_1), h_2(X_2))$. The marginals of $\mathbf{G}$ are:

$$\begin{aligned}
\mathbf{G}_1(y_1) &= \Pr\{Y_1 \leq y_1\} \\
&= \Pr\{h_1(X_1) \leq y_1\} \\
&= \Pr\left\{X_1 \leq h_1^{-1}(y_1)\right\} \quad \text{(because } h_1 \text{ is strictly increasing)} \\
&= \mathbf{F}_1\left(h_1^{-1}(y_1)\right)
\end{aligned}$$

and $\mathbf{G}_2(y_2) = \mathbf{F}_2\left(h_2^{-1}(y_2)\right)$. We deduce that $\mathbf{G}_1^{-1}(u_1) = h_1\left(\mathbf{F}_1^{-1}(u_1)\right)$ and $\mathbf{G}_2^{-1}(u_2) = h_2\left(\mathbf{F}_2^{-1}(u_2)\right)$. By definition, we have:

$$\mathbf{C}\langle Y_1, Y_2\rangle(u_1, u_2) = \mathbf{G}\left(\mathbf{G}_1^{-1}(u_1), \mathbf{G}_2^{-1}(u_2)\right)$$

Moreover, it follows that:

$$\begin{aligned}
\mathbf{G}\left(\mathbf{G}_1^{-1}(u_1), \mathbf{G}_2^{-1}(u_2)\right) &= \Pr\left\{Y_1 \leq \mathbf{G}_1^{-1}(u_1), Y_2 \leq \mathbf{G}_2^{-1}(u_2)\right\} \\
&= \Pr\left\{h_1(X_1) \leq \mathbf{G}_1^{-1}(u_1), h_2(X_2) \leq \mathbf{G}_2^{-1}(u_2)\right\} \\
&= \Pr\left\{X_1 \leq h_1^{-1}\left(\mathbf{G}_1^{-1}(u_1)\right), X_2 \leq h_2^{-1}\left(\mathbf{G}_2^{-1}(u_2)\right)\right\} \\
&= \Pr\left\{X_1 \leq \mathbf{F}_1^{-1}(u_1), X_2 \leq \mathbf{F}_2^{-1}(u_2)\right\} \\
&= \mathbf{F}\left(\mathbf{F}_1^{-1}(u_1), \mathbf{F}_2^{-1}(u_2)\right)
\end{aligned}$$

Because we have $\mathbf{C}\langle X_1, X_2\rangle(u_1, u_2) = \mathbf{F}\left(\mathbf{F}_1^{-1}(u_1), \mathbf{F}_2^{-1}(u_2)\right)$, we deduce that $\mathbf{C}\langle Y_1, Y_2\rangle = \mathbf{C}\langle X_1, X_2\rangle$.

Example 115 *If X_1 and X_2 are two positive random variables, the previous theorem implies that:*

$$\begin{aligned}
\mathbf{C}\langle X_1, X_2\rangle &= \mathbf{C}\langle \ln X_1, X_2\rangle \\
&= \mathbf{C}\langle \ln X_1, \ln X_2\rangle \\
&= \mathbf{C}\langle X_1, \exp X_2\rangle \\
&= \mathbf{C}\left\langle \sqrt{X_1}, \exp X_2\right\rangle
\end{aligned}$$

Applying an increasing transformation does not change the copula function, only the marginals. Thus, the copula of the multivariate log-normal distribution is the same than the copula of the multivariate normal distribution.

The scale invariance property is perhaps not surprising if we consider the canonical decomposition of the bivariate probability distribution. Indeed, the copula $\mathbf{C}\langle U_1, U_2\rangle$ is equal to the copula $\mathbf{C}\langle X_1, X_2\rangle$ where $U_1 = \mathbf{F}_1(X_1)$ and $U_2 = \mathbf{F}_2(X_2)$. In some sense, Sklar's theorem is an application of the scale invariance property by considering $h_1(x_1) = \mathbf{F}_1(x_1)$ and $h_2(x_2) = \mathbf{F}_2(x_2)$.

Example 116 *We assume that $X_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ and $X_2 \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$. If the copula of (X_1, X_2) is $\mathbf{C}^-$, we have $U_2 = 1 - U_1$. This implies that:*

$$\begin{aligned}
\Phi\left(\frac{X_2 - \mu_2}{\sigma_2}\right) &= 1 - \Phi\left(\frac{X_1 - \mu_1}{\sigma_1}\right) \\
&= \Phi\left(-\frac{X_1 - \mu_1}{\sigma_1}\right)
\end{aligned}$$

We deduce that X_1 and X_2 are countermonotonic if:

$$X_2 = \mu_2 - \frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)$$

By applying the same reasoning to the copula function $\mathbf{C}^+$, we show that X_1 and X_2 are comonotonic if:

$$X_2 = \mu_2 + \frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)$$

We now consider the log-normal random variables $Y_1 = \exp(X_1)$ and $Y_2 = \exp(X_2)$. For the countermonotonicity case, we obtain:

$$\ln Y_2 = \mu_2 - \frac{\sigma_2}{\sigma_1}(\ln Y_1 - \mu_1)$$

or:

$$Y_2 = \exp\left(\mu_2 + \frac{\sigma_2}{\sigma_1}\mu_1\right) \cdot Y_1^{-\sigma_2/\sigma_1}$$

For the comonotonicity case, the relationship becomes:

$$Y_2 = \exp\left(\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1\right) \cdot Y_1^{\sigma_2/\sigma_1}$$

If we assume that $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, the log-normal random variables Y_1 and Y_2 are countermonotonic if $Y_2 = Y_1^{-1}$ and comonotonic if $Y_2 = Y_1$.

11.2.2 Dependence measures

We can interpret the copula function $\mathbf{C}\langle X_1, X_2\rangle$ as a standardization of the joint distribution after eliminating the effects of marginals. Indeed, it is a comprehensive statistic of the dependence function between X_1 and X_2. Therefore, a non-comprehensive statistic will be a dependence measure if it can be expressed using $\mathbf{C}\langle X_1, X_2\rangle$.

11.2.2.1 Concordance measures

Following Nelsen (2006), a numeric measure m of association between X_1 and X_2 is a measure of concordance if it satisfies the following properties:

1. $-1 = m\langle X, -X\rangle \leq m\langle \mathbf{C}\rangle \leq m\langle X, X\rangle = 1$;
2. $m\left\langle \mathbf{C}^{\perp}\right\rangle = 0$;
3. $m\langle -X_1, X_2\rangle = m\langle X_1, -X_2\rangle = -m\langle X_1, X_2\rangle$;
4. if $\mathbf{C}_1 \prec \mathbf{C}_2$, then $m\langle \mathbf{C}_1\rangle \leq m\langle \mathbf{C}_2\rangle$;

Using this last property, we have: $\mathbf{C} \prec \mathbf{C}^{\perp} \Rightarrow m\langle \mathbf{C}\rangle < 0$ and $\mathbf{C} \succ \mathbf{C}^{\perp} \Rightarrow m\langle \mathbf{C}\rangle > 0$. The concordance measure can then be viewed as a generalization of the linear correlation when the dependence function is not normal. Indeed, a positive quadrant dependence (PQD) copula will have a positive concordance measure whereas a negative quadrant dependence (NQD) copula will have a negative concordance measure. Moreover, the bounds -1 and $+1$ are reached when the copula function is countermonotonic and comonotonic.

Among the several concordance measures, we find Kendall's tau and Spearman's rho, which play an important role in non-parametric statistics. Let us consider a sample of n observations $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ of the random vector (X, Y). Kendall's tau is the

probability of concordance – $(X_i - X_j)\cdot(Y_i - Y_j) > 0$ – minus the probability of discordance – $(X_i - X_j)\cdot(Y_i - Y_j) < 0$:

$$\tau = \Pr\{(X_i - X_j)\cdot(Y_i - Y_j) > 0\} - \Pr\{(X_i - X_j)\cdot(Y_i - Y_j) < 0\}$$

Spearman's rho is the linear correlation of the rank statistics $(X_{i:n}, Y_{i:n})$. We can also show that Spearman's rho has the following expression:

$$\varrho = \frac{\operatorname{cov}(\mathbf{F}_X(X), \mathbf{F}_Y(Y))}{\sigma(\mathbf{F}_X(X))\cdot\sigma(\mathbf{F}_Y(Y))}$$

Schweizer and Wolff (1981) showed that Kendall's tau and Spearman's rho are concordance measures and have the following expressions:

$$\begin{aligned}\tau &= 4\iint_{[0,1]^2}\mathbf{C}(u_1,u_2)\,\mathrm{d}\mathbf{C}(u_1,u_2) - 1\\ \varrho &= 12\iint_{[0,1]^2}u_1u_2\,\mathrm{d}\mathbf{C}(u_1,u_2) - 3\end{aligned}$$

From a numerical point of view, the following formulas should be preferred (Nelsen, 2006):

$$\begin{aligned}\tau &= 1 - 4\iint_{[0,1]^2}\partial_{u_1}\mathbf{C}(u_1,u_2)\,\partial_{u_2}\mathbf{C}(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2\\ \varrho &= 12\iint_{[0,1]^2}\mathbf{C}(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2 - 3\end{aligned}$$

For some copulas, we have analytical formulas. For instance, we have:

Copula	ϱ	τ
Normal	$6\pi^{-1}\arcsin(\rho/2)$	$2\pi^{-1}\arcsin(\rho)$
Gumbel	✓	$(\theta-1)/\theta$
FGM	$\theta/3$	$2\theta/9$
Frank	$1 - 12\theta^{-1}(\mathbf{D}_1(\theta) - \mathbf{D}_2(\theta))$	$1 - 4\theta^{-1}(1 - \mathbf{D}_1(\theta))$

where $\mathbf{D}_k(x)$ is the Debye function. The Gumbel (or Gumbel-Hougaard) copula is equal to:

$$\mathbf{C}(u_1,u_2;\theta) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right)$$

for $\theta \geq 1$, whereas the expression of the Farlie-Gumbel-Morgenstern (or FGM) copula is:

$$\mathbf{C}(u_1,u_2;\theta) = u_1u_2(1 + \theta(1-u_1)(1-u_2))$$

for $-1 \leq \theta \leq 1$.

For illustration, we report in Figures 11.5, 11.6 and 11.7 the level curves of several density functions built with Normal, Frank and Gumbel copulas. In order to compare them, the parameter of each copula is calibrated such that Kendall's tau is equal to 50%. This means that these 12 distributions functions have the same dependence with respect to Kendall's tau. However, the dependence is different from one figure to another, because their copula function is not the same. This is why Kendall's tau is not an exhaustive statistic of the dependence between two random variables.

We could build bivariate probability distributions, which are even less comparable. Indeed, the set of these three copula families (Normal, Frank and Gumbel) is very small

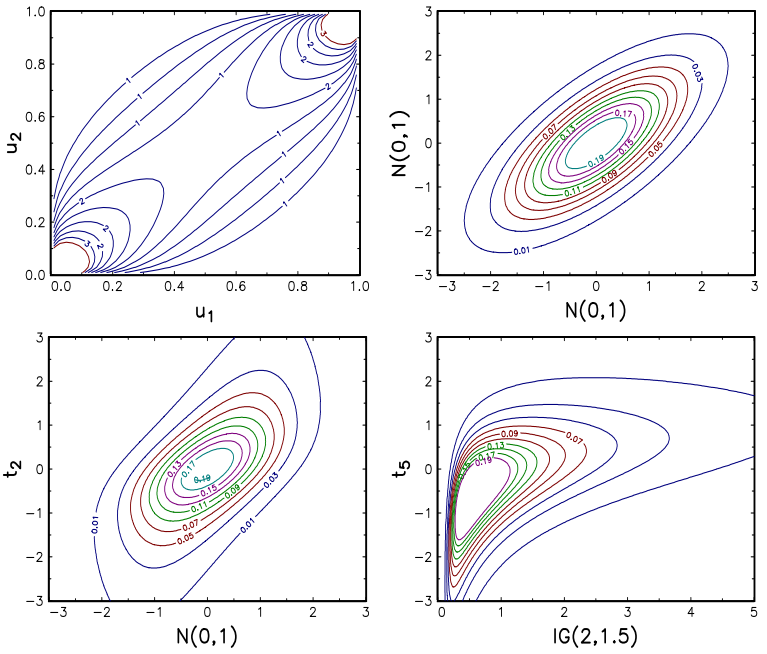


FIGURE 11.5: Contour lines of bivariate densities (Normal copula)

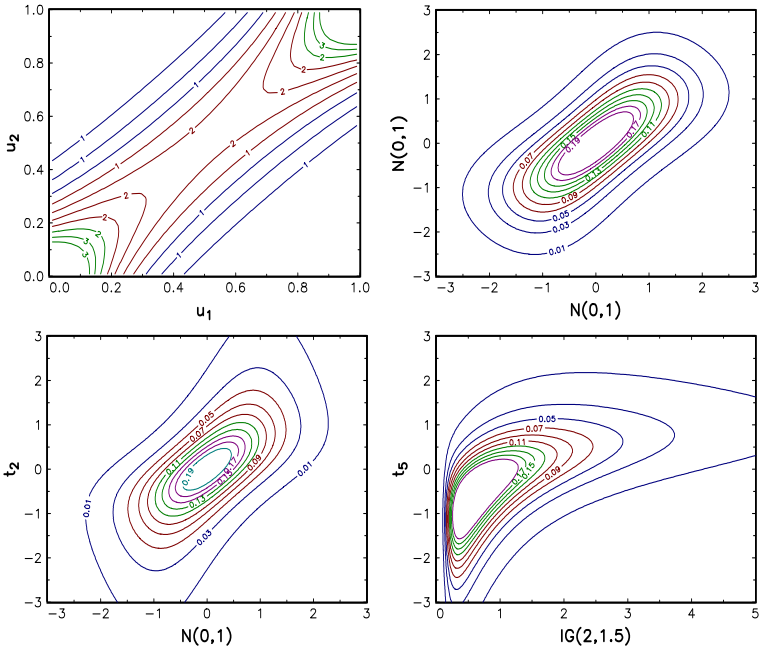


FIGURE 11.6: Contour lines of bivariate densities (Frank copula)

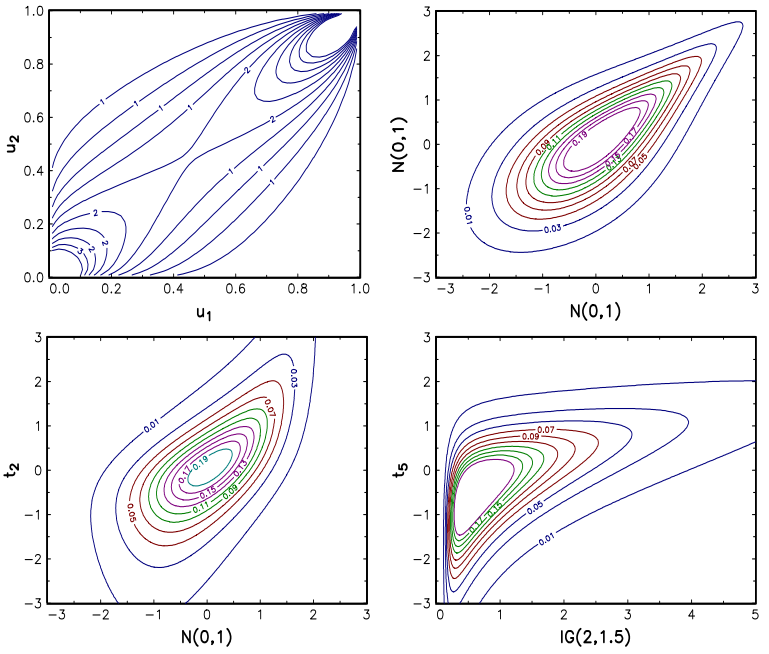


FIGURE 11.7: Contour lines of bivariate densities (Gumbel copula)

compared to the set $\mathcal{C}$ of copulas. However, there exist other dependence functions that are very far from the previous copulas. For instance, we consider the region $\mathcal{B}(\tau, \varrho)$ defined by:

$$(\tau, \varrho) \in \mathcal{B}(\tau, \varrho) \Leftrightarrow \begin{cases} (3\tau - 1)/2 \leq \varrho \leq \left(1 + 2\tau - \tau^2\right)/2 & \text{if} \quad \tau \geq 0 \\ \left(\tau^2 + 2\tau - 1\right)/2 \leq \varrho \leq (1 + 3\tau)/2 & \text{if} \quad \tau \leq 0 \end{cases}$$

Nelsen (2006) shows that these bounds cannot be improved and there is always a copula function that corresponds to a point of the boundary $\mathcal{B}(\tau, \varrho)$. In Figure 11.8 we report the bounds $\mathcal{B}(\tau, \varrho)$ and the area reached by 8 copula families (Normal, Plackett, Frank, Clayton, Gumbel, Galambos, Hüsler-Reiss, FGM). These copulas covered a small surface of the $\tau - \varrho$ region. These copula families are then relatively similar if we consider these concordance measures. Obtaining copulas that have a different behavior requires that the dependence is not monotone[4] on the whole domain $[0, 1]^2$.

11.2.2.2 Linear correlation

We recall that the linear correlation (or Pearson's correlation) is defined as follows:

$$\rho\langle X_1, X_2\rangle = \frac{\mathbb{E}[X_1 \cdot X_2] - \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]}{\sigma(X_1) \cdot \sigma(X_2)}$$

Tchen (1980) showed the following properties of this measure:

- if the dependence of the random vector (X_1, X_2) is the product copula $\mathbf{C}^{\perp}$, then $\rho\langle X_1, X_2\rangle = 0$;
- ρ is an increasing function with respect to the concordance measure:

$$\mathbf{C}_1 \succ \mathbf{C}_2 \Rightarrow \rho_1\langle X_1, X_2\rangle \geq \rho_2\langle X_1, X_2\rangle$$

[4] For instance, the dependence can be positive in one region and negative in another region.

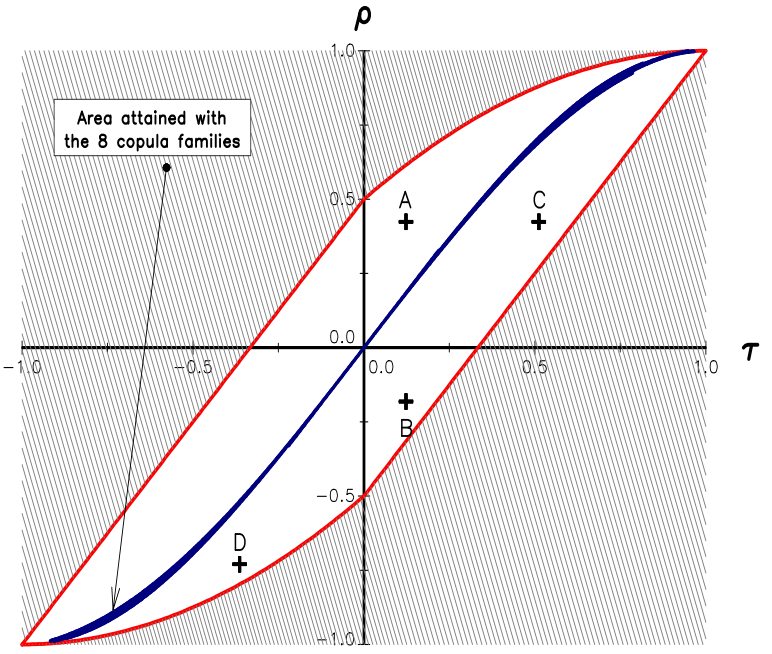


FIGURE 11.8: Bounds of (τ, ϱ) statistics

- $\rho\langle X_1, X_2\rangle$ is bounded:

$$\rho^-\langle X_1, X_2\rangle \leq \rho\langle X_1, X_2\rangle \leq \rho^+\langle X_1, X_2\rangle$$

  and the bounds are reached for the Fréchet copulas $\mathbf{C}^-$ and $\mathbf{C}^+$.

However, a concordance measure must satisfy $m\langle\mathbf{C}^-\rangle = -1$ and $m\langle\mathbf{C}^+\rangle = +1$. If we use the stochastic representation of Fréchet bounds, we have:

$$\rho^-\langle X_1, X_2\rangle = \rho^+\langle X_1, X_2\rangle = \frac{\mathbb{E}[f_1(X)\cdot f_2(X)] - \mathbb{E}[f_1(X)]\cdot\mathbb{E}[f_2(X)]}{\sigma(f_1(X))\cdot\sigma(f_2(X))}$$

The solution of the equation $\rho^-\langle X_1, X_2\rangle = -1$ is $f_1(x) = a_1x + b_1$ and $f_2(x) = a_2x + b_2$ where $a_1a_2 < 0$. For the equation $\rho^+\langle X_1, X_2\rangle = +1$, the condition becomes $a_1a_2 > 0$. Except for Gaussian random variables, there are few probability distributions that can satisfy these conditions. Moreover, if the linear correlation is a concordance measure, it is an invariant measure by increasing transformations:

$$\rho\langle X_1, X_2\rangle = \rho\langle f_1(X_1), f_2(X_2)\rangle$$

Again, the solution of this equation is $f_1(x) = a_1x + b_1$ and $f_2(x) = a_2x + b_2$ where $a_1a_2 > 0$. We now have a better understanding why we say that this dependence measure is linear. In summary, the copula function generalizes the concept of linear correlation in a non-Gaussian non-linear world.

Example 117 *We consider the bivariate log-normal random vector* (X_1, X_2) *where* $X_1 \sim \mathcal{LN}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{LN}(\mu_2, \sigma_2^2)$ *and* $\rho = \rho\langle\ln X_1, \ln X_2\rangle$.

We can show that:

$$\mathbb{E}[X_1^{p_1}\cdot X_2^{p_2}] = \exp\left(p_1\mu_1 + p_2\mu_2 + \frac{p_1^2\sigma_1^2 + p_2^2\sigma_2^2}{2} + p_1p_2\rho\sigma_1\sigma_2\right)$$

It follows that:

$$\rho\left[X_1, X_2\right] = \frac{\exp\left(\rho\sigma_1\sigma_2\right) - 1}{\sqrt{\exp\left(\sigma_1^2\right) - 1} \cdot \sqrt{\exp\left(\sigma_2^2\right) - 1}}$$

We deduce that $\rho\left[X_1, X_2\right] \in [\rho^-, \rho^+]$, but the bounds are not necessarily -1 and $+1$. For instance, when we use the parameters $\sigma_1 = 1$ and $\sigma_2 = 3$, we obtain the following results:

Copula	$\rho\left[X_1, X_2\right]$	$\tau\left[X_1, X_2\right]$	$\varrho\left[X_1, X_2\right]$
$\mathbf{C}^-$	-0.008	-1.000	-1.000
$\rho = -0.7$	-0.007	-0.494	-0.683
$\mathbf{C}^\perp$	0.000	0.000	0.000
$\rho = 0.7$	0.061	0.494	0.683
$\mathbf{C}^+$	0.162	1.000	1.000

When the copula function is $\mathbf{C}^-$, the linear correlation takes a value close to zero! In Figure 11.9, we show that the bounds ρ^- and ρ^+ of $\rho\left[X_1, X_2\right]$ are not necessarily -1 and $+1$. When the marginals are log-normal, the upper bound $\rho^+ = +1$ is reached only when $\sigma_1 = \sigma_2$ and the lower bound $\rho^- = -1$ is never reached. This poses a problem to interpret the value of a correlation. Let us consider two random vectors (X_1, X_2) and (Y_1, Y_2). What could we say about the dependence function when $\rho\left[X_1, X_2\right] \geq \rho\left[Y_1, Y_2\right]$? The answer is nothing if the marginals are not Gaussian. Indeed, we have seen previously that a 70% linear correlation between two Gaussian random vectors becomes a 6% linear correlation if we apply an exponential transformation. However, the two copulas of (X_1, X_2) and (Y_1, Y_2) are exactly the same. In fact, the drawback of the linear correlation is that this measure depends on the marginals and not only on the copula function.

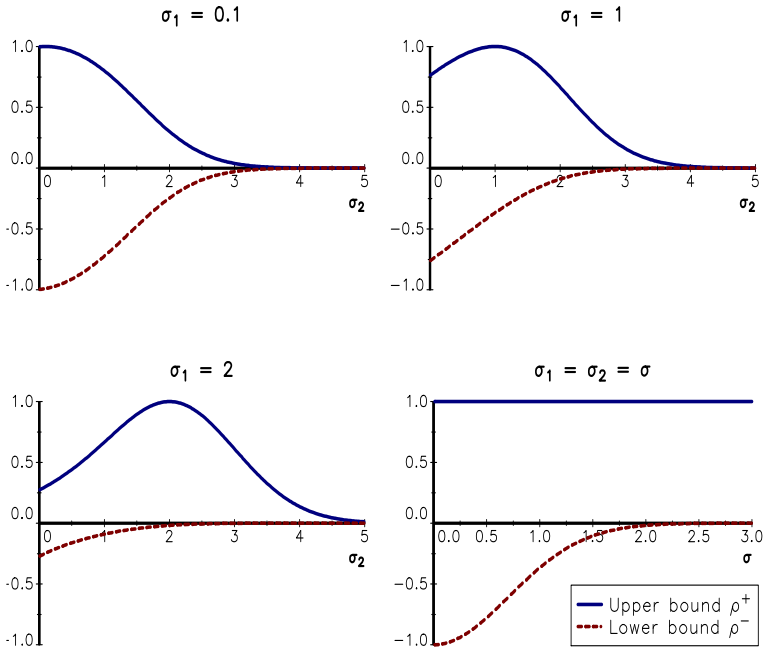


FIGURE 11.9: Bounds of the linear correlation between two log-normal random variables

11.2.2.3 Tail dependence

Contrary to concordance measures, tail dependence is a local measure that characterizes the joint behavior of the random variables X_1 and X_2 at the extreme points $x^- = \inf\{x : \mathbf{F}(x) > 0\}$ and $x^+ = \sup\{x : \mathbf{F}(x) < 1\}$. Let $\mathbf{C}$ be a copula function such that the following limit exists:

$$\lambda^+ = \lim_{u \to 1^-} \frac{1 - 2u + \mathbf{C}(u,u)}{1-u}$$

We say that $\mathbf{C}$ has an upper tail dependence when $\lambda^+ \in (0,1]$ and $\mathbf{C}$ has no upper tail dependence when $\lambda^+ = 0$ (Joe, 1997). For the lower tail dependence λ^-, the limit becomes:

$$\lambda^- = \lim_{u \to 0^+} \frac{\mathbf{C}(u,u)}{u}$$

We notice that λ^+ and λ^- can also be defined as follows:

$$\lambda^+ = \lim_{u \to 1^-} \Pr\{U_2 > u \mid U_1 > u\}$$

and:

$$\lambda^- = \lim_{u \to 0^+} \Pr\{U_2 < u \mid U_1 < u\}$$

To compute the upper tail dependence, we consider the joint survival function $\bar{\mathbf{C}}$ defined by:

$$\begin{aligned} \bar{\mathbf{C}}(u_1,u_2) &= \Pr\{U_1 > u_1, U_2 > u_2\} \\ &= 1 - u_1 - u_2 + \mathbf{C}(u_1,u_2) \end{aligned}$$

The expression of the upper tail dependence is then equal to:

$$\begin{aligned} \lambda^+ &= \lim_{u \to 1^-} \frac{\bar{\mathbf{C}}(u,u)}{1-u} \\ &= -\lim_{u \to 1^-} \frac{\mathrm{d}\bar{\mathbf{C}}(u,u)}{\mathrm{d}u} \\ &= -\lim_{u \to 1^-} \left(-2 + \partial_1 \mathbf{C}(u,u) + \partial_2 \mathbf{C}(u,u)\right) \\ &= \lim_{u \to 1^-} \left(\Pr\{U_2 > u \mid U_1 = u\} + \Pr\{U_1 > u \mid U_2 = u\}\right) \end{aligned}$$

By assuming that the copula is symmetric, we finally obtain:

$$\begin{aligned} \lambda^+ &= 2 \lim_{u \to 1^-} \Pr\{U_2 > u \mid U_1 = u\} \\ &= 2 - 2 \lim_{u \to 1^-} \Pr\{U_2 < u \mid U_1 = u\} \\ &= 2 - 2 \lim_{u \to 1^-} \mathbf{C}_{2|1}(u,u) \end{aligned} \tag{11.4}$$

In a similar way, we find that the lower tail dependence of a symmetric copula is equal to:

$$\lambda^- = 2 \lim_{u \to 0^+} \mathbf{C}_{2|1}(u,u) \tag{11.5}$$

For the copula functions $\mathbf{C}^-$ and $\mathbf{C}^\perp$, we have $\lambda^- = \lambda^+ = 0$. For the copula $\mathbf{C}^+$, we obtain $\lambda^- = \lambda^+ = 1$. However, there exist copulas such that $\lambda^- \neq \lambda^+$. This is the case of the

Gumbel copula $\mathbf{C}(u_1, u_2; \theta) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right)$, because we have $\lambda^- = 0$ and $\lambda^+ = 2 - 2^{1/\theta}$. The Gumbel copula has then an upper tail dependence, but no lower tail dependence. If we consider the Clayton copula $\mathbf{C}(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$, we obtain $\lambda^- = 2^{-1/\theta}$ and $\lambda^+ = 0$.

Coles *et al.* (1999) define the quantile-quantile dependence function as follows:

$$\lambda^+(\alpha) = \Pr\left\{X_2 > \mathbf{F}_2^{-1}(\alpha) \mid X_1 > \mathbf{F}_1^{-1}(\alpha)\right\}$$

It is the conditional probability that X_2 is larger than the quantile $\mathbf{F}_2^{-1}(\alpha)$ given that X_1 is larger than the quantile $\mathbf{F}_1^{-1}(\alpha)$. We have:

$$\begin{aligned}
\lambda^+(\alpha) &= \Pr\left\{X_2 > \mathbf{F}_2^{-1}(\alpha) \mid X_1 > \mathbf{F}_1^{-1}(\alpha)\right\} \\
&= \frac{\Pr\left\{X_2 > \mathbf{F}_2^{-1}(\alpha), X_1 > \mathbf{F}_1^{-1}(\alpha)\right\}}{\Pr\left\{X_1 > \mathbf{F}_1^{-1}(\alpha)\right\}} \\
&= \frac{1 - \Pr\left\{X_1 \leq \mathbf{F}_1^{-1}(\alpha)\right\} - \Pr\left\{X_2 \leq \mathbf{F}_2^{-1}(\alpha)\right\}}{1 - \Pr\left\{X_1 \leq \mathbf{F}_1^{-1}(\alpha)\right\}} + \\
&\quad \frac{\Pr\left\{X_2 \leq \mathbf{F}_2^{-1}(\alpha), X_1 \leq \mathbf{F}_1^{-1}(\alpha)\right\}}{1 - \Pr\left\{\mathbf{F}_1(X_1) \leq \alpha\right\}} \\
&= \frac{1 - 2\alpha + \mathbf{C}(\alpha, \alpha)}{1 - \alpha}
\end{aligned}$$

The tail dependence λ^+ is then the limit of the conditional probability $\lambda^+(\alpha)$ when the confidence level α tends to 1. It is also the probability of one variable being extreme given that the other is extreme. Because $\lambda^+(\alpha)$ is a probability, we verify that $\lambda^+ \in [0, 1]$. If the probability is zero, the extremes are independent. If λ^+ is equal to 1, the extremes are perfectly dependent. To illustrate the measures[5] $\lambda^+(\alpha)$ and $\lambda^-(\alpha)$, we represent their values for the Gumbel and Clayton copulas in Figure 11.10. The parameters are calibrated with respect to Kendall's tau.

Remark 138 *We consider two portfolios, whose losses correspond to the random variables L_1 and L_2 with probability distributions $\mathbf{F}_1$ and $\mathbf{F}_2$. The probability that the loss of the second portfolio is larger than its value-at-risk knowing that the value-at-risk of the first portfolio is exceeded is exactly equal to the quantile-quantile dependence measure $\lambda^+(\alpha)$:*

$$\begin{aligned}
\lambda^+(\alpha) &= \Pr\left\{L_2 > \mathbf{F}_2^{-1}(\alpha) \mid L_1 > \mathbf{F}_1^{-1}(\alpha)\right\} \\
&= \Pr\left\{L_2 > \mathrm{VaR}_\alpha(L_2) \mid L_1 > \mathrm{VaR}_\alpha(L_1)\right\}
\end{aligned}$$

11.3 Parametric copula functions

In this section, we study the copula families, which are commonly used in risk management. They are parametric copulas, which depend on a set of parameters. Statistical inference, in particular parameter estimation, is developed in the next section.

[5] We have $\lambda^-(\alpha) = \Pr\left\{X_2 < \mathbf{F}_2^{-1}(\alpha) \mid X_1 < \mathbf{F}_1^{-1}(\alpha)\right\}$ and $\lim_{\alpha \to 0} \lambda^-(\alpha) = \lambda^-$.

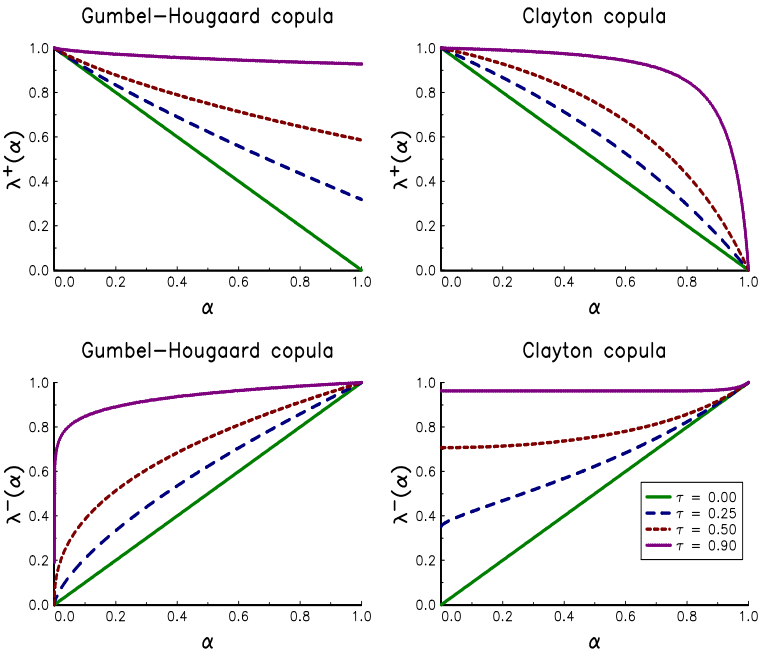


FIGURE 11.10: Quantile-quantile dependence measures $\lambda^{+}(\alpha)$ and $\lambda^{-}(\alpha)$

11.3.1 Archimedean copulas

11.3.1.1 Definition

Genest and MacKay (1986b) define Archimedean copulas as follows:

$$\mathbf{C}(u_1, u_2) = \begin{cases} \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) & \text{if } \varphi(u_1) + \varphi(u_2) \leq \varphi(0) \\ 0 & \text{otherwise} \end{cases}$$

where φ a C^2 is a function which satisfies $\varphi(1) = 0$, $\varphi'(u) < 0$ and $\varphi''(u) > 0$ for all $u \in [0,1]$. $\varphi(u)$ is called the generator of the copula function. If $\varphi(0) = \infty$, the generator is said to be strict. Genest and MacKay (1986a) link the construction of Archimedean copulas to the independence of random variables. Indeed, by considering the multiplicative generator $\lambda(u) = \exp(-\varphi(u))$, the authors show that:

$$\mathbf{C}(u_1, u_2) = \lambda^{-1}(\lambda(u_1)\lambda(u_2))$$

This means that:

$$\lambda(\Pr\{U_1 \leq u_1, U_2 \leq u_2\}) = \lambda(\Pr\{U_1 \leq u_1\}) \times \lambda(\Pr\{U_2 \leq u_2\})$$

In this case, the random variables (U_1, U_2) become independent when the scale of probabilities has been transformed.

Example 118 *If* $\varphi(u) = u^{-1} - 1$, *we have* $\varphi^{-1}(u) = (1+u)^{-1}$ *and:*

$$\mathbf{C}(u_1, u_2) = \left(1 + \left(u_1^{-1} - 1 + u_2^{-1} - 1\right)\right)^{-1} = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}$$

The Gumbel logistic copula is then an Archimedean copula.

Example 119 *The product copula* $\mathbf{C}^{\perp}$ *is Archimedean and the associated generator is* $\varphi(u) = -\ln u$. *Concerning Fréchet copulas, only* $\mathbf{C}^{-}$ *is Archimedean with* $\varphi(u) = 1 - u$.

In Table 11.1, we provide another examples of Archimedean copulas[6].

TABLE 11.1: Archimedean copula functions

Copula	$\varphi(u)$	$\mathbf{C}(u_1, u_2)$
$\mathbf{C}^{\perp}$	$-\ln u$	$u_1 u_2$
Clayton	$u^{-\theta} - 1$	$\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$
Frank	$-\ln \dfrac{e^{-\theta u} - 1}{e^{-\theta} - 1}$	$-\dfrac{1}{\theta} \ln \left(1 + \dfrac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1}\right)$
Gumbel	$(-\ln u)^{\theta}$	$\exp\left(-\left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta}\right)$
Joe	$-\ln\left(1 - (1-u)^{\theta}\right)$	$1 - \left(\bar{u}_1^{\theta} + \bar{u}_2^{\theta} - \bar{u}_1^{\theta}\bar{u}_2^{\theta}\right)^{1/\theta}$

11.3.1.2 Properties

Archimedean copulas play an important role in statistics, because they present many interesting properties, for example:

- $\mathbf{C}$ is symmetric, meaning that $\mathbf{C}(u_1, u_2) = \mathbf{C}(u_2, u_1)$;
- $\mathbf{C}$ is associative, implying that $\mathbf{C}(u_1, \mathbf{C}(u_1, u_3)) = \mathbf{C}(\mathbf{C}(u_1, u_2), u_3)$;
- the diagonal section $\delta(u) = \mathbf{C}(u, u)$ satisfies $\delta(u) < u$ for all $u \in (0, 1)$;
- if a copula $\mathbf{C}$ is associative and $\delta(u) < u$ for all $u \in (0, 1)$, then $\mathbf{C}$ is Archimedean.

Genest and MacKay (1986a) also showed that the expression of Kendall's tau is:

$$\tau\langle \mathbf{C} \rangle = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} \, \mathrm{d}u$$

whereas the copula density is:

$$c(u_1, u_2) = -\frac{\varphi''(\mathbf{C}(u_1, u_2))\, \varphi'(u_1)\, \varphi'(u_2)}{\left[\varphi'(\mathbf{C}(u_1, u_2))\right]^3}$$

Example 120 *With the Clayton copula, we have* $\varphi(u) = u^{-\theta} - 1$ *and* $\varphi'(u) = -\theta u^{-\theta-1}$. *We deduce that:*

$$\begin{aligned} \tau &= 1 + 4 \int_0^1 \frac{1 - u^{-\theta}}{\theta u^{-\theta-1}} \, \mathrm{d}u \\ &= \frac{\theta}{\theta + 2} \end{aligned}$$

[6] We use the notations $\bar{u} = 1 - u$ and $\tilde{u} = -\ln u$.

11.3.1.3 Two-parameter Archimedean copulas

Nelsen (2006) showed that if $\varphi(t)$ is a strict generator, then we can build two-parameter Archimedean copulas by considering the following generator:

$$\varphi_{\alpha,\beta}(t) = (\varphi(t^{\alpha}))^{\beta}$$

where $\alpha > 0$ and $\beta > 1$. For instance, if $\varphi(t) = t^{-1} - 1$, the two-parameter generator is $\varphi_{\alpha,\beta}(t) = (t^{-\alpha} - 1)^{\beta}$. Therefore, the corresponding copula function is defined by:

$$\mathbf{C}(u_1, u_2) = \left(\left[\left(u_1^{-\alpha} - 1\right)^{\beta} + \left(u_2^{-\alpha} - 1\right)^{\beta} \right]^{1/\beta} + 1 \right)^{-1/\alpha}$$

This is a generalization of the Clayton copula, which is obtained when the parameter β is equal to 1.

11.3.1.4 Extension to the multivariate case

We can build multivariate Archimedean copulas in the following way:

$$\mathbf{C}(u_1, \ldots, u_n) = \varphi^{-1}(\varphi(u_1) + \ldots + \varphi(u_n))$$

However, $\mathbf{C}$ is a copula function if and only if the function $\varphi^{-1}(u)$ is completely monotone (Nelsen, 2006):

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}u^k} \varphi^{-1}(u) \geq 0 \qquad \forall\, k \geq 1$$

For instance, the multivariate Gumbel copula is defined by:

$$\mathbf{C}(u_1, \ldots, u_n) = \exp\left(-\left((-\ln u_1)^{\theta} + \ldots + (-\ln u_n)^{\theta} \right)^{1/\theta} \right)$$

The previous construction is related to an important class of multivariate distributions, which are called frailty models (Oakes, 1989). Let $\mathbf{F}_1, \ldots, \mathbf{F}_n$ be univariate distribution functions, and let $\mathbf{G}$ be an n-variate distribution function with univariate marginals $\mathbf{G}_i$, such that $\bar{\mathbf{G}}(0, \ldots, 0) = 1$. We denote by ψ_i the Laplace transform of $\mathbf{G}_i$. Marshall and Olkin (1988) showed that the function defined by:

$$\mathbf{F}(x_1, \ldots, x_n) = \int \cdots \int \mathbf{C}\left(\mathbf{H}_1^{t_1}(x_1), \ldots, \mathbf{H}_n^{t_n}(x_n)\right) \mathrm{d}\mathbf{G}(t_1, \ldots, t_n)$$

is a multivariate probability distribution with marginals $\mathbf{F}_1, \ldots, \mathbf{F}_n$ if $\mathbf{H}_i(x) = \exp\left(-\psi_i^{-1}(\mathbf{F}_i(x))\right)$. If we assume that the univariate distributions $\mathbf{G}_i$ are the same and equal to $\mathbf{G}_1$, $\mathbf{G}$ is the upper Fréchet bound and $\mathbf{C}$ is the product copula $\mathbf{C}^{\perp}$, the previous expression becomes:

$$\begin{aligned}
\mathbf{F}(x_1, \ldots, x_n) &= \int \prod_{i=1}^{n} \mathbf{H}_i^{t_1}(x_i)\, \mathrm{d}\mathbf{G}_1(t_1) \\
&= \int \exp\left(-t_1 \sum_{i=1}^{n} \psi^{-1}(\mathbf{F}_i(x_i)) \right) \mathrm{d}\mathbf{G}_1(t_1) \\
&= \psi\left(\psi^{-1}(\mathbf{F}_1(x_1)) + \ldots + \psi^{-1}(\mathbf{F}_n(x_n))\right)
\end{aligned}$$

The corresponding copula is then given by:

$$\mathbf{C}(u_1, \ldots, u_n) = \psi\left(\psi^{-1}(u_1) + \ldots + \psi^{-1}(u_n)\right)$$

This is a special case of Archimedean copulas where the generator φ is the inverse of a Laplace transform. For instance, the Clayton copula is a frailty copula where $\psi(x) = (1+x)^{-1/\theta}$ is the Laplace transform of a Gamma random variable. The Gumbel-Hougaard copula is frailty too and we have $\psi(x) = \exp\left(-x^{1/\theta}\right)$. This is the Laplace transform of a positive stable distribution.

For frailty copulas, Joe (1997) showed that upper and lower tail dependence measures are given by:

$$\lambda^+ = 2 - 2\lim_{x\to 0}\frac{\psi'(2x)}{\psi'(x)}$$

and:

$$\lambda^- = 2\lim_{x\to\infty}\frac{\psi'(2x)}{\psi'(x)}$$

Example 121 *In the case of the Clayton copula, the Laplace transform is* $\psi(x) = (1+x)^{-1/\theta}$*. We have:*

$$\frac{\psi'(2x)}{\psi'(x)} = \frac{(1+2x)^{-1/\theta-1}}{(1+x)^{-1/\theta-1}}$$

We deduce that:

$$\begin{aligned}\lambda^+ &= 2 - 2\lim_{x\to 0}\frac{(1+2x)^{-1/\theta-1}}{(1+x)^{-1/\theta-1}}\\ &= 2-2\\ &= 0\end{aligned}$$

and:

$$\begin{aligned}\lambda^- &= 2\lim_{x\to\infty}\frac{(1+2x)^{-1/\theta-1}}{(1+x)^{-1/\theta-1}}\\ &= 2\times 2^{-1/\theta-1}\\ &= 2^{-1/\theta}\end{aligned}$$

11.3.2 Normal copula

The Normal copula is the dependence function of the multivariate normal distribution with a correlation matrix ρ:

$$\mathbf{C}(u_1,\ldots,u_n;\rho) = \Phi_n\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n);\rho\right)$$

By using the canonical decomposition of the multivariate density function:

$$f(x_1,\ldots,x_n) = c\left(\mathbf{F}_1(x_1),\ldots,\mathbf{F}_n(x_n)\right)\prod_{i=1}^{n} f_i(x_i)$$

we deduce that the probability density function of the Normal copula is:

$$c(u_1,\ldots,u_n,;\rho) = \frac{1}{|\rho|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}x^\top\left(\rho^{-1}-I_n\right)x\right)$$

where $x_i = \Phi^{-1}(u_i)$. In the bivariate case, we obtain[7]:

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)} + \frac{x_1^2 + x_2^2}{2}\right)$$

It follows that the expression of the bivariate Normal copula function is also equal to:

$$\mathbf{C}(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \phi_2(x_1, x_2; \rho)\, \mathrm{d}x_1\, \mathrm{d}x_2 \tag{11.6}$$

where $\phi_2(x_1, x_2; \rho)$ is the bivariate normal density:

$$\phi_2(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right)$$

Example 122 *Let (X_1, X_2) be a standardized Gaussian random vector, whose cross-correlation is ρ. Using the Cholesky decomposition, we write X_2 as follows:*

$$X_2 = \rho X_1 + \sqrt{1-\rho^2} X_3$$

where $X_3 \sim \mathcal{N}(0,1)$ is independent from X_1 and X_2. We have:

$$\begin{aligned}
\Phi_2(x_1, x_2; \rho) &= \Pr\{X_1 \leq x_1, X_2 \leq x_2\} \\
&= \mathbb{E}\left[\Pr\left\{X_1 \leq x_1, \rho X_1 + \sqrt{1-\rho^2} X_3 \leq x_2 \mid X_1\right\}\right] \\
&= \int_{-\infty}^{x_1} \Phi\left(\frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right) \phi(x)\, \mathrm{d}x
\end{aligned}$$

It follows that:

$$\mathbf{C}(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho x}{\sqrt{1-\rho^2}}\right) \phi(x)\, \mathrm{d}x$$

We finally obtain that the bivariate Normal copula function is equal to:

$$\mathbf{C}(u_1, u_2; \rho) = \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho \Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \mathrm{d}u \tag{11.7}$$

This expression is more convenient to use than Equation (11.6).

Like the normal distribution, the Normal copula is easy to manipulate for computational purposes. For instance, Kendall's tau and Spearman's rho are equal to:

$$\tau = \frac{2}{\pi} \arcsin \rho$$

and:

$$\varrho = \frac{6}{\pi} \arcsin \frac{\rho}{2}$$

[7] In the bivariate case, the parameter ρ is the cross-correlation between X_1 and X_2, that is the element $(1,2)$ of the correlation matrix.

The conditional distribution $\mathbf{C}_{2|1}(u_1, u_2)$ has the following expression:

$$\begin{aligned}\mathbf{C}_{2|1}(u_1, u_2) &= \partial_1 \mathbf{C}(u_1, u_2) \\ &= \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right)\end{aligned}$$

To compute the tail dependence, we apply Equation (11.4) and obtain:

$$\begin{aligned}\lambda^+ &= 2 - 2\lim_{u\to 1^-} \Phi\left(\frac{\Phi^{-1}(u) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \\ &= 2 - 2\lim_{u\to 1^-} \Phi\left(\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\Phi^{-1}(u)\right)\end{aligned}$$

We finally deduce that:

$$\lambda^+ = \lambda^- = \begin{cases} 0 & \text{if } \rho < 1 \\ 1 & \text{if } \rho = 1 \end{cases}$$

In Figure 11.11, we have represented the quantile-quantile dependence measure $\lambda^+(\alpha)$ for several values of the parameter ρ. When ρ is equal to 90% and α is close to one, we notice that $\lambda^+(\alpha)$ dramatically decreases. This means that even if the correlation is high, the extremes are independent.

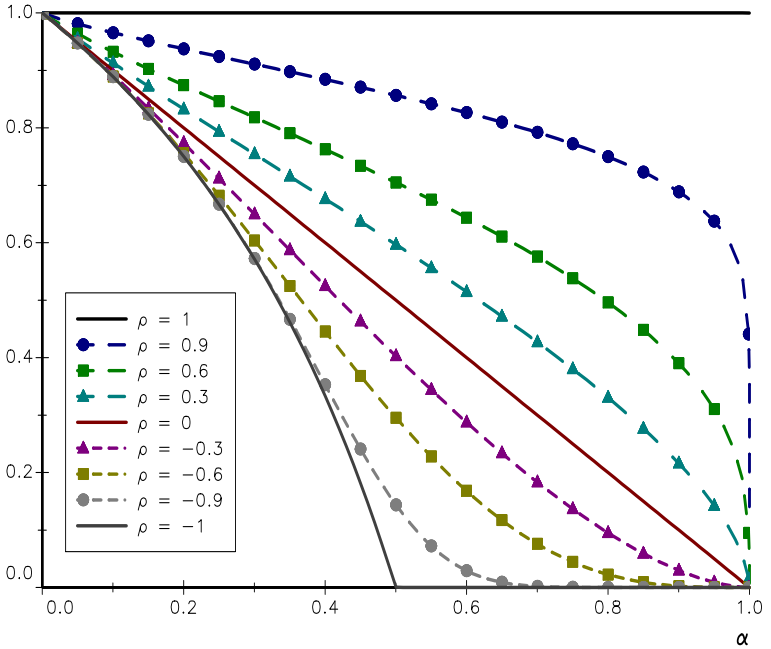


FIGURE 11.11: Tail dependence $\lambda^+(\alpha)$ for the Normal copula

11.3.3 Student's t copula

In a similar way, the Student's t copula is the dependence function associated with the multivariate Student's t probability distribution:

$$\mathbf{C}(u_1, \ldots, u_n; \rho, \nu) = \mathbf{T}_n\left(\mathbf{T}_\nu^{-1}(u_1), \ldots, \mathbf{T}_\nu^{-1}(u_n); \rho, \nu\right)$$

By using the definition of the cumulative distribution function:

$$\mathbf{T}_n(x_1,\ldots,x_n;\rho,\nu) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}\frac{\Gamma\left(\frac{\nu+n}{2}\right)|\rho|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{n}{2}}}\left(1+\frac{1}{\nu}x^\top\rho^{-1}x\right)^{-\frac{\nu+n}{2}}\mathrm{d}x$$

we can show that the copula density is then:

$$c(u_1,\ldots,u_n,;\rho,\nu) = |\rho|^{-\frac{1}{2}}\frac{\Gamma\left(\frac{\nu+n}{2}\right)\left[\Gamma\left(\frac{\nu}{2}\right)\right]^n}{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^n\Gamma\left(\frac{\nu}{2}\right)}\frac{\left(1+\frac{1}{\nu}x^\top\rho^{-1}x\right)^{-\frac{\nu+n}{2}}}{\prod_{i=1}^n\left(1+\frac{x_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}}$$

where $x_i = \mathbf{T}_\nu^{-1}(u_i)$. In the bivariate case, we deduce that the t copula has the following expression:

$$\begin{aligned}\mathbf{C}(u_1,u_2;\rho,\nu) &= \int_{-\infty}^{\mathbf{T}_\nu^{-1}(u_1)}\int_{-\infty}^{\mathbf{T}_\nu^{-1}(u_2)}\frac{1}{2\pi\sqrt{1-\rho^2}}\cdot\\ &\quad\left(1+\frac{x_1^2+x_2^2-2\rho x_1x_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

Like the Normal copula, we can obtain another expression, which is easier to manipulate. Let (X_1,X_2) be a random vector whose probability distribution is $\mathbf{T}_2(x_1,x_2;\rho,\nu)$. Conditionally to $X_1 = x_1$, we have:

$$\left(\frac{\nu+1}{\nu+x_1^2}\right)^{1/2}\frac{X_2-\rho x_1}{\sqrt{1-\rho^2}} \sim \mathbf{T}_{\nu+1}$$

The conditional distribution $\mathbf{C}_{2|1}(u_1,u_2)$ is then equal to:

$$\mathbf{C}_{2|1}(u_1,u_2;\rho,\nu) = \mathbf{T}_{\nu+1}\left(\left(\frac{\nu+1}{\nu+\left[\mathbf{T}_\nu^{-1}(u_1)\right]^2}\right)^{1/2}\frac{\mathbf{T}_\nu^{-1}(u_2)-\rho\mathbf{T}_\nu^{-1}(u_1)}{\sqrt{1-\rho^2}}\right)$$

We deduce that:

$$\mathbf{C}(u_1,u_2;\rho,\nu) = \int_0^{u_1}\mathbf{C}_{2|1}(u,u_2;\rho,\nu)\,\mathrm{d}u$$

We can show that the expression of Kendall's tau for the t copula is the one obtained for the Normal copula. In the case of Spearman's rho, there is no analytical expression. We denote by $\varrho_t(\rho,\nu)$ and $\varrho_n(\rho)$ the values of Spearman's rho for Student's t and Normal copulas with same parameter ρ. We can show that $\varrho_t(\rho,\nu) > \varrho_n(\rho)$ for negative values of ρ and $\varrho_t(\rho,\nu) < \varrho_n(\rho)$ for positive values of ρ. In Figure 11.12, we report the relationship between τ and ϱ for different degrees of freedom ν.

Because the t copula is symmetric, we can apply Equation (11.4) and obtain:

$$\begin{aligned}\lambda^+ &= 2-2\lim_{u\to1^-}\mathbf{T}_{\nu+1}\left(\left(\frac{\nu+1}{\nu+\left[\mathbf{T}_\nu^{-1}(u)\right]^2}\right)^{1/2}\frac{\mathbf{T}_\nu^{-1}(u)-\rho\mathbf{T}_\nu^{-1}(u)}{\sqrt{1-\rho^2}}\right)\\ &= 2-2\cdot\mathbf{T}_{\nu+1}\left(\left(\frac{(\nu+1)(1-\rho)}{(1+\rho)}\right)^{1/2}\right)\end{aligned}$$

We finally deduce that:

$$\lambda^+ = \begin{cases}0 & \text{if } \rho=-1\\ >0 & \text{if } \rho>-1\end{cases}$$

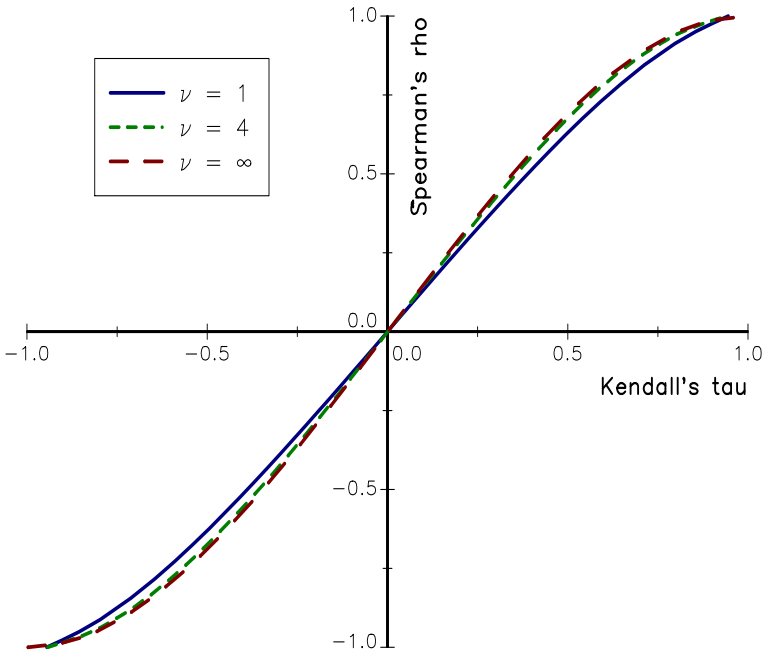


FIGURE 11.12: Relationship between τ and ϱ of the Student's t copula

Contrary to the Normal copula, the t copula has an upper tail dependence. In Figures 11.13 and 11.14, we represent the quantile-quantile dependence measure $\lambda^+(\alpha)$ for two degrees of freedom ν. We observe that the behavior of $\lambda^+(\alpha)$ is different than the one obtained in Figure 11.11 with the Normal copula. In Table 11.2, we give the numerical values of the coefficient λ^+ for various values of ρ and ν. We notice that it is strictly positive for small degrees of freedom even if the parameter ρ is negative. For instance, λ^+ is equal to 13.40% when ν and ρ are equal to 1 and -50%. We also observe that the convergence to the Gaussian case is low when the parameter ρ is positive.

TABLE 11.2: Values in % of the upper tail dependence λ^+ for the Student's t copula

ν	Parameter ρ (in %)					
	-70.00	-50.00	0.00	50.00	70.00	90.00
1	7.80	13.40	29.29	50.00	61.27	77.64
2	2.59	5.77	18.17	39.10	51.95	71.77
3	0.89	2.57	11.61	31.25	44.81	67.02
4	0.31	1.17	7.56	25.32	39.07	62.98
6	0.04	0.25	3.31	17.05	30.31	56.30
10	0.00	0.01	0.69	8.19	19.11	46.27
∞	0.00	0.00	0.00	0.00	0.00	0.00

Remark 139 *The Normal copula is a particular case of the Student's t copula when ν tends to ∞. This is why these two copulas are often compared for a given value of ρ. However, we must be careful because the previous analysis of the tail dependence has shown that these two copulas are very different. Let us consider the bivariate case. We can write the Student's t*

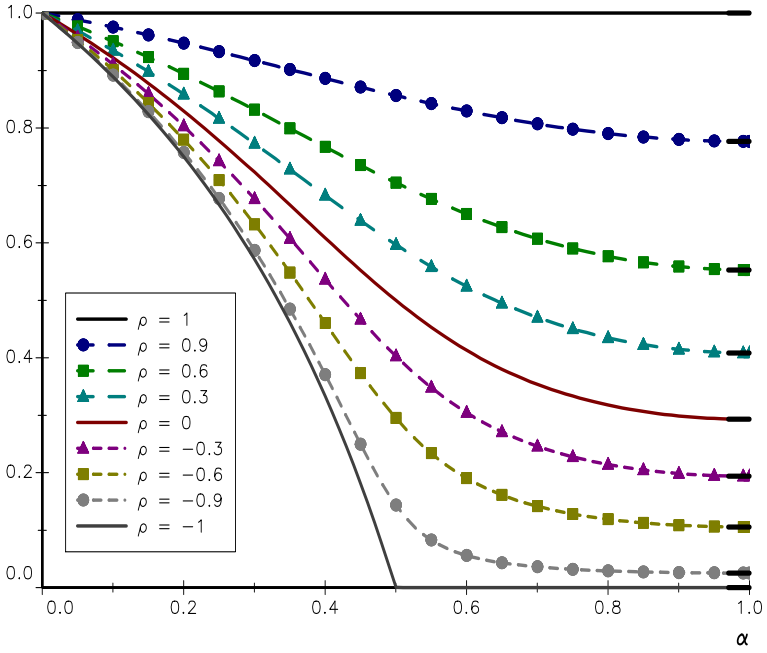


FIGURE 11.13: Tail dependence $\lambda^+(\alpha)$ for the Student's t copula ($\nu = 1$)

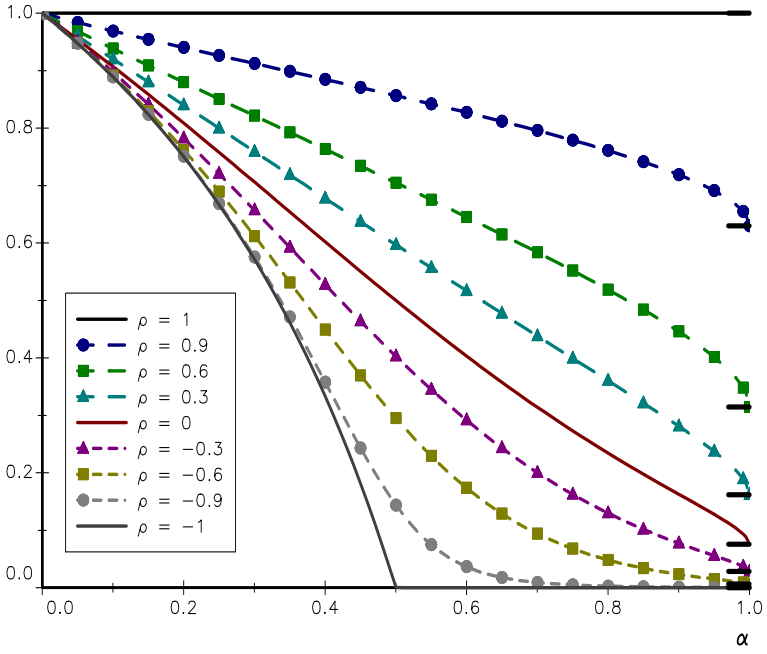


FIGURE 11.14: Tail dependence $\lambda^+(\alpha)$ for the Student's t copula ($\nu = 4$)

random vector (T_1, T_2) *as follows:*

$$\begin{aligned}(T_1, T_2) &= \frac{(N_1, N_2)}{\sqrt{X/\nu}} \\ &= \left(\frac{N_1}{\sqrt{X/\nu}}, \rho\frac{N_1}{\sqrt{X/\nu}} + \sqrt{1-\rho^2}\frac{N_3}{\sqrt{X/\nu}}\right)\end{aligned}$$

where N_1 *and* N_3 *are two independent Gaussian random variables and* X *is a random variable, whose probability distribution is* $\chi^2(\nu)$*. This is the introduction of the random variable* X *that produces a strong dependence between* T_1 *and* T_2*, and correlates the extremes. Even if the parameter* ρ *is equal to zero, we obtain:*

$$(T_1, T_2) = \left(\frac{N_1}{\sqrt{X/\nu}}, \frac{N_3}{\sqrt{X/\nu}}\right)$$

This implies that the product copula $\mathbf{C}^{\perp}$ *can never be attained by the t copula.*

11.4 Statistical inference and estimation of copula functions

We now consider the estimation problem of copula functions. We first introduce the empirical copula, which may viewed as a non-parametric estimator of the copula function. Then, we discuss the method of moments to estimate the parameters of copula functions. Finally, we apply the method of maximum likelihood and show the different forms of implementation.

11.4.1 The empirical copula

Let $\hat{\mathbf{F}}$ be the empirical distribution associated to a sample of T observations of the random vector $(X_1, \ldots, X_n)$. Following Deheuvels (1979), any copula $\hat{\mathbf{C}} \in \mathcal{C}$ defined on the lattice $\mathfrak{L}$:

$$\mathfrak{L} = \left\{\left(\frac{t_1}{T}, \ldots, \frac{t_n}{T}\right) : 1 \leq j \leq n, t_j = 0, \ldots, T\right\}$$

by the function:

$$\hat{\mathbf{C}}\left(\frac{t_1}{T}, \ldots, \frac{t_n}{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{n}\mathbb{1}\left\{\mathfrak{R}_{t,i} \leq t_i\right\}$$

is an empirical copula. Here $\mathfrak{R}_{t,i}$ is the rank statistic of the random variable X_i meaning that $X_{\mathfrak{R}_{t,i}:T,i} = X_{t,i}$. We notice that $\hat{\mathbf{C}}$ is the copula function associated to the empirical distribution $\hat{\mathbf{F}}$. However, $\hat{\mathbf{C}}$ is not unique because $\hat{\mathbf{F}}$ is not continuous. In the bivariate case, we obtain:

$$\begin{aligned}\hat{\mathbf{C}}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left\{\mathfrak{R}_{t,1} \leq t_1, \mathfrak{R}_{t,2} \leq t_2\right\} \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left\{x_{t,1} \leq x_{t_1:T,1}, x_{t,2} \leq x_{t_2:T,2}\right\}\end{aligned}$$

where $\{(x_{t,1}, x_{t,2}), t = 1, \ldots, T\}$ denotes the sample of (X_1, X_2). Nelsen (2006) defines the empirical copula frequency function as follows:

$$\begin{aligned}\hat{c}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) &= \hat{\mathbf{C}}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) - \hat{\mathbf{C}}\left(\frac{t_1-1}{T}, \frac{t_2}{T}\right) - \\ &\quad \hat{\mathbf{C}}\left(\frac{t_1}{T}, \frac{t_2-1}{T}\right) + \hat{\mathbf{C}}\left(\frac{t_1-1}{T}, \frac{t_2-1}{T}\right) \\ &= \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left\{x_{t,1} = x_{t_1:T,1}, x_{t,2} = x_{t_2:T,2}\right\}\end{aligned}$$

We have then:

$$\hat{\mathbf{C}}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \sum_{j_1=1}^{t_1}\sum_{j_2=1}^{t_2} \hat{c}\left(\frac{j_1}{T}, \frac{j_2}{T}\right)$$

We can interpret $\hat{c}$ as the probability density function of the sample.

Example 123 *We consider the daily returns of European (EU) and American (US) MSCI equity indices from January 2006 to December 2015. In Figure 11.15, we represent the level lines of the empirical copula and compare them with the level lines of the Normal copula. For this copula function, the parameter ρ is estimated by the linear correlation between the daily returns of the two MSCI equity indices. We notice that the Normal copula does not exactly fit the empirical copula.*

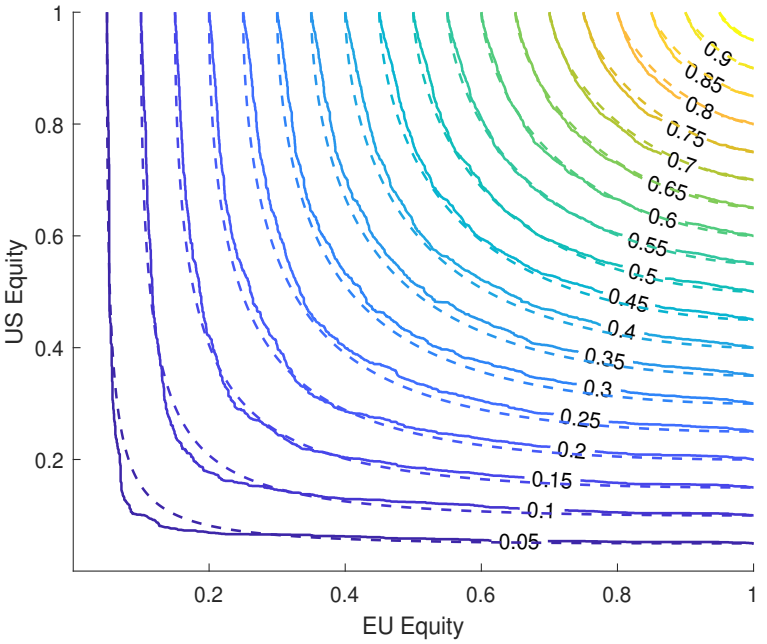


FIGURE 11.15: Comparison of the empirical copula (solid line) and the Normal copula (dashed line)

Like the histogram of the empirical distribution function $\hat{\mathbf{F}}$, it is difficult to extract information from $\hat{\mathbf{C}}$ or $\hat{c}$, because these functions are not smooth[8]. It is better to use a dependogram. This representation has been introduced by Deheuvels (1981), and consists in transforming the sample $\{(x_{t,1}, x_{t,2}), t = 1, \ldots, T\}$ of the random vector (X_1, X_2) into a sample $\{(u_{t,1}, u_{t,2}), t = 1, \ldots, T\}$ of uniform random variables (U_1, U_2) by considering the rank statistics:

$$u_{t,i} = \frac{1}{T}\mathfrak{R}_{t,i}$$

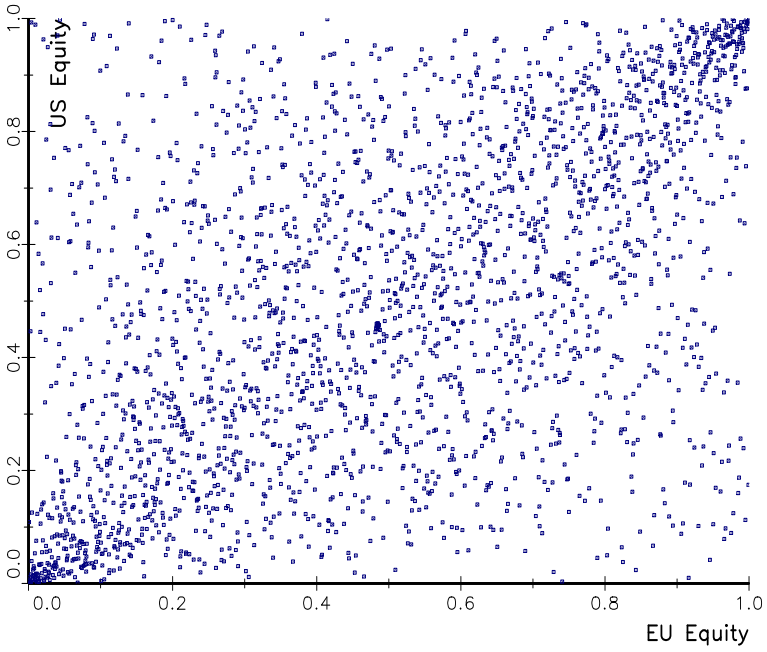


FIGURE 11.16: Dependogram of EU and US equity returns

The dependogram is then the scatter plot between $u_{t,1}$ and $u_{t,2}$. For instance, Figure 11.16 shows the dependogram of EU and US equity returns. We can compare this figure with the one obtained by assuming that equity returns are Gaussian. Indeed, Figure 11.17 shows the dependogram of a simulated bivariate Gaussian random vector when the correlation is equal to 57.8%, which is the estimated value between EU and US equity returns during the study period.

11.4.2 The method of moments

When it is applied to copulas, this method is different than the one presented in Chapter 10. Indeed, it consists in estimating the parameters θ of the copula function from the population version of concordance measures. For instance, if $\tau = f_\tau(\theta)$ is the relationship between θ and Kendall's tau, the MM estimator is simply the inverse of this relationship:

$$\hat{\theta} = f_\tau^{-1}(\hat{\tau})$$

[8] This is why they are generally coupled with approximation methods based on Bernstein polynomials (Sancetta and Satchell, 2004).

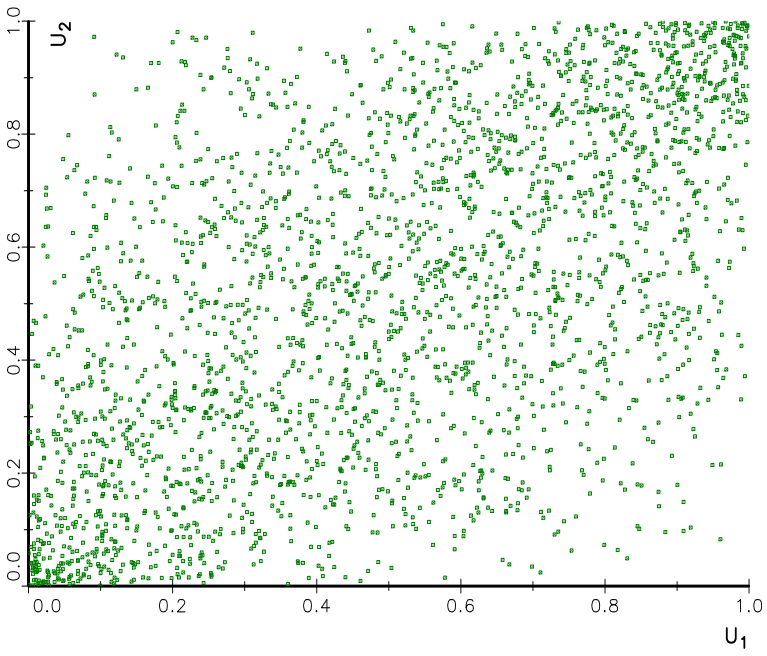


FIGURE 11.17: Dependogram of simulated Gaussian returns

where $\hat{\tau}$ is the estimate of Kendall's tau based on the sample[9]. For instance, in the case of the Gumbel copula, we obtain:

$$\hat{\theta} = \frac{1}{1-\hat{\tau}}$$

Remark 140 *This approach is also valid for other concordance measures like Spearman's rho. We have then:*

$$\hat{\theta} = f_{\varrho}^{-1}(\hat{\varrho})$$

where $\hat{\varrho}$ is the estimate[10] of Spearman's rho and f_{ϱ} is the theoretical relationship between θ and Spearman's rho.

Example 124 *We consider the daily returns of 5 asset classes from January 2006 to December 2015. These asset classes are represented by the European MSCI equity index, the American MSCI equity index, the Barclays sovereign bond index, the Barclays corporate investment grade bond index and the Bloomberg commodity index. In Table 11.3, we report the correlation matrix. In Tables 11.4 and 11.5, we assume that the dependence function is a Normal copula and give the matrix $\hat{\rho}$ of estimated parameters using the method of moments based on Kendall's tau and Spearman's rhorho. We notice that these two matrices are very close, but we also observe some important differences with the correlation matrix reported in Table 11.3.*

[9] We have:

$$\hat{\tau} = \frac{c-d}{c+d}$$

where c and d are respectively the number of concordant and discordant pairs.

[10] It is equal to the linear correlation between the rank statistics.

TABLE 11.3: Matrix of linear correlations $\hat{\rho}_{i,j}$

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	57.8	100.0			
Sovereign	−34.0	−32.6	100.0		
Credit	−15.1	−28.6	69.3	100.0	
Commodity	51.8	34.3	−22.3	−14.4	100.0

TABLE 11.4: Matrix of parameters $\hat{\rho}_{i,j}$ estimated using Kendall's tau

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	57.7	100.0			
Sovereign	−31.8	−32.1	100.0		
Credit	−17.6	−33.8	73.9	100.0	
Commodity	43.4	30.3	−19.6	−15.2	100.0

TABLE 11.5: Matrix of parameters $\hat{\rho}_{i,j}$ estimated using Spearman's rho

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	55.4	100.0			
Sovereign	−31.0	−31.3	100.0		
Credit	−17.1	−32.7	73.0	100.0	
Commodity	42.4	29.4	−19.2	−14.9	100.0

11.4.3 The method of maximum likelihood

Let us denote by $\{(x_{t,1}, \ldots, x_{t,n}), t = 1 \ldots, T\}$ the sample of the random vector $(X_1, \ldots, X_n)$, whose multivariate distribution function has the following canonical decomposition:

$$\mathbf{F}(x_1, \ldots, x_n) = \mathbf{C}(\mathbf{F}_1(x_1; \theta_1), \ldots, \mathbf{F}_n(x_n; \theta_n); \theta_c)$$

This means that this statistical model depends on two types of parameters:

- the parameters $(\theta_1, \ldots, \theta_n)$ of univariate distribution functions;
- the parameters θ_c of the copula function.

The expression of the log-likelihood function is:

$$\begin{aligned} \boldsymbol{\ell}(\theta_1, \ldots, \theta_n, \theta_c) &= \sum_{t=1}^{T} \ln c(\mathbf{F}_1(x_{t,1}; \theta_1), \ldots, \mathbf{F}_n(x_{t,n}; \theta_n); \theta_c) + \\ &\quad \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(x_{t,i}; \theta_i) \end{aligned}$$

where c is the copula density and f_i is the probability density function associated to $\mathbf{F}_i$. The ML estimator is then defined as follows:

$$\left(\hat{\theta}_1, \ldots, \hat{\theta}_n, \hat{\theta}_c\right) = \arg\max \boldsymbol{\ell}(\theta_1, \ldots, \theta_n, \theta_c)$$

The estimation by maximum likelihood method can be time-consuming when the number of parameters is large. However, the copula approach suggests a two-stage parametric method (Shih and Louis, 1995):

1. the first stage involves maximum likelihood from univariate marginals, meaning that we estimate the parameters $\theta_1, \ldots, \theta_n$ separately for each marginal:

$$\hat{\theta}_i = \arg\max \sum_{t=1}^{T} \ln f_i\left(x_{t,i}; \theta_i\right)$$

2. the second stage involves maximum likelihood of the copula parameters θ_c with the univariate parameters $\hat{\theta}_1, \ldots, \hat{\theta}_n$ held fixed from the first stage:

$$\hat{\theta}_c = \arg\max \sum_{t=1}^{T} \ln c\left(\mathbf{F}_1\left(x_{t,1}; \hat{\theta}_1\right), \ldots, \mathbf{F}_n\left(x_{t,n}; \hat{\theta}_n\right); \theta_c\right)$$

This approach is known as the method of inference functions for marginals or IFM (Joe, 1997). Let $\hat{\theta}_{\text{IFM}}$ be the IFM estimator obtained with this two-stage procedure. We have:

$$T^{1/2}\left(\hat{\theta}_{\text{IFM}} - \theta_0\right) \to \mathcal{N}\left(\mathbf{0}, \mathcal{V}^{-1}\left(\theta_0\right)\right)$$

where $\mathcal{V}\left(\theta_0\right)$ is the Godambe matrix (Joe, 1997).

Genest *et al.* (1995) propose a third estimation method, which consists in estimating the copula parameters θ_c by considering the non-parametric estimates of the marginals $\mathbf{F}_1, \ldots, \mathbf{F}_n$:

$$\hat{\theta}_c = \arg\max \sum_{t=1}^{T} \ln c\left(\hat{\mathbf{F}}_1\left(x_{t,1}\right), \ldots, \hat{\mathbf{F}}_n\left(x_{t,n}\right); \theta_c\right)$$

In this case, $\hat{\mathbf{F}}_i\left(x_{t,i}\right)$ is the normalized rank $\mathfrak{R}_{t,i}/T$. This estimator called omnibus or OM is then the ML estimate applied to the dependogram.

Example 125 *Let us assume that the dependence function of asset returns (X_1, X_2) is the Frank copula whereas the marginals are Gaussian. The log-likelihood function for observation t is then equal to:*

$$\begin{aligned}
\boldsymbol{\ell}_t &= \ln\left(\theta_c\left(1 - e^{-\theta_c}\right) e^{-\theta_c(\Phi(y_{t,1}) + \Phi(y_{t,2}))}\right) - \\
&\quad \ln\left(\left(1 - e^{-\theta_c}\right) - \left(1 - e^{-\theta_c \Phi(y_{t,1})}\right)\left(1 - e^{-\theta_c \Phi(y_{t,2})}\right)\right)^2 - \\
&\quad \left(\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\sigma_1^2 + \frac{1}{2}y_{t,1}^2\right) - \\
&\quad \left(\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\sigma_2^2 + \frac{1}{2}y_{t,2}^2\right)
\end{aligned}$$

where $y_{t,i} = \sigma_i^{-1}\left(x_{t,i} - \mu_i\right)$ is the standardized return of asset i for the observation t. The vector of parameters to estimate is $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \theta_c)$. In the case of the IFM approach, the parameters $(\mu_1, \sigma_1, \mu_2, \sigma_2)$ are estimated in a first step. Then, we estimate the copula parameter θ_c by considering the following log-likelihood function:

$$\begin{aligned}
\boldsymbol{\ell}_t &= \ln\left(\theta_c\left(1 - e^{-\theta_c}\right) e^{-\theta_c(\Phi(\hat{y}_{t,1}) + \Phi(\hat{y}_{t,2}))}\right) - \\
&\quad \ln\left(\left(1 - e^{-\theta_c}\right) - \left(1 - e^{-\theta_c \Phi(\hat{y}_{t,1})}\right)\left(1 - e^{-\theta_c \Phi(\hat{y}_{t,2})}\right)\right)^2
\end{aligned}$$

where $\hat{y}_{t,i}$ *is equal to* $\hat{\sigma}_i^{-1}(x_{t,i} - \hat{\mu}_i)$. *Finally, the OM approach uses the uniform variates* $u_{t,i} = \mathfrak{R}_{t,i}/T$ *in the expression of the log-likelihood function function:*

$$\begin{aligned}\boldsymbol{\ell}_t &= \ln\left(\theta_c\left(1-e^{-\theta_c}\right)e^{-\theta_c(u_{t,1}+u_{t,2})}\right) - \\ &\quad \ln\left(\left(1-e^{-\theta_c}\right)-\left(1-e^{-\theta_c u_{t,1}}\right)\left(1-e^{-\theta_c u_{t,2}}\right)\right)^2\end{aligned}$$

Using the returns of MSCI Europe and US indices for the last 10 years, we obtain the following results for the parameter θ_c of the Frank copula:

	ML	IFM	OM	Method of Moments	
				Kendall	Spearman
$\hat{\theta}_c$	6.809	6.184	4.149	3.982	3.721
$\hat{\tau}$	0.554	0.524	0.399	0.387	0.367
$\hat{\varrho}$	0.754	0.721	0.571	0.555	0.529

We obtain $\hat{\theta}_c = 6.809$ for the method of maximum likelihood and $\hat{\theta}_c = 6.184$ for the IFM approach. These results are very close, that is not the case with the omnibus approach where we obtain $\hat{\theta}_c = 4.149$. This means that the assumption of Gaussian marginals is far to be verified. The specification of wrong marginals in ML and IFM approaches induces then a bias in the estimation of the copula parameter. With the omnibus approach, we do not face this issue because we consider non-parametric marginals. This explains that we obtain a value, which is close to the MM estimates (Kendall's tau and Spearman's rho).

For IFM and OM approaches, we can obtain a semi-analytical expression of $\hat{\theta}_c$ for some specific copula functions. In the case of the Normal copula, the matrix ρ of the parameters is estimated with the following algorithm:

1. we first transform the uniform variates $u_{t,i}$ into Gaussian variates:
$$n_{t,i} = \Phi^{-1}(u_{t,i})$$
2. we then calculate the correlation matrix of the Gaussian variates $n_{t,i}$.

For the Student's t copula, Bouyé *et al.* (2000) suggest the following algorithm:

1. let $\hat{\rho}_0$ be the estimated value of ρ for the Normal copula;
2. $\hat{\rho}_{k+1}$ is obtained using the following equation:
$$\hat{\rho}_{k+1} = \frac{1}{T}\sum_{t=1}^{T}\frac{(\nu+n)\,\varsigma_t\varsigma_t^\top}{\nu+\varsigma_t^\top\hat{\rho}_k^{-1}\varsigma_t}$$
where:
$$\varsigma_t = \begin{pmatrix}\mathbf{t}_\nu^{-1}(u_{t,1})\\ \vdots \\ \mathbf{t}_\nu^{-1}(u_{t,n})\end{pmatrix}$$
3. repeat the second step until convergence: $\hat{\rho}_{k+1} = \hat{\rho}_k := \hat{\rho}_\infty$.

Let us consider Example 124. We have estimated the parameter matrix ρ of Normal and Student's t copulas using the omnibus approach. Results are given in Tables 11.6, 11.7 and 11.8. We notice that these matrices are different than the correlation matrix calculated in Table 11.3. The reason is that we have previously assumed that the marginals were Gaussian. In this case, the ML estimate introduced a bias in the copula parameter in order to compensate the bias induced by the wrong specification of the marginals.

TABLE 11.6: Omnibus estimate $\hat{\rho}$ (Normal copula)

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	56.4	100.0			
Sovereign	−32.5	−32.1	100.0		
Credit	−16.3	−30.3	70.2	100.0	
Commodity	46.5	30.7	−21.1	−14.7	100.0

TABLE 11.7: Omnibus estimate $\hat{\rho}$ (Student's t copula with $\nu = 1$)

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	47.1	100.0			
Sovereign	−20.3	−18.9	100.0		
Credit	−9.3	−22.1	57.6	100.0	
Commodity	28.0	17.1	−7.4	−6.2	100.0

TABLE 11.8: Omnibus estimate $\hat{\rho}$ (Student's t copula with $\nu = 4$)

	EU Equity	US Equity	Sovereign	Credit	Commodity
EU Equity	100.0				
US Equity	59.6	100.0			
Sovereign	−31.5	−31.9	100.0		
Credit	−18.3	−32.9	71.3	100.0	
Commodity	43.0	30.5	−17.2	−13.4	100.0

Remark 141 *The discrepancy between the ML or IFM estimate and the OM estimate is interesting information for knowing if the specification of the marginals are right or not. In particular, a large discrepancy indicates that the estimated marginals are far from the empirical marginals.*

11.5 Exercises

11.5.1 Gumbel logistic copula

1. Calculate the density of the Gumbel logistic copula.
2. Show that it has a lower tail dependence, but no upper tail dependence.

11.5.2 Farlie-Gumbel-Morgenstern copula

We consider the following function:

$$\mathbf{C}(u_1, u_2) = u_1 u_2 (1 + \theta (1 - u_1)(1 - u_2)) \tag{11.8}$$

1. Show that $\mathbf{C}$ is a copula function for $\theta \in [-1, 1]$.
2. Calculate the tail dependence coefficient λ, the Kendall's τ statistic and the Spearman's ϱ statistic.
3. Let $X = (X_1, X_2)$ be a bivariate random vector. We assume that $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ and $X_2 \sim \mathcal{E}(\lambda)$. Propose an algorithm to simulate (X_1, X_2) when the copula is the function (11.8).
4. Calculate the log-likelihood function of the sample $\left\{(x_{1,i}, x_{2,i})_{i=1}^{i=n}\right\}$.

11.5.3 Survival copula

Let $\mathbf{S}$ be the bivariate function defined by:

$$\mathbf{S}(x_1, x_2) = \exp\left(-\left(x_1 + x_2 - \theta\frac{x_1 x_2}{x_1 + x_2}\right)\right)$$

with $\theta \in [0, 1]$, $x_1 \geq 0$ et $x_2 \geq 0$.

1. Verify that $\mathbf{S}$ is a survival function.
2. Define the survival copula associated to $\mathbf{S}$.

11.5.4 Method of moments

Let (X_1, X_2) be a bivariate random vector such that $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. We consider that the dependence function is given by the following copula:

$$\mathbf{C}(u_1, u_2) = \theta \cdot \mathbf{C}^-(u_1, u_2) + (1 - \theta) \cdot \mathbf{C}^+(u_1, u_2)$$

where $\theta \in [0, 1]$ is the copula parameter.

1. We assume that $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$. Find the parameter θ such that the linear correlation of X_1 and X_2 is equal to zero. Show that there exists a function f such that $X_1 = f(X_2)$. Comment on this result.
2. Calculate the linear correlation of X_1 and X_2 as a function of the parameters μ_1, μ_2, σ_1, σ_2 and θ.
3. Propose a method of moments to estimate θ.

11.5.5 Correlated loss given default rates

We assume that the probability distribution of the (annual) loss given default rate associated to a risk class $\mathcal{C}$ is given by:

$$\begin{aligned}\mathbf{F}(x) &= \Pr\{\text{LGD} \leq x\} \\ &= x^\gamma\end{aligned}$$

1. Find the conditions on the parameter γ that are necessary for $\mathbf{F}$ to be a probability distribution.
2. Let $\{x_1, \ldots, x_n\}$ be a sample of loss given default rates. Calculate the log-likelihood function and deduce the ML estimator $\hat{\gamma}_{\text{ML}}$.

3. Calculate the first moment $\mathbb{E}[\text{LGD}]$. Then find the method of moments estimator $\hat{\gamma}_{\text{MM}}$.

4. We assume that $x_i = 50\%$ for all i. Calculate the numerical values taken by $\hat{\gamma}_{\text{ML}}$ and $\hat{\gamma}_{\text{MM}}$. Comment on these results.

5. We now consider two risk classes $\mathcal{C}_1$ and $\mathcal{C}_2$ and note LGD_1 and LGD_2 the corresponding LGD rates. We assume that the dependence function between LGD_1 and LGD_2 is given by the Gumbel-Barnett copula:

$$\mathbf{C}(u_1, u_2) = u_1 u_2 e^{-\theta \ln u_1 \ln u_2}$$

where θ is the copula parameter. Show that the density function of the copula is equal to:

$$c(u_1, u_2; \theta) = \left(1 - \theta - \theta \ln(u_1 u_2) + \theta^2 \ln u_1 \ln u_2\right) e^{-\theta \ln u_1 \ln u_2}$$

6. Deduce the log-likelihood function of the historical sample $\left\{(x_i, y_i)_{i=1}^{i=n}\right\}$.

7. We note $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\theta}$ the ML estimators of the parameters γ_1 (risk class $\mathcal{C}_1$), γ_2 (risk class $\mathcal{C}_2$) and θ (copula parameter). Why the ML estimator $\hat{\gamma}_1$ does not correspond to the ML estimator $\hat{\gamma}_{\text{ML}}$ except in the case $\hat{\theta} = 0$? Illustrate with an example.

11.5.6 Calculation of correlation bounds

1. Give the mathematical definition of the copula functions $\mathbf{C}^-$, $\mathbf{C}^\perp$ and $\mathbf{C}^+$. What is the probabilistic interpretation of these copulas?

2. We note $\boldsymbol{\tau}$ and LGD the default time and the loss given default of a counterparty. We assume that $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$ and $\text{LGD} \sim \mathcal{U}_{[0,1]}$.

   (a) Show that the dependence between $\boldsymbol{\tau}$ and LGD is maximum when the following equality holds:

   $$\text{LGD} + e^{-\lambda \boldsymbol{\tau}} - 1 = 0$$

   (b) Show that the linear correlation $\rho(\boldsymbol{\tau}, \text{LGD})$ verifies the following inequality:

   $$|\rho\langle \boldsymbol{\tau}, \text{LGD}\rangle| \leq \frac{\sqrt{3}}{2}$$

   (c) Comment on these results.

3. We consider two exponential default times $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ with parameters λ_1 and λ_2.

   (a) We assume that the dependence function between $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ is $\mathbf{C}^+$. Demonstrate that the following relationship is true:

   $$\boldsymbol{\tau}_1 = \frac{\lambda_2}{\lambda_1} \boldsymbol{\tau}_2$$

   (b) Show that there exists a function f such that $\boldsymbol{\tau}_2 = f(\boldsymbol{\tau}_2)$ when the dependence function is $\mathbf{C}^-$.

   (c) Show that the lower and upper bounds of the linear correlation satisfy the following relationship:

   $$-1 < \rho\langle \boldsymbol{\tau}_1, \boldsymbol{\tau}_2\rangle \leq 1$$

(d) In the more general case, show that the linear correlation of a random vector (X_1, X_2) cannot be equal to -1 if the support of the random variables X_1 and X_2 is $[0, +\infty]$.

4. We assume that (X_1, X_2) is a Gaussian random vector where $X_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$, $X_2 \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$ and ρ is the linear correlation between X_1 and X_2. We note $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ the set of parameters.

   (a) Find the probability distribution of $X_1 + X_2$.

   (b) Then show that the covariance between $Y_1 = e^{X_1}$ and $Y_2 = e^{X_2}$ is equal to:

$$\operatorname{cov}(Y_1, Y_2) = e^{\mu_1 + \frac{1}{2}\sigma_1^2} \cdot e^{\mu_2 + \frac{1}{2}\sigma_2^2} \cdot \left(e^{\rho\sigma_1\sigma_2} - 1\right)$$

   (c) Deduce the correlation between Y_1 and Y_2.

   (d) For which values of θ does the equality $\rho\langle Y_1, Y_2\rangle = +1$ hold? Same question when $\rho\langle Y_1, Y_2\rangle = -1$.

   (e) We consider the bivariate Black-Scholes model:

$$\begin{cases} \mathrm{d}S_1(t) = \mu_1 S_1(t)\,\mathrm{d}t + \sigma_1 S_1(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}S_2(t) = \mu_2 S_2(t)\,\mathrm{d}t + \sigma_2 S_2(t)\,\mathrm{d}W_2(t) \end{cases}$$

   with $\mathbb{E}\left[W_1(t) W_2(t)\right] = \rho t$. Deduce the linear correlation between $S_1(t)$ and $S_2(t)$. Find the limit case $\lim_{t\to\infty} \rho\langle S_1(t), S_2(t)\rangle$.

   (f) Comment on these results.

11.5.7 The bivariate Pareto copula

We consider the bivariate Pareto distribution:

$$\begin{aligned} \mathbf{F}(x_1, x_2) &= 1 - \left(\frac{\theta_1 + x_1}{\theta_1}\right)^{-\alpha} - \left(\frac{\theta_2 + x_2}{\theta_2}\right)^{-\alpha} + \\ &\quad \left(\frac{\theta_1 + x_1}{\theta_1} + \frac{\theta_2 + x_2}{\theta_2} - 1\right)^{-\alpha} \end{aligned}$$

where $x_1 \geq 0$, $x_2 \geq 0$, $\theta_1 > 0$, $\theta_2 > 0$ and $\alpha > 0$.

1. Show that the marginal functions of $\mathbf{F}(x_1, x_2)$ correspond to univariate Pareto distributions.

2. Find the copula function associated to the bivariate Pareto distribution.

3. Deduce the copula density function.

4. Show that the bivariate Pareto copula function has no lower tail dependence, but an upper tail dependence.

5. Do you think that the bivariate Pareto copula family can reach the copula functions $\mathbf{C}^-$, $\mathbf{C}^\perp$ and $\mathbf{C}^+$? Justify your answer.

6. Let X_1 and X_2 be two Pareto distributed random variables, whose parameters are (α_1, θ_1) and (α_2, θ_2).

(a) Show that the linear correlation between X_1 and X_2 is equal to 1 if and only if the parameters α_1 and α_2 are equal.

(b) Show that the linear correlation between X_1 and X_2 can never reached the lower bound -1.

(c) Build a new bivariate Pareto distribution by assuming that the marginal distributions are $\mathcal{P}(\alpha_1, \theta_1)$ and $\mathcal{P}(\alpha_2, \theta_2)$ and the dependence function is a bivariate Pareto copula with parameter α. What is the relevance of this approach for building bivariate Pareto distributions?

Chapter 12

Extreme Value Theory

This chapter is dedicated to tail (or extreme) risk modeling. Tail risk recovers two notions. The first one is related to rare events, meaning that a severe loss may occur with a very small probability. The second one concerns the magnitude of a loss that is difficult to reconciliate with the observed volatility of the portfolio. Of course, the two notions are connected, but the second is more frequent. For instance, stock market crashes are numerous since the end of the eighties. The study of these rare or abnormal events needs an appropriate framework to analyze their risk. This is the subject of this chapter. In a first section, we consider order statistics, which are very useful to understand the underlying concept of tail risk. Then, we present the extreme value theory (EVT) in the unidimensional case. Finally, the last section deals with the correlation issue between extreme risks.

12.1 Order statistics

12.1.1 Main properties

Let $X_1, \ldots, X_n$ be *iid* random variables, whose probability distribution is denoted by $\mathbf{F}$. We rank these random variables by increasing order:

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n-1:n} \leq X_{n:n}$$

$X_{i:n}$ is called the i^{th} order statistic in the sample of size n. We note $x_{i:n}$ the corresponding random variate or the value taken by $X_{i:n}$. We have:

$$\begin{aligned}
\mathbf{F}_{i:n}(x) &= \Pr\{X_{i:n} \leq x\} \\
&= \Pr\{\text{at least } i \text{ variables among } X_1, \ldots, X_n \text{ are less or equal to } x\} \\
&= \sum_{k=i}^{n} \Pr\{k \text{ variables among } X_1, \ldots, X_n \text{ are less or equal to } x\} \\
&= \sum_{k=i}^{n} \binom{n}{k} \mathbf{F}(x)^k \left(1 - \mathbf{F}(x)\right)^{n-k} \qquad (12.1)
\end{aligned}$$

We note f the density function of $\mathbf{F}$. We deduce that the density function of $X_{i:n}$ has the following expression:

$$\begin{aligned}
f_{i:n}(x) &= \sum_{k=i}^{n} \binom{n}{k} k\mathbf{F}(x)^{k-1} \left(1 - \mathbf{F}(x)\right)^{n-k} f(x) - \\
&\quad \sum_{k=i}^{n-1} \binom{n}{k} \mathbf{F}(x)^k (n-k) \left(1 - \mathbf{F}(x)\right)^{n-k-1} f(x) \qquad (12.2)
\end{aligned}$$

It follows that:

$$\begin{aligned} f_{i:n}(x) &= \sum_{k=i}^{n} \frac{n!}{(k-1)!\,(n-k)!} \mathbf{F}(x)^{k-1} (1-\mathbf{F}(x))^{n-k} f(x) - \\ &\quad \sum_{k=i}^{n-1} \frac{n!}{k!\,(n-k-1)!} \mathbf{F}(x)^{k} (1-\mathbf{F}(x))^{n-k-1} f(x) \\ &= \sum_{k=i}^{n} \frac{n!}{(k-1)!\,(n-k)!} \mathbf{F}(x)^{k-1} (1-\mathbf{F}(x))^{n-k} f(x) - \\ &\quad \sum_{k=i+1}^{n} \frac{n!}{(k-1)!\,(n-k)!} \mathbf{F}(x)^{k-1} (1-\mathbf{F}(x))^{n-k} f(x) \\ &= \frac{n!}{(i-1)!\,(n-i)!} \mathbf{F}(x)^{i-1} (1-\mathbf{F}(x))^{n-i} f(x) \end{aligned} \tag{12.3}$$

Remark 142 *When k is equal to n, the derivative of $(1-\mathbf{F}(x))^{n-k}$ is equal to zero. This explains that the second summation in Equation (12.2) does not include the case $k=n$.*

Example 126 *If $X_1,\ldots,X_n$ follow a uniform distribution $\mathcal{U}_{[0,1]}$, we obtain:*

$$\begin{aligned} \mathbf{F}_{i:n}(x) &= \sum_{k=i}^{n} \binom{n}{k} x^k (1-x)^{n-k} \\ &= \mathcal{IB}(x; i, n-i+1) \end{aligned}$$

where $\mathcal{IB}(x;\alpha,\beta)$ is the regularized incomplete beta function[1]*:*

$$\mathcal{IB}(x;\alpha,\beta) = \frac{1}{\mathfrak{B}(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$$

We deduce that $X_{i:n} \sim \mathcal{B}(i, n-i+1)$. It follows that the expected value of the order statistic $X_{i:n}$ is equal to:

$$\begin{aligned} \mathbb{E}[X_{i:n}] &= \mathbb{E}[\mathcal{B}(i, n-i+1)] \\ &= \frac{i}{n+1} \end{aligned}$$

We verify the stochastic ordering:

$$j > i \Rightarrow \mathbf{F}_{i:n} \succ \mathbf{F}_{j:n}$$

Indeed, we have:

$$\begin{aligned} \mathbf{F}_{i:n}(x) &= \sum_{k=i}^{n} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} \\ &= \sum_{k=i}^{j-1} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} + \sum_{k=j}^{n} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} \\ &= \mathbf{F}_{j:n}(x) + \sum_{k=i}^{j-1} \binom{n}{k} \mathbf{F}(x)^k (1-\mathbf{F}(x))^{n-k} \end{aligned}$$

meaning that $\mathbf{F}_{i:n}(x) \geq \mathbf{F}_{j:n}(x)$. In Figure 12.1, we illustrate this property when the random variables $X_1,\ldots,X_n$ follow the normal distribution $\mathcal{N}(0,1)$. We verify that $\mathbf{F}_{i:n}(x)$ increases with the ordering value i.

[1] It is also the Beta probability distribution $\mathcal{IB}(x;\alpha,\beta) = \Pr\{\mathcal{B}(\alpha,\beta) \leq x\}$.

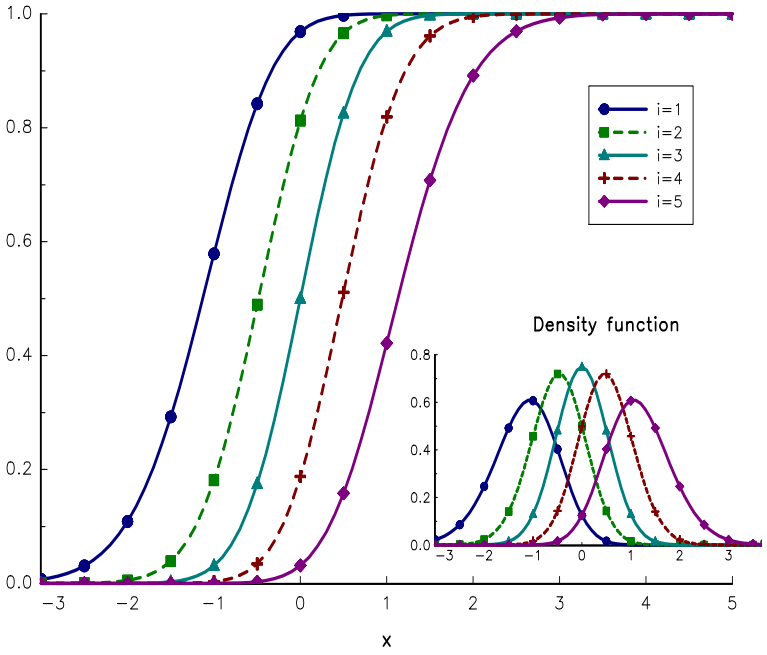


FIGURE 12.1: Distribution function $\mathbf{F}_{i:n}$ when the random variables $X_1, \ldots, X_n$ are Gaussian

12.1.2 Extreme order statistics

Two order statistics are particularly interesting for the study of rare events. They are the lowest and highest order statistics:

$$X_{1:n} = \min(X_1, \ldots, X_n)$$

and:

$$X_{n:n} = \max(X_1, \ldots, X_n)$$

We can find their probability distributions by setting $i = 1$ and $i = n$ in Formula (12.1). We can also retrieve their expression by noting that:

$$\begin{aligned}
\mathbf{F}_{1:n}(x) = \Pr\{\min(X_1, \ldots, X_n) \leq x\} &= 1 - \Pr\{\min(X_1, \ldots, X_n) \geq x\} \\
&= 1 - \Pr\{X_1 \geq x, X_2 \geq x, \ldots, X_n \geq x\} \\
&= 1 - \prod_{i=1}^{n} \Pr\{X_i \geq x\} \\
&= 1 - \prod_{i=1}^{n} (1 - \Pr\{X_i \leq x\}) \\
&= 1 - (1 - \mathbf{F}(x))^n
\end{aligned}$$

and:

$$\begin{aligned}
\mathbf{F}_{n:n}(x) = \Pr\{\max(X_1, \ldots, X_n) \leq x\} &= \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} \\
&= \prod_{i=1}^{n} \Pr\{X_i \leq x\} \\
&= \mathbf{F}(x)^n
\end{aligned}$$

We deduce that the density functions are equal to:

$$f_{1:n}(x) = n(1 - \mathbf{F}(x))^{n-1} f(x)$$

and

$$f_{n:n}(x) = n\mathbf{F}(x)^{n-1} f(x)$$

Let us consider an example with the Gaussian distribution $\mathcal{N}(0,1)$. Figure 12.2 shows the evolution of the density function $f_{n:n}$ with respect to the sample size n. We verify the stochastic ordering: $n > m \Rightarrow \mathbf{F}_{n:n} \succ \mathbf{F}_{m:m}$.

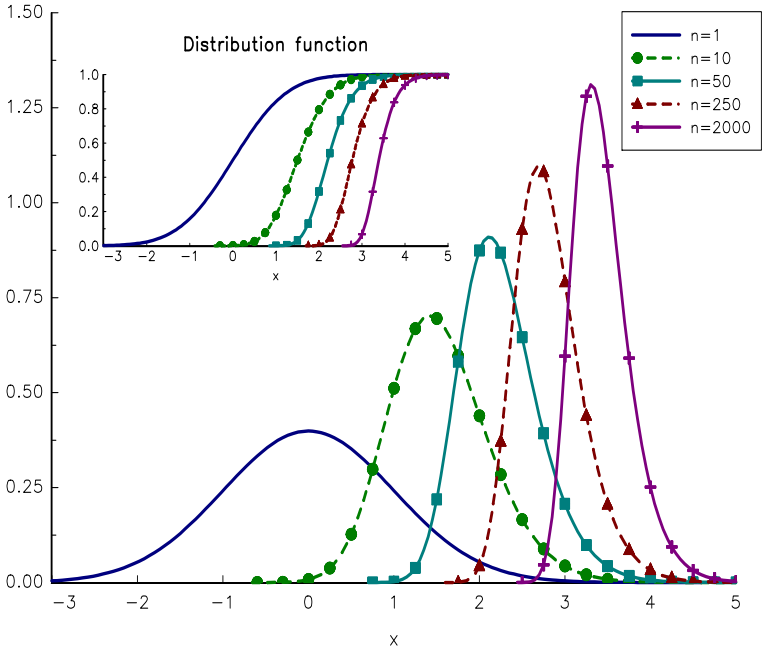


FIGURE 12.2: Density function $f_{n:n}$ of the Gaussian random variable $\mathcal{N}(0,1)$

Let us now illustrate the impact of the probability distribution tails on order statistics. We consider the daily returns of the MSCI USA index from 1995 to 2015. We consider three hypotheses:

$\mathcal{H}_1$ Daily returns are Gaussian, meaning that:

$$R_t = \hat{\mu} + \hat{\sigma} X_t$$

where $X_t \sim \mathcal{N}(0,1)$, $\hat{\mu}$ is the empirical mean of daily returns and $\hat{\sigma}$ is the daily standard deviation.

$\mathcal{H}_2$ Daily returns follow a Student's t distribution[2]:

$$R_t = \hat{\mu} + \hat{\sigma}\sqrt{\frac{\nu - 2}{\nu}} X_t$$

where $X_t \sim \mathbf{t}_\nu$. We consider two alternative assumptions: $\mathcal{H}_{2a} : \nu = 3$ and $\mathcal{H}_{2b} : \nu = 6$.

[2]We add the factor $\sqrt{\frac{\nu-2}{\nu}}$ in order to verify that $\text{var}(R_t) = \hat{\sigma}^2$.

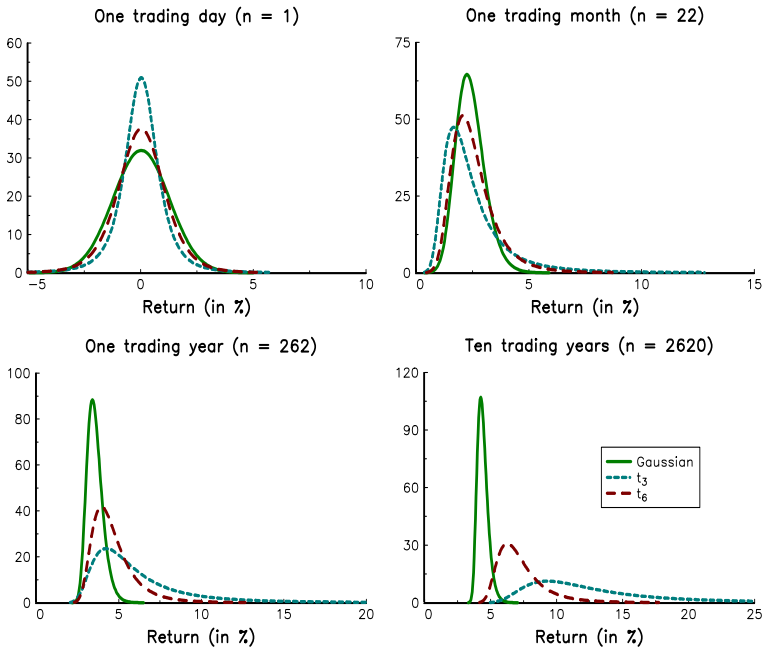


FIGURE 12.3: Density function of the maximum order statistic (daily return of the MSCI USA index, 1995-2015)

We represent the probability density function of $R_{n:n}$ for several values of n in Figure 12.3. When n is equal to one trading day, $R_{n:n}$ is exactly the daily return. We notice that it is difficult to observe the impact of the probability distribution tail. However, when n increases, the impact becomes more and more important. Order statistics allow amplifying local phenomena of probability distributions. In particular, extreme order statistics are a very useful tool to analyze left and right tails.

Remark 143 *The limit distributions of minima and maxima are given by the following results:*

$$\begin{aligned}\lim_{n\to\infty} \mathbf{F}_{1:n}(x) &= \lim_{n\to\infty} 1-(1-\mathbf{F}(x))^n \\ &= \begin{cases} 0 & \text{if } \mathbf{F}(x)=0 \\ 1 & \text{if } \mathbf{F}(x)>0 \end{cases}\end{aligned}$$

and:

$$\begin{aligned}\lim_{n\to\infty} \mathbf{F}_{n:n}(x) &= \lim_{n\to\infty} \mathbf{F}(x)^n \\ &= \begin{cases} 0 & \text{if } \mathbf{F}(x)<1 \\ 1 & \text{if } \mathbf{F}(x)=1 \end{cases}\end{aligned}$$

We deduce that the limit distributions are degenerate as they only take values of 0 *and* 1. *This property is very important, because it means that we cannot study extreme events by considering these limit distributions. This is why the extreme value theory is based on another convergence approach of extreme order statistics.*

12.1.3 Inference statistics

The common approach to estimate the parameters θ of the probability density function $f(x;\theta)$ is to maximize the log-likelihood function of a given sample $\{x_1,\ldots,x_T\}$:

$$\hat{\theta} = \arg\max \sum_{t=1}^{T} \ln f(x_t;\theta)$$

In a similar way, we can consider the sample[3] $\left\{x'_1,\ldots,x'_{n_S}\right\}$ of the order statistic $X_{i:n}$ and estimate the parameters θ by the method of maximum likelihood:

$$\hat{\theta}_{i:n} = \arg\max \boldsymbol{\ell}_{i:n}(\theta)$$

where:

$$\begin{aligned}\boldsymbol{\ell}_{i:n}(\theta) &= \sum_{s=1}^{n_S} \ln f_{i:n}(x'_s;\theta) \\ &= \sum_{s=1}^{n_S} \ln \frac{n!}{(i-1)!\,(n-i)!} \mathbf{F}(x'_s;\theta)^{i-1} \left(1-\mathbf{F}(x'_s;\theta)\right)^{n-i} f(x'_s;\theta)\end{aligned}$$

The computation of the log-likelihood function gives:

$$\begin{aligned}\boldsymbol{\ell}_{i:n}(\theta) &= n_S \ln n! - n_S \ln (i-1)! - n_S \ln (n-i)! + \\ &\quad (i-1)\sum_{s=1}^{n_S} \ln \mathbf{F}(x'_s;\theta) + (n-i)\sum_{s=1}^{n_S} \ln\left(1-\mathbf{F}(x'_s;\theta)\right) + \\ &\quad \sum_{s=1}^{n_S} \ln f(x'_s;\theta)\end{aligned}$$

By definition, the traditional ML estimator is equal to new ML estimator when $n=1$ and $i=1$:

$$\hat{\theta} = \hat{\theta}_{1:1}$$

In the other cases ($n>1$), there is no reason that the two estimators coincide exactly:

$$\hat{\theta}_{i:n} \neq \hat{\theta}$$

However, if the random variates are drawn from the distribution function $X \sim \mathbf{F}(x;\theta)$, we can test the hypothesis $\mathcal{H}: \hat{\theta}_{i:n} = \theta$ for all n and $i \leq n$. If two estimates $\hat{\theta}_{i:n}$ and $\hat{\theta}_{i':n'}$ are very different, this indicates that the distribution function is certainly not appropriate for modeling the random variable X.

Let us consider the previous example with the returns of the MSCI USA index. We assume that the daily returns can be modeled with the Student's t distribution:

$$\frac{R_t - \mu}{\sigma} \sim \mathbf{t}_\nu$$

The vector of parameters to estimate is then $\theta = (\mu,\sigma)$. In Tables 12.1, 12.2 and 12.3, we report the values taken by the ML estimator $\hat{\sigma}_{i:n}$ obtained by considering several order statistics and three values of ν. For instance, the ML estimate $\hat{\sigma}_{1:1}$ in the case of the $\mathbf{t}_1$ distribution is equal to 50 bps. We notice that the values taken by $\hat{\sigma}_{i:n}$ are not very stable

[3] The size of the sample n_S is equal to the size of the original sample T divided by n.

TABLE 12.1: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_1$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	50									
2	48	49								
3	44	54	44							
4	41	53	53	41						
5	38	52	55	51	37					
6	35	51	56	56	48	33				
7	32	49	55	56	55	45	29			
8	31	48	53	55	54	50	43	26		
9	29	46	55	56	57	55	49	40	25	
10	28	43	53	58	57	56	53	48	37	20

TABLE 12.2: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_6$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	88									
2	89	87								
3	91	91	85							
4	95	92	89	87						
5	98	99	87	90	88					
6	101	104	95	88	92	89				
7	101	112	100	88	94	95	89			
8	102	116	103	89	85	89	98	89		
9	105	121	117	97	85	86	94	101	88	
10	105	123	120	108	91	87	92	99	104	88

TABLE 12.3: ML estimate of σ (in bps) for the probability distribution $\mathbf{t}_\infty$

Size n	Order i									
	1	2	3	4	5	6	7	8	9	10
1	125									
2	125	124								
3	136	116	129							
4	147	116	112	140						
5	155	133	103	114	150					
6	163	142	118	107	122	157				
7	171	152	125	105	117	134	162			
8	175	165	130	106	99	111	139	170		
9	180	174	155	122	95	99	128	152	171	
10	183	182	162	136	110	100	111	127	155	181

with respect to i and n. This indicates that the three probability distribution functions ($\mathbf{t}_1$, $\mathbf{t}_6$ and $\mathbf{t}_\infty$) are not well appropriate to represent the index returns. In Figure 12.4, we have reported the corresponding annualized volatility[4] calculated from the order statistics $R_{i:10}$. In the case of the $\mathbf{t}_1$ distribution, we notice that it is lower for median order statistics than extreme order statistics. The $\mathbf{t}_1$ distribution has then the property to overestimate extreme events. In the case of the Gaussian (or $\mathbf{t}_\infty$) distribution, we obtain contrary results. The Gaussian distribution has the property to underestimate extreme events. In order to compensate this bias, the method of maximum likelihood applied to extreme order statistics will overestimate the volatility.

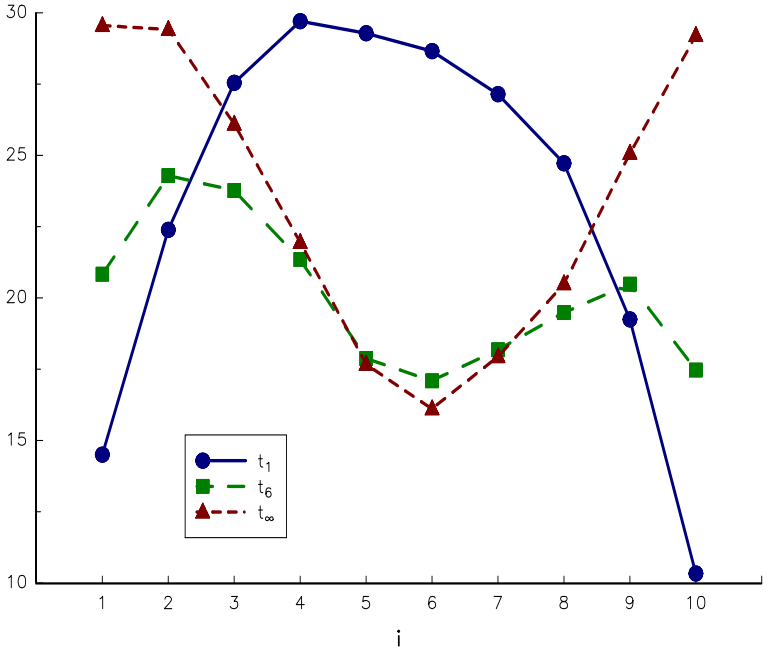


FIGURE 12.4: Annualized volatility (in %) calculated from the order statistics $R_{i:10}$

Remark 144 *The approach based on extreme order statistics to calculate the volatility is then a convenient way to reduce the under-estimation of the Gaussian value-at-risk.*

12.1.4 Extension to dependent random variables

Let us now assume that $X_1, \ldots, X_n$ are not *iid*. We note $\mathbf{C}$ the copula of the corresponding random vector. It follows that:

$$\begin{aligned} \mathbf{F}_{n:n}(x) &= \Pr\{X_{n:n} \leq x\} \\ &= \Pr\{X_1 \leq x, \ldots, X_n \leq x\} \\ &= \mathbf{C}(\mathbf{F}_1(x), \ldots, \mathbf{F}_n(x)) \end{aligned}$$

[4] The annualized volatility takes the value $\sqrt{260} \cdot c \cdot \hat{\sigma}_{i:n}$ where the constant c is equal to $\sqrt{\nu/(\nu-2)}$. In the case of the $\mathbf{t}_1$ distribution, c is equal to 3.2.

and:

$$\begin{aligned}\mathbf{F}_{1:n}(x) &= \Pr\{X_{1:n} \leq x\} \\ &= 1 - \Pr\{X_{1:n} \geq x\} \\ &= 1 - \Pr\{X_1 \geq x, \ldots, X_n \geq x\} \\ &= 1 - \breve{\mathbf{C}}(1 - \mathbf{F}_1(x), \ldots, 1 - \mathbf{F}_n(x))\end{aligned}$$

where $\breve{\mathbf{C}}$ is the survival copula associated to $\mathbf{C}$.

Remark 145 *In the case of the product copula and identical probability distributions, we retrieve the previous results:*

$$\begin{aligned}\mathbf{F}_{n:n}(x) &= \mathbf{C}^{\perp}(\mathbf{F}(x), \ldots, \mathbf{F}(x)) \\ &= \mathbf{F}(x)^n\end{aligned}$$

and:

$$\begin{aligned}\mathbf{F}_{1:n}(x) &= 1 - \mathbf{C}^{\perp}(1 - \mathbf{F}(x), \ldots, 1 - \mathbf{F}(x)) \\ &= 1 - (1 - \mathbf{F}(x))^n\end{aligned}$$

If we are interested in other order statistics, we use the following formula given in Georges *et al.* (2001):

$$\mathbf{F}_{i:n}(x) = \sum_{k=i}^{n} \left[\sum_{l=i}^{k} (-1)^{k-l} \binom{k}{l} \sum_{\mathbf{v}(\mathbf{F}_1(x),\ldots,\mathbf{F}_n(x)) \in \mathcal{Z}(n-k,n)} \mathbf{C}(u_1, \ldots, u_n) \right]$$

where:

$$\mathcal{Z}(m,n) = \left\{ \mathbf{v} \in [0,1]^n \mid v_i \in \{u_i, 1\}, \sum_{i=1}^{n} \mathbf{1}\{v_i = 1\} = m \right\}$$

In order to understand this formula, we consider the case $n = 3$. We have[5]:

$$\begin{aligned}\mathbf{F}_{1:3}(x) &= \mathbf{F}_1(x) + \mathbf{F}_2(x) + \mathbf{F}_3(x) - \\ &\quad \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), 1) - \mathbf{C}(\mathbf{F}_1(x), 1, \mathbf{F}_3(x)) - \mathbf{C}(1, \mathbf{F}_2(x), \mathbf{F}_3(x)) + \\ &\quad \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x)) \\ \mathbf{F}_{2:3}(x) &= \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), 1) + \mathbf{C}(\mathbf{F}_1(x), 1, \mathbf{F}_3(x)) + \mathbf{C}(1, \mathbf{F}_2(x), \mathbf{F}_3(x)) - \\ &\quad 2\mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x)) \\ \mathbf{F}_{3:3}(x) &= \mathbf{C}(\mathbf{F}_1(x), \mathbf{F}_2(x), \mathbf{F}_3(x))\end{aligned}$$

We verify that:

$$\mathbf{F}_{1:3}(x) + \mathbf{F}_{2:3}(x) + \mathbf{F}_{3:3}(x) = \mathbf{F}_1(x) + \mathbf{F}_2(x) + \mathbf{F}_3(x)$$

The dependence structure has a big impact on the distribution of order statistics. For instance, if we assume that $X_1, \ldots, X_n$ are *iid*, we obtain:

$$\mathbf{F}_{n:n}(x) = \mathbf{F}(x)^n$$

[5] Because $\mathbf{C}(\mathbf{F}_1(x), 1, 1) = \mathbf{F}_1(x)$.

If the copula function is the upper Fréchet copula, this result becomes:

$$\begin{aligned}\mathbf{F}_{n:n}(x) &= \mathbf{C}^{+}(\mathbf{F}(x),\ldots,\mathbf{F}(x)) \\ &= \min(\mathbf{F}(x),\ldots,\mathbf{F}(x)) \\ &= \mathbf{F}(x)\end{aligned}$$

This implies that the occurrence probability of extreme events is lower in this second case.

We consider n Weibull default times $\boldsymbol{\tau}_i \sim \mathcal{W}(\lambda_i,\gamma_i)$. The survival function is equal to $\mathbf{S}_i(t) = \exp(-\lambda_i t^{\gamma_i})$. The hazard rate $\lambda_i(t)$ is then $\lambda_i\gamma_i t^{\gamma_i-1}$ and the expression of the density is $f_i(t) = \lambda_i(t)\,\mathbf{S}_i(t)$. If we assume that the survival copula is the Gumbel-Hougaard copula with parameter $\theta \geq 1$, the survival function of the first-to-default time is equal to:

$$\begin{aligned}\mathbf{S}_{1:n}(t) &= \exp\left(-\left((-\ln\mathbf{S}_1(t))^{\theta}+\ldots+(-\ln\mathbf{S}_n(t))^{\theta}\right)^{1/\theta}\right) \\ &= \exp\left(-\left(\sum_{i=1}^{n}\lambda_i^{\theta}t^{\theta\gamma_i}\right)^{1/\theta}\right)\end{aligned}$$

We deduce the expression of the density function:

$$\begin{aligned}f_{1:n}(t) &= \left(\sum_{i=1}^{n}\lambda_i^{\theta}t^{\theta\gamma_i}\right)^{1/\theta-1}\cdot\left(\sum_{i=1}^{n}\gamma_i\lambda_i^{\theta}t^{\theta\gamma_i-1}\right)\cdot \\ &\quad \exp\left(-\left(\sum_{i=1}^{n}\lambda_i^{\theta}t^{\theta\gamma_i}\right)^{1/\theta}\right)\end{aligned}$$

In the case where the default times are identically distributed, the first-to-default time is a Weibull default time: $\boldsymbol{\tau}_{1:n} \sim \mathcal{W}\left(n^{1/\theta}\lambda,\gamma\right)$. In Figure 12.5, we report the density function $f_{1:10}(t)$ for the parameters $\lambda = 3\%$ and $\gamma = 2$. We notice that the parameter θ of the copula function has a big influence on the first-to-default time. The case $\theta = 1$ corresponds to the product copula and we retrieve the previous result:

$$\mathbf{S}_{1:n}(t) = \mathbf{S}(t)^n$$

When the Gumbel-Hougaard is the upper Fréchet copula ($\theta \to \infty$), we verify that the density function of $\boldsymbol{\tau}_{1:n}$ is this of any default time $\boldsymbol{\tau}_i$.

12.2 Univariate extreme value theory

The extreme value theory consists in studying the limit distribution of extreme order statistics $X_{1:n}$ and $X_{n:n}$ when the sample size tends to infinity. We will see that the limit distribution converges to three probability distributions. This result will help to evaluate stress scenarios and to build a stress testing framework.

Remark 146 *In what follows, we only consider the largest order statistic $X_{n:n}$. Indeed, the minimum order statistic $X_{1:n}$ can be defined with respect to the maximum order statistic $Y_{n:n}$ by setting $Y_i = -X_i$:*

$$\begin{aligned}X_{1:n} &= \min(X_1,\ldots,X_n) \\ &= \min(-Y_1,\ldots,-Y_n) \\ &= -\max(Y_1,\ldots,Y_n) \\ &= -Y_{n:n}\end{aligned}$$

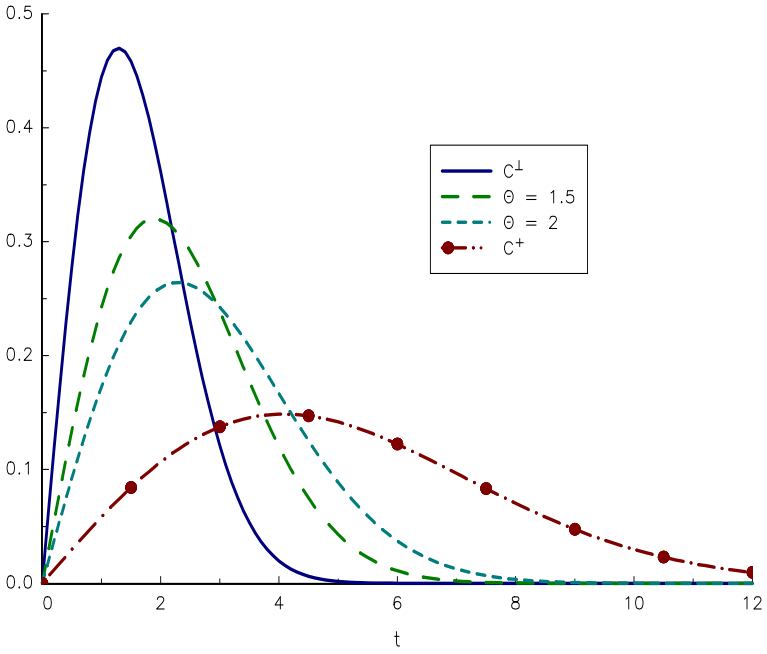


FIGURE 12.5: Density function of the first-to-default time $\boldsymbol{\tau}_{1:10}$

12.2.1 Fisher-Tippet theorem

We follow Embrechts *et al.* (1997) for the formulation of the Fisher-Tippet theorem. Let $X_1, \ldots, X_n$ be a sequence of *iid* random variables, whose distribution function is $\mathbf{F}$. If there exist two constants a_n and b_n and a non-degenerate distribution function $\mathbf{G}$ such that:

$$\lim_{n\to\infty} \Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} = \mathbf{G}(x) \tag{12.4}$$

then $\mathbf{G}$ can be classified as one of the following three types[6]:

Type I	(Gumbel)	$\boldsymbol{\Lambda}(x) = \exp\left(-e^{-x}\right)$
Type II	(Fréchet)	$\boldsymbol{\Phi}_\alpha(x) = \mathbb{1}(x \geq 0) \cdot \exp\left(-x^{-\alpha}\right)$
Type III	(Weibull)	$\boldsymbol{\Psi}_\alpha(x) = \mathbb{1}(x \leq 0) \cdot \exp\left(-(-x)^{\alpha}\right)$

The distribution functions $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_\alpha$ et $\boldsymbol{\Psi}_\alpha$ are called extreme value distributions. The Fisher-Tippet theorem is very important, because the set of extreme value distributions is very small although the set of distribution functions is very large. We can draw a parallel with the normal distribution and the sum of random variables. In some sense, the Fisher-Tippet theorem provides an extreme value analog of the central limit theorem.

[6] In terms of probability density functions, we have:

$$g(x) = \begin{cases} \exp\left(-x - e^{-x}\right) & \text{(Gumbel)} \\ \mathbb{1}(x \geq 0) \cdot \alpha x^{-(1+\alpha)} \cdot \exp\left(-x^{-\alpha}\right) & \text{(Fréchet)} \\ \mathbb{1}(x \leq 0) \cdot \alpha(-x)^{\alpha-1} \cdot \exp\left(-(-x)^{\alpha}\right) & \text{(Weibull)} \end{cases}$$

Let us consider the case of exponential random variables, whose probability distribution is $\mathbf{F}(x) = 1 - \exp(-\lambda x)$. We have[7]:

$$\begin{aligned}\lim_{n\to\infty} \mathbf{F}_{n:n}(x) &= \lim_{n\to\infty} \left(1 - e^{-\lambda x}\right)^n \\ &= \lim_{n\to\infty} \left(1 - \frac{ne^{-\lambda x}}{n}\right)^n \\ &= \lim_{n\to\infty} \exp\left(-ne^{-\lambda x}\right) \\ &= 0\end{aligned}$$

We verify that the limit distribution is degenerate. If we consider the affine transformation with $a_n = 1/\lambda$ et $b_n = (\ln n)/\lambda$, we obtain:

$$\begin{aligned}\Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \Pr\{X_{n:n} \leq a_n x + b_n\} \\ &= \left(1 - e^{-\lambda(a_n x + b_n)}\right)^n \\ &= \left(1 - e^{-x - \ln n}\right)^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathbf{G}(x) &= \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \\ &= \exp\left(-e^{-x}\right)\end{aligned}$$

It follows that the limit distribution of the affine transformation is not degenerate. In Figure 12.6, we illustrate the convergence of $\mathbf{F}^n(a_n x + b_n)$ to the Gumbel distribution $\mathbf{\Lambda}(x)$.

Example 127 *If we consider the Pareto distribution, we have:*

$$\mathbf{F}(x) = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

The normalizing constants are $a_n = x_- n^{1/\alpha}$ *and* $b_n = 0$. *We obtain:*

$$\begin{aligned}\Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \left(1 - \left(\frac{x_- n^{1/\alpha} x}{x_-}\right)^{-\alpha}\right)^n \\ &= \left(1 - \frac{x^{-\alpha}}{n}\right)^n\end{aligned}$$

We deduce that the law of the maximum tends to the Fréchet distribution:

$$\lim_{n\to\infty} \left(1 - \frac{x^{-\alpha}}{n}\right)^n = \exp\left(-x^{-\alpha}\right)$$

[7] Because we have:

$$\begin{aligned}\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \\ &= \exp(x)\end{aligned}$$

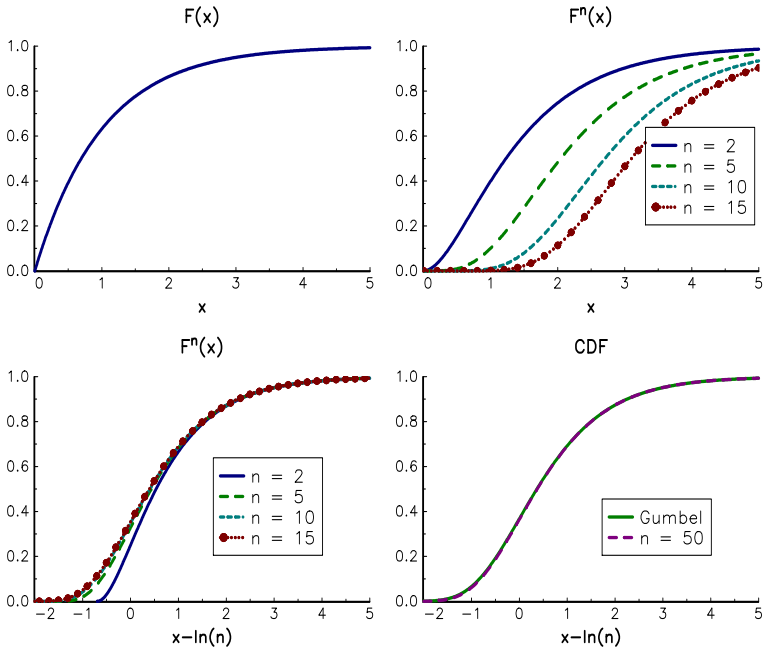


FIGURE 12.6: Max-convergence of the exponential distribution $\mathcal{E}(1)$ to the Gumbel distribution

Example 128 *For the uniform distribution, the normalizing constants become* $a_n = n^{-1}$ *and* $b_n = 1$ *and we obtain the Weibull distribution with* $\alpha = 1$*:*

$$\begin{aligned}\lim_{n\to\infty} \Pr\left\{\frac{X_{n:n} - b_n}{a_n} \leq x\right\} &= \left(1 + \frac{x}{n}\right)^n \\ &= \exp(x)\end{aligned}$$

12.2.2 Maximum domain of attraction

The application of the Fisher-Tippet theorem is limited because it can be extremely difficult to find the normalizing constants and the extreme value distribution for a given probability distribution $\mathbf{F}$. However, the graphical representation of $\mathbf{\Lambda}$, $\mathbf{\Phi}_\alpha$ and $\mathbf{\Psi}_\alpha$ given in Figure 12.7 already provides some information. For instance, the Weibull probability distribution concerns random variables that are right bounded. This is why it has less interest in finance than the Fréchet or Gumbel distribution functions[8]. We also notice some differences in the shape of the curves. In particular, the Gumbel distribution is more '*normal*' than the Fréchet distribution, whose shape and tail depend on the parameter α (see Figure 12.8).

We say that the distribution function $\mathbf{F}$ belongs to the max-domain of attraction of the distribution function $\mathbf{G}$ and we write $\mathbf{F} \in \text{MDA}(\mathbf{G})$ if the distribution function of the normalized maximum converges to $\mathbf{G}$. For instance, we have already seen that $\mathcal{E}(\lambda) \in \text{MDA}(\mathbf{\Lambda})$. In what follows, we indicate how to characterize the set $\text{MDA}(\mathbf{G})$ and which normalizing constants are[9].

[8] However, the Weibull probability distribution is related to the Fréchet probability distribution thanks to the relationship $\mathbf{\Psi}_\alpha(x) = \mathbf{\Phi}_\alpha\left(-x^{-1}\right)$.

[9] Most of the following results come from Resnick (1987).

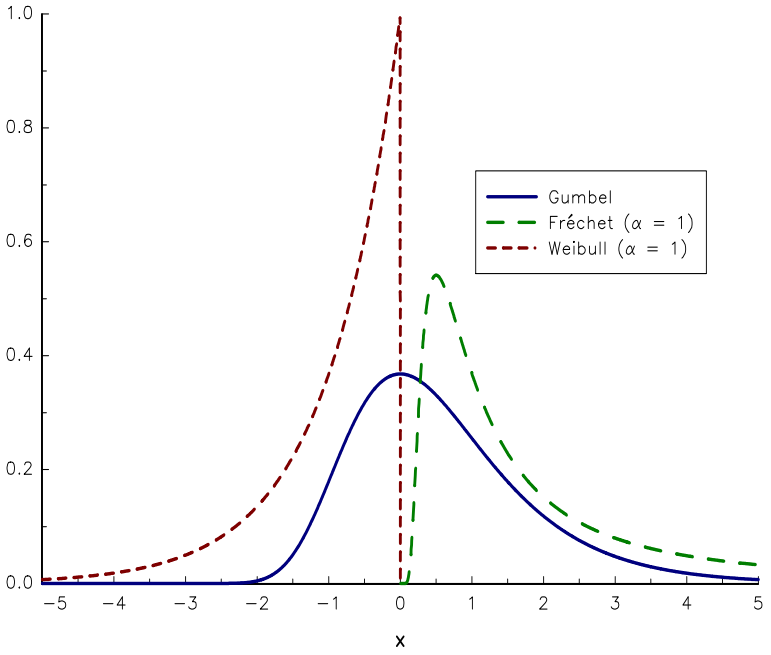


FIGURE 12.7: Density function of $\mathbf{\Lambda}$, $\mathbf{\Phi}_1$ and $\mathbf{\Psi}_1$

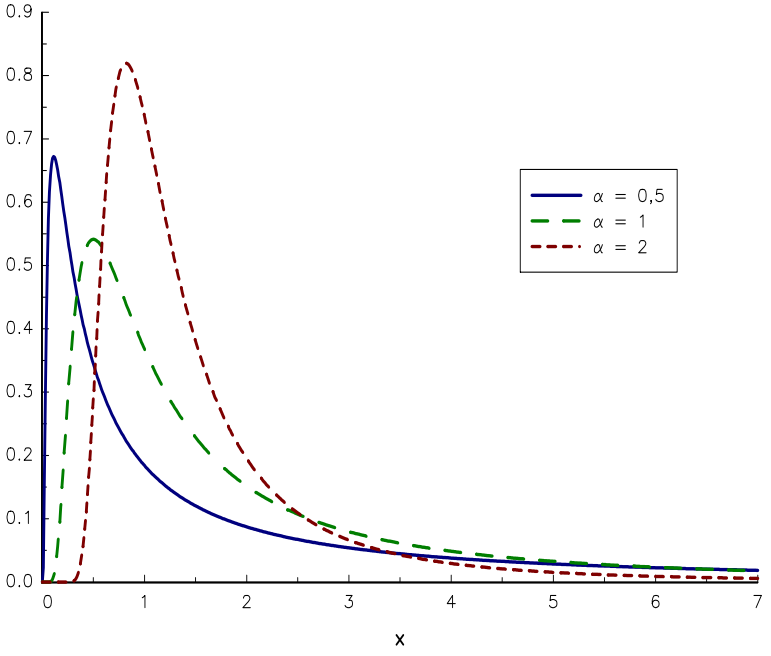


FIGURE 12.8: Density function of the Fréchet probability distribution

12.2.2.1 MDA of the Gumbel distribution

$\mathbf{F} \in \text{MDA}(\mathbf{\Lambda})$ if and only if there exists a function $h(t)$ such that:

$$\lim_{t \to x_0} \frac{1 - \mathbf{F}(t + x \cdot h(t))}{1 - \mathbf{F}(t)} = \exp(-x)$$

where $x_0 \leq \infty$. The normalizing constants are then $a_n = h\left(\mathbf{F}^{-1}\left(1 - n^{-1}\right)\right)$ and $b_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right)$.

The previous characterization of $\text{MDA}(\mathbf{\Lambda})$ is difficult to use because we have to define the function $h(t)$. However, we can show that if the distribution function $\mathbf{F}$ is C^2, a sufficient condition is:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = -1$$

For instance, in the case of the exponential distribution, we have $\mathbf{F}(x) = 1 - \exp(-\lambda x)$, $\partial_x \mathbf{F}(x) = \lambda \exp(-\lambda x)$ and $\partial_x^2 \mathbf{F}(x) = -\lambda^2 \exp(-\lambda x)$. We verify that:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = \lim_{x \to \infty} \frac{\exp(-\lambda x) \cdot \left(-\lambda^2 \exp(-\lambda x)\right)}{(\lambda \exp(-\lambda x))^2} = -1$$

If we consider the Gaussian distribution $\mathcal{N}(0,1)$, we have $\mathbf{F}(x) = \Phi(x)$, $\partial_x \mathbf{F}(x) = \phi(x)$ and $\partial_x^2 \mathbf{F}(x) = -x\phi(x)$. Using L'Hospital's rule, we deduce that:

$$\lim_{x \to \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial_x^2 \mathbf{F}(x)}{(\partial_x \mathbf{F}(x))^2} = \lim_{x \to \infty} -\frac{x \cdot \Phi(-x)}{\phi(x)} = -1$$

12.2.2.2 MDA of the Fréchet distribution

We say that a function f is regularly varying with index α and we write $f \in \text{RV}_\alpha$ if we have:

$$\lim_{t \to \infty} \frac{f(t \cdot x)}{f(t)} = x^\alpha$$

for every $x > 0$. We can then show the following theorem: $\mathbf{F} \in \text{MDA}(\mathbf{\Phi}_\alpha)$ if and only if $1 - \mathbf{F} \in \text{RV}_{-\alpha}$, and the normalizing constants are $a_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right)$ and $b_n = 0$.

Using the previous theorem, we deduce that the distribution function $\mathbf{F} \in \text{MDA}(\mathbf{\Phi}_\alpha)$ if it satisfies the following condition:

$$\lim_{t \to \infty} \frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} = x^{-\alpha}$$

If we apply this result to the Pareto distribution, we obtain:

$$\begin{aligned} \lim_{t \to \infty} \frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} &= \lim_{t \to \infty} \frac{(t \cdot x / x_-)^{-\alpha}}{(t / x_-)^{-\alpha}} \\ &= x^{-\alpha} \end{aligned}$$

We deduce that $1 - \mathbf{F} \in \text{RV}_{-\alpha}$, $\mathbf{F} \in \text{MDA}(\mathbf{\Phi}_\alpha)$, $a_n = \mathbf{F}^{-1}\left(1 - n^{-1}\right) = x_- n^{1/\alpha}$ and $b_n = 0$.

Remark 147 *The previous theorem suggests that:*

$$\frac{1 - \mathbf{F}(t \cdot x)}{1 - \mathbf{F}(t)} \approx x^{-\alpha}$$

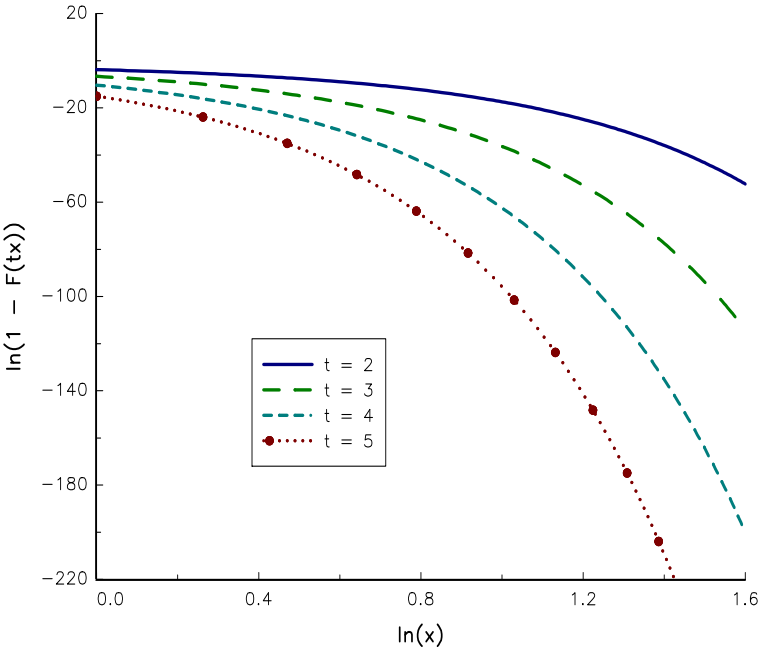


FIGURE 12.9: Graphical validation of the regular variation property for the normal distribution $\mathcal{N}(0,1)$

when t is sufficiently large. This means that we must observe a linear relationship between $\ln(x)$ *and* $\ln(1-\mathbf{F}(t\cdot x))$*:*

$$\ln(1-\mathbf{F}(t\cdot x)) \approx \ln(1-\mathbf{F}(t)) - \alpha\ln(x)$$

This property can be used to check graphically if a given distribution function belongs or not to the maximum domain of attraction of the Fréchet distribution. For instance, we observe that $\mathcal{N}(0,1) \notin \text{MDA}(\mathbf{\Phi}_\alpha)$ *in Figure 12.9, because the curve is not a straight line.*

12.2.2.3 MDA of the Weibull distribution

For the Weibull distribution, we can show that $\mathbf{F} \in \text{MDA}(\mathbf{\Psi}_\alpha)$ if and only if $1-\mathbf{F}\left(x_0 - x^{-1}\right) \in \text{RV}_{-\alpha}$ and $x_0 < \infty$. The normalizing constants are $a_n = x_0 - \mathbf{F}^{-1}\left(1-n^{-1}\right)$ and $b_n = x_0$.

If we consider the uniform distribution with $x_0 = 1$, we have:

$$\mathbf{F}\left(x_0 - x^{-1}\right) = 1 - \frac{1}{x}$$

and:

$$\begin{aligned}\lim_{t\to\infty}\frac{1-\mathbf{F}\left(1-t^{-1}x^{-1}\right)}{1-\mathbf{F}\left(1-t^{-1}\right)} &= \lim_{t\to\infty}\frac{t^{-1}x^{-1}}{t^{-1}}\\ &= x^{-1}\end{aligned}$$

We deduce that $\mathbf{F} \in \text{MDA}(\mathbf{\Psi}_1)$, $a_n = 1-\mathbf{F}^{-1}\left(1-n^{-1}\right) = n^{-1}$ and $b_n = 1$.

TABLE 12.4: Maximum domain of attraction and normalizing constants of some distribution functions

Distribution	$\mathbf{G}(x)$	a_n	b_n
$\mathcal{E}(\lambda)$	$\mathbf{\Lambda}$	λ^{-1}	$\lambda^{-1}\ln n$
$\mathcal{G}(\alpha,\beta)$	$\mathbf{\Lambda}$	β^{-1}	$\beta^{-1}(\ln n+(\alpha-1)\ln(\ln n)-\ln\Gamma(\alpha))$
$\mathcal{N}(0,1)$	$\mathbf{\Lambda}$	$(2\ln n)^{-1/2}$	$\frac{4\ln n-\ln 4\pi-\ln(\ln n)}{2\sqrt{2\ln n}}$
$\mathcal{LN}(\mu,\sigma^2)$	$\mathbf{\Lambda}$	$\sigma(2\ln n)^{-1/2}b_n$	$\exp\left(\mu+\sigma\left(\frac{4\ln n-\ln 4\pi+\ln(\ln n)}{2\sqrt{2\ln n}}\right)\right)$
$\mathcal{P}(\alpha,x_-)$	$\mathbf{\Phi}_\alpha$	$x_-n^{1/\alpha}$	0
$\mathcal{LG}(\alpha,\beta)$	$\mathbf{\Phi}_\beta$	$\frac{\left(n(\ln n)^{\alpha-1}\right)^{1/\beta}}{\Gamma(\alpha)}$	0
t_ν	$\mathbf{\Phi}_\nu$	$\mathbf{T}_\nu^{-1}\left(1-n^{-1}\right)$	0
$\mathcal{U}_{[0,1]}$	$\mathbf{\Psi}_1$	n^{-1}	1
$\mathcal{B}(\alpha,\beta)$	$\mathbf{\Psi}_\alpha$	$\left(\frac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta+1)}\right)^{-1/\beta}$	1

Source: Embrechts *et al.* (1997).

12.2.2.4 Main results

In Table 12.4, we report the maximum domain of attraction and normalizing constants of some well-known distribution functions.

Remark 148 *Let $\mathbf{G}(x)$ be the non-degenerate distribution of $X_{n:n}$. We note a_n and b_n the normalizing constants. We consider the linear transformation $Y=cX+d$ with $c>0$. Because we have $Y_{n:n}=cX_{n:n}+d$, we deduce that:*

$$\begin{aligned}\mathbf{G}(x) &= \lim_{n\to\infty}\Pr\{X_{n:n}\leq a_nx+b_n\}\\ &= \lim_{n\to\infty}\Pr\left\{\frac{Y_{n:n}-d}{c}\leq a_nx+b_n\right\}\\ &= \lim_{n\to\infty}\Pr\{Y_{n:n}\leq a_ncx+b_nc+d\}\\ &= \lim_{n\to\infty}\Pr\{Y_{n:n}\leq a'_nx+b'_n\}\end{aligned}$$

where $a'_n=a_nc$ and $b'_n=b_nc+d$. This means that $\mathbf{G}(x)$ is also the non-degenerate distribution of $Y_{n:n}$, and a'_n and b'_n are the normalizing constants. For instance, if we consider the distribution function $\mathcal{N}(\mu,\sigma^2)$, we deduce that the normalizing constants are:

$$a_n=\sigma(2\ln n)^{-1/2}$$

and:

$$b_n = \mu + \sigma\left(\frac{4\ln n - \ln 4\pi + \ln(\ln n)}{2\sqrt{2\ln n}}\right)$$

The normalizing constants are uniquely defined. In the case of the Gaussian distribution $\mathcal{N}(0,1)$, they are equal to $a_n = h(b_n) = b_n/\left(1+b_n^2\right)$ and $b_n = \Phi^{-1}\left(1-n^{-1}\right)$. In Table 12.4, we report an approximation which is not necessarily unique. For instance, Gasull *et al.* (2015) propose the following alternative value of b_n:

$$b_n \approx \sqrt{\ln\left(\frac{n^2}{2\pi}\right) - \ln\left(\ln\left(\frac{n^2}{2\pi}\right)\right) + \frac{\ln(0.5+\ln n^2)-2}{\ln n^2 - \ln 2\pi}}$$

and show that this solution is more accurate than the classical approximation.

12.2.3 Generalized extreme value distribution

12.2.3.1 Definition

From a statistical point of view, the previous results of the extreme value theory are difficult to use. Indeed, they are many issues concerning the choice of the distribution function, the normalizing constants or the convergence rate as explained by Coles (2001):

> "*The three types of limits that arise in Theorem 12.2.1 have distinct forms of behavior, corresponding to the different forms of tail behaviour for the distribution function* $\mathbf{F}$ *of the* X_i*. This can be made precise by considering the behavior of the limit distribution* $\mathbf{G}$ *at* x_+*, its upper end-point. For the Weibull distribution* x_+ *is finite, while for both the Fréchet and Gumbel distributions* $x_+ = \infty$*. However, the density of* $\mathbf{G}$ *decays exponentially for the Gumbel distribution and polynomially for the Fréchet distribution, corresponding to relatively different rates of decay in the tail of* $\mathbf{F}$*. It follows that in applications the three different families give quite different representations of extreme value behavior. In early applications of extreme value theory, it was usual to adopt one of the three families, and then to estimate the relevant parameters of that distribution. But there are two weakness: first, a technique is required to choose which of the three families is most appropriate for the data at hand; second, once such a decision is made, subsequent inferences presume this choice to be correct, and do not allow for the uncertainty such a selection involves, even though this uncertainty may be substantial*".

In practice, the statistical inference on extreme values takes another route. Indeed, the three types can be combined into a single distribution function:

$$\mathbf{G}(x) = \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

defined on the support $\Delta = \left\{x : 1+\xi\sigma^{-1}(x-\mu) > 0\right\}$. It is known as the '*generalized extreme value*' distribution and we denote it by $\mathcal{GEV}(\mu,\sigma,\xi)$. We obtain the following cases:

- the limit case $\xi \to 0$ corresponds to the Gumbel distribution;
- $\xi = -\alpha^{-1} > 0$ defines the Fréchet distribution;
- the Weibull distribution is obtained by considering $\xi = -\alpha^{-1} < 0$.

We also notice that the parameters μ and σ are the limits of the normalizing constants b_n and a_n. The corresponding density function is equal to:

$$g(x) = \frac{1}{\sigma} \cdot \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-(1+\xi)/\xi} \cdot \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

It is represented in Figure 12.10 for various values of the parameters. We notice that μ is a parameter of localization, σ controls the standard deviation and ξ is related to the tail of the distribution. The parameters can be estimated using the method of maximum likelihood and we obtain:

$$\boldsymbol{\ell}_t = -\ln\sigma - \left(\frac{1+\xi}{\xi}\right)\ln\left(1 + \xi\left(\frac{x_t-\mu}{\sigma}\right)\right) - \left(1 + \xi\left(\frac{x_t-\mu}{\sigma}\right)\right)^{-1/\xi}$$

where x_t is the observed maximum for the t^{th} period.

We consider again the example of the MSCI USA index. Using daily returns, we calculate the block maximum for each period of 22 trading days. We then estimate the GEV distribution using the method of maximum likelihood. For the period 1995-2015, we obtain $\hat{\mu} = 0.0149$, $\hat{\sigma} = 0.0062$ and $\hat{\xi} = 0.3736$. In Figure 12.11, we compared the estimated GEV distribution with the distribution function $\mathbf{F}_{22:22}(x)$ when we assume that daily returns are Gaussian. We notice that the Gaussian hypothesis largely underestimates extreme events as illustrated by the quantile function in the table below:

α	90%	95%	96%	97%	98%	99%
Gaussian	3.26%	3.56%	3.65%	3.76%	3.92%	4.17%
GEV	3.66%	4.84%	5.28%	5.91%	6.92%	9.03%

For instance, the probability is 1% to observe a maximum daily return during a period of one month larger than 4.17% in the case of the Gaussian distribution and 9.03% in the case of the GEV distribution.

12.2.3.2 Estimating the value-at-risk

Let us consider a portfolio w, whose mark-to-market value is $P_t(w)$ at time t. We recall that the P&L between t and $t+1$ is equal to:

$$\begin{aligned}\Pi(w) &= P_{t+1}(w) - P_t(w) \\ &= P_t(w) \cdot R(w)\end{aligned}$$

where $R(w)$ is the daily return of the portfolio. If we note $\hat{\mathbf{F}}$ the estimated probability distribution of $R(w)$, the expression of the value-at-risk at the confidence level α is equal to:

$$\text{VaR}_\alpha(w) = -P_t(w) \cdot \hat{\mathbf{F}}^{-1}(1-\alpha)$$

We now estimate the GEV distribution $\hat{\mathbf{G}}$ of the maximum of $-R(w)$ for a period of n trading days[10]. We have to define the confidence level α_{GEV} when we consider block minima of daily returns that corresponds to the same confidence level α when we consider daily returns. For that, we assume that the two exception events have the same return period, implying that:

$$\frac{1}{1-\alpha} \times 1 \text{ day} = \frac{1}{1-\alpha_{\text{GEV}}} \times n \text{ days}$$

[10] We model the maximum of the opposite of daily returns, because we are interested in extreme losses, and not in extreme profits.

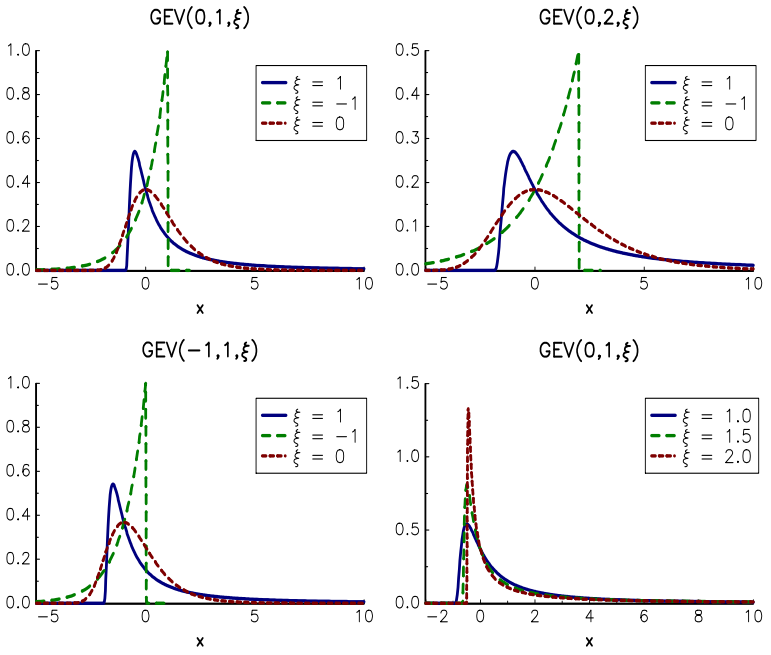


FIGURE 12.10: Probability density function of the GEV distribution

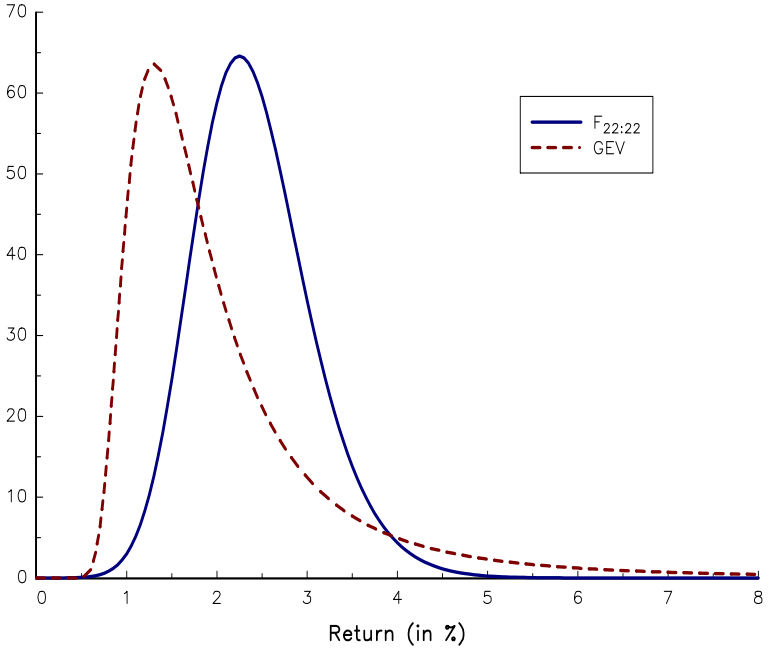


FIGURE 12.11: Probability density function of the maximum return $R_{22:22}$

We deduce that:

$$\alpha_{\text{GEV}} = 1 - (1 - \alpha) \cdot n$$

It follows that the value-at-risk calculated with the GEV distribution is equal to[11]:

$$\text{VaR}_\alpha(w) = P(t) \cdot \hat{\mathbf{G}}^{-1}(\alpha_{\text{GEV}})$$

We consider four portfolios invested in the MSCI USA index and the MSCI EM index: (1) long on the MSCI USA, (2) long on the MSCI EM index, (3) long on the MSCI USA and short on the MSCI EM index and (4) long on the MSCI EM index and short on the MSCI USA index. Using daily returns from January 1995 to December 2015, we estimate the daily value-at-risk of these portfolios for different confidence levels α. We report the results in Table 12.5 for Gaussian and historical value-at-risk measures and compare them with those calculated with the GEV approach. In this case, we estimate the parameters of the extreme value distribution using block maxima of 22 trading days. When we consider a 99% confidence level, the lowest value is obtained by the GEV method followed by Gaussian and historical methods. For a higher quantile, the GEV VaR is between the Gaussian VaR and the historical VaR. The value-at-risk calculated with the GEV approach can therefore be interpreted as a parametric value-at-risk, which is estimated using only tail events.

TABLE 12.5: Comparing Gaussian, historical and GEV value-at-risk measures

VaR	α	Long US	Long EM	Long US Short EM	Long EM Short US
Gaussian	99.0%	2.88%	2.83%	3.06%	3.03%
	99.5%	3.19%	3.14%	3.39%	3.36%
	99.9%	3.83%	3.77%	4.06%	4.03%
Historical	99.0%	3.46%	3.61%	3.37%	3.81%
	99.5%	4.66%	4.73%	3.99%	4.74%
	99.9%	7.74%	7.87%	6.45%	7.27%
GEV	99.0%	2.64%	2.61%	2.72%	2.93%
	99.5%	3.48%	3.46%	3.41%	3.82%
	99.9%	5.91%	6.05%	5.35%	6.60%

12.2.4 Peak over threshold

12.2.4.1 Definition

The estimation of the GEV distribution is a '*block component-wise*' approach. This means that from a sample of random variates, we build a sample of maxima by considering blocks with the same length. This implies a loss of information, because some blocks may contain several extreme events whereas some other blocks may not be impacted by extremes. Another approach consists in using the '*peak over threshold*' (POT) method. In this case, we are interested in estimating the distribution of exceedance over a certain threshold u:

$$\mathbf{F}_u(x) = \Pr\{X - u \leq x \mid X > u\}$$

[11] The inverse function of the probability distribution $\mathcal{GEV}(\mu, \sigma, \xi)$ is equal to:

$$\mathbf{G}^{-1}(\alpha) = \mu - \frac{\sigma}{\xi}\left(1 - (-\ln \alpha)^{-\xi}\right)$$

where $0 \leq x < x_0 - u$ and $x_0 = \sup\{x \in \mathbb{R} : \mathbf{F}(x) < 1\}$. $\mathbf{F}_u(x)$ is also called the conditional excess distribution function. It is also equal to:

$$\begin{aligned}\mathbf{F}_u(x) &= 1 - \Pr\{X - u \leq x \mid X \leq u\} \\ &= 1 - \left(\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)}\right) \\ &= \frac{\mathbf{F}(u+x) - \mathbf{F}(u)}{1 - \mathbf{F}(u)}\end{aligned}$$

Pickands (1975) showed that, for very large u, $\mathbf{F}_u(x)$ follows a generalized Pareto distribution (GPD): $\mathbf{F}_u(x) \approx \mathbf{H}(x)$ where[12]:

$$\mathbf{H}(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}$$

The distribution function $\mathcal{GPD}(\sigma, \xi)$ depends on two parameters: σ is the scale parameter and ξ is the shape parameter.

Example 129 *If* $\mathbf{F}$ *is an exponential distribution* $\mathcal{E}(\lambda)$*, we have:*

$$\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} = \exp(-\lambda x)$$

This is the generalized Pareto distribution when $\sigma = 1/\lambda$ *and* $\xi \to 0$*.*

Example 130 *If* $\mathbf{F}$ *is a uniform distribution, we have:*

$$\frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} = 1 - \frac{x}{1-u}$$

It corresponds to the generalized Pareto distribution with the following parameters: $\sigma = 1 - u$ *and* $\xi = -1$*.*

In fact, there is a strong link between the block maxima approach and the peak over threshold method. Suppose that $X_{n:n} \sim \mathcal{GEV}(\mu, \sigma, \xi)$. It follows that:

$$\mathbf{F}^n(x) \approx \exp\left\{-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right\}$$

We deduce that:

$$n \ln \mathbf{F}(x) \approx -\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}$$

Using the approximation $\ln \mathbf{F}(x) \approx -(1 - \mathbf{F}(x))$ for large x, we obtain:

$$1 - \mathbf{F}(x) \approx \frac{1}{n}\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}$$

We find that $\mathbf{F}_u(x)$ is a generalized Pareto distribution $\mathcal{GPD}(\tilde{\sigma}, \xi)$:

$$\begin{aligned}\Pr\{X > u + x \mid X > u\} &= \frac{1 - \mathbf{F}(u+x)}{1 - \mathbf{F}(u)} \\ &= \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}\end{aligned}$$

[12] If $\xi \to 0$, we have $\mathbf{H}(x) = 1 - \exp(-x/\sigma)$.

where:

$$\tilde{\sigma} = \sigma + \xi (u - \mu)$$

Therefore, we have a duality between GEV and GPD distribution functions:

> "*[...] if block maxima have approximating distribution* $\mathbf{G}$*, then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter* ξ *is equal to that of the corresponding GEV distribution. Choosing a different, but still large, block size* n *would affect the values of the GEV parameters, but not those of the corresponding generalized Pareto distribution of threshold excesses:* ξ *is invariant to block size, while the calculation of* $\tilde{\sigma}$ *is unperturbed by the changes in* μ *and* σ *which are self-compensating*" (Coles, 2001, page 75).

The estimation of the parameters (σ, ξ) is not obvious because it depends on the value taken by the threshold u. It must be sufficiently large to apply the previous theorem, but we also need enough data to obtain good estimates. We notice that the mean residual life $e(u)$ is a linear function of u:

$$\begin{aligned} e(u) &= \mathbb{E}[X - u \mid X > u] \\ &= \frac{\sigma + \xi u}{1 - \xi} \end{aligned}$$

when $\xi < 1$. If the GPD approximation is valid for a value u_0, it is therefore valid for any value $u > u_0$. To determine u_0, we can use a mean residual life plot, which consists in plotting u against the empirical mean excess $\hat{e}(u)$:

$$\hat{e}(u) = \frac{\sum_{i=1}^{n} (x_i - u)^+}{\sum_{i=1}^{n} \mathbb{1}\{x_i > u\}}$$

Once u_0 is found, we estimate the parameters (σ, ξ) by the method of maximum likelihood or the linear regression[13].

Let us consider our previous example. In Figure 12.12, we have reported the mean residual life plot for the left tail of the four portfolios[14]. The determination of u_0 consists in finding linear relationships. We have a first linear relationship between $u = -3\%$ and $u = -1\%$, but it is not valid because it is followed by a change in slope. We prefer to consider that the linear relationship is valid for $u \geq 2\%$. By assuming that $u_0 = 2\%$ for all the four portfolios, we obtain the estimates given in Table 12.6.

12.2.4.2 Estimating the expected shortfall

We recall that:

$$\mathbf{F}_u(x) = \frac{\mathbf{F}(u + x) - \mathbf{F}(u)}{1 - \mathbf{F}(u)} \approx \mathbf{H}(x)$$

where $\mathbf{H} \sim \mathcal{GPD}(\sigma, \xi)$. We deduce that:

$$\begin{aligned} \mathbf{F}(x) &= \mathbf{F}(u) + (1 - \mathbf{F}(u)) \cdot \mathbf{F}_u(x - u) \\ &\approx \mathbf{F}(u) + (1 - \mathbf{F}(u)) \cdot \mathbf{H}(x - u) \end{aligned}$$

[13] In this case, we estimate the linear model $\hat{e}(u) = a + b \cdot u + \varepsilon$ for $u \geq u_0$ and deduce that $\hat{\sigma} = \hat{a} / \left(1 + \hat{b}\right)$ and $\hat{\xi} = \hat{b} / \left(1 + \hat{b}\right)$.

[14] This means that $\hat{e}(u)$ is calculated using the portfolio loss, that is the opposite of the portfolio return.

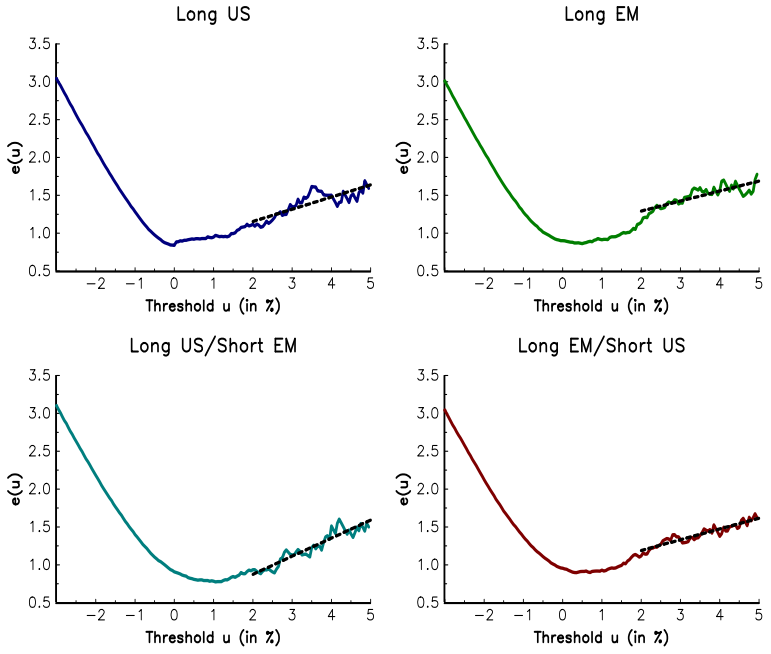


FIGURE 12.12: Mean residual life plot

TABLE 12.6: Estimation of the generalized Pareto distribution

Parameter	Long US	Long EM	Long US Short EM	Long EM Short US
$\hat{a}$	0.834	1.029	0.394	0.904
$\hat{b}$	0.160	0.132	0.239	0.142
$\hat{\sigma}$	0.719	0.909	0.318	0.792
$\hat{\xi}$	0.138	0.117	0.193	0.124

We consider a sample of size n. We note n_u the number of observations whose value x_i is larger than the threshold u. The non-parametric estimate of $\mathbf{F}(u)$ is then equal to:

$$\hat{\mathbf{F}}(u) = 1 - \frac{n_u}{n}$$

Therefore, we obtain the following semi-parametric estimate of $\mathbf{F}(x)$ for x larger than u:

$$\begin{aligned}
\hat{\mathbf{F}}(x) &= \hat{\mathbf{F}}(u) + \left(1 - \hat{\mathbf{F}}(u)\right) \cdot \hat{\mathbf{H}}(x-u) \\
&= \left(1 - \frac{n_u}{n}\right) + \frac{n_u}{n}\left(1 - \left(1 + \frac{\hat{\xi}(x-u)}{\hat{\sigma}}\right)^{-1/\hat{\xi}}\right) \\
&= 1 - \frac{n_u}{n}\left(1 + \frac{\hat{\xi}(x-u)}{\hat{\sigma}}\right)^{-1/\hat{\xi}}
\end{aligned}$$

We can interpret $\hat{\mathbf{F}}(x)$ as the historical estimate of the probability distribution tail that is improved by the extreme value theory. We deduce that:

$$\begin{aligned}\text{VaR}_\alpha &= \hat{\mathbf{F}}^{-1}(\alpha) \\ &= u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{n}{n_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right)\end{aligned}$$

and:

$$\begin{aligned}\text{ES}_\alpha &= \mathbb{E}[X \mid X > \text{VaR}_\alpha] \\ &= \text{VaR}_\alpha + \mathbb{E}[X - \text{VaR}_\alpha \mid X > \text{VaR}_\alpha] \\ &= \text{VaR}_\alpha + \frac{\hat{\sigma} + \hat{\xi}(\text{VaR}_\alpha - u)}{1-\hat{\xi}} \\ &= \frac{\text{VaR}_\alpha}{1-\hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi}u}{1-\hat{\xi}} \\ &= u - \frac{\hat{\sigma}}{\hat{\xi}} + \frac{\hat{\sigma}}{\left(1-\hat{\xi}\right)\hat{\xi}}\left(\frac{n}{n_u}(1-\alpha)\right)^{-\hat{\xi}}\end{aligned}$$

We consider again the example of the four portfolios with exposures on US and EM equities. In the sample, we have 3 815 observations, whereas the value taken by n_u when u is equal to 2% is 171, 161, 174 and 195 respectively. Using the estimates given in Table 12.6, we calculate the daily value-at-risk and expected shortfall of the four portfolios. The results are reported in Table 12.7. If we compare them with those obtained in Table 12.5 on page 773, we notice that the GPD VaR is close to the GEV VaR.

TABLE 12.7: Estimating value-at-risk and expected shortfall risk measures using the generalized Pareto distribution

Risk measure	α	Long US	Long EM	Long US Short EM	Long EM Short US
VaR	99.0%	3.20%	3.42%	2.56%	3.43%
	99.5%	3.84%	4.20%	2.88%	4.13%
	99.9%	5.60%	6.26%	3.80%	6.02%
ES	99.0%	4.22%	4.64%	3.09%	4.54%
	99.5%	4.97%	5.52%	3.48%	5.34%
	99.9%	7.01%	7.86%	4.62%	7.49%

12.3 Multivariate extreme value theory

The extreme value theory is generally formulated and used in the univariate case. It can be easily extended to the multivariate case, but its implementation is more difficult. This section is essentially based on the works of Deheuvels (1978), Galambos (1987) and Joe (1997).

12.3.1 Multivariate extreme value distributions

12.3.1.1 Extreme value copulas

An extreme value (EV) copula satisfies the following relationship:

$$\mathbf{C}\left(u_1^t,\ldots,u_n^t\right)=\mathbf{C}^t\left(u_1,\ldots,u_n\right)$$

for all $t>0$. For instance, the Gumbel copula is an EV copula:

$$\begin{aligned}\mathbf{C}\left(u_1^t,u_2^t\right)&=\exp\left(-\left(\left(-\ln u_1^t\right)^\theta+\left(-\ln u_2^t\right)^\theta\right)^{1/\theta}\right)\\&=\exp\left(-\left(t^\theta\left(\left(-\ln u_1\right)^\theta+\left(-\ln u_2\right)^\theta\right)\right)^{1/\theta}\right)\\&=\left(\exp\left(-\left(\left(-\ln u_1\right)^\theta+\left(-\ln u_2\right)^\theta\right)^{1/\theta}\right)\right)^t\\&=\mathbf{C}^t\left(u_1,u_2\right)\end{aligned}$$

but it is not the case of the Farlie-Gumbel-Morgenstern copula:

$$\begin{aligned}\mathbf{C}\left(u_1^t,u_2^t\right)&=u_1^tu_2^t+\theta u_1^tu_2^t\left(1-u_1^t\right)\left(1-u_2^t\right)\\&=u_1^tu_2^t\left(1+\theta-\theta u_1^t-\theta u_2^t+\theta u_1^tu_2^t\right)\\&\neq u_1^tu_2^t\left(1+\theta-\theta u_1-\theta u_2+\theta u_1u_2\right)^t\\&\neq\mathbf{C}^t\left(u_1,u_2\right)\end{aligned}$$

The term '*extreme value copula*' suggests a relationship between the extreme value theory and these copula functions. Let $X=(X_1,\ldots,X_n)$ be a random vector of dimension n. We note $X_{m:m}$ the random vector of maxima:

$$X_{m:m}=\begin{pmatrix}X_{m:m,1}\\\vdots\\X_{m:m,n}\end{pmatrix}$$

and $\mathbf{F}_{m:m}$ the corresponding distribution function:

$$\mathbf{F}_{m:m}\left(x_1,\ldots,x_n\right)=\Pr\left\{X_{m:m,1}\leq x_1,\ldots,X_{m:m,n}\leq x_n\right\}$$

The multivariate extreme value (MEV) theory considers the asymptotic behavior of the non-degenerate distribution function $\mathbf{G}$ such that:

$$\lim_{m\to\infty}\Pr\left(\frac{X_{m:m,1}-b_{m,1}}{a_{m,1}}\leq x_1,\ldots,\frac{X_{m:m,n}-b_{m,n}}{a_{m,n}}\leq x_n\right)=\mathbf{G}\left(x_1,\ldots,x_n\right)$$

Using Sklar's theorem, there exists a copula function $\mathbf{C}\left\langle\mathbf{G}\right\rangle$ such that:

$$\mathbf{G}\left(x_1,\ldots,x_n\right)=\mathbf{C}\left\langle\mathbf{G}\right\rangle\left(\mathbf{G}_1\left(x_1\right),\ldots,\mathbf{G}_n\left(x_n\right)\right)$$

It is obvious that the marginals $\mathbf{G}_1,\ldots,\mathbf{G}_n$ satisfy the Fisher-Tippet theorem, meaning that the marginals of a multivariate extreme value distribution can only be Gumbel, Fréchet or Weibull distribution functions. For the copula $\mathbf{C}\left\langle\mathbf{G}\right\rangle$, we have the following result: $\mathbf{C}\left\langle\mathbf{G}\right\rangle$ is an extreme value copula.

With the copula representation, we can then easily define MEV distributions. For instance, if we consider the random vector (X_1, X_2), whose joint distribution function is:

$$\mathbf{F}(x_1, x_2) = \exp\left(-\left((-\ln \Phi(x_1))^{\theta} + (-\ln x_2)^{\theta}\right)^{1/\theta}\right)$$

we notice that X_1 is a Gaussian random variable and X_2 is a uniform random variable. We conclude that the corresponding limit distribution function of maxima is:

$$\mathbf{G}(x_1, x_2) = \exp\left(-\left((-\ln \mathbf{\Lambda}(x_1))^{\theta} + (-\ln \mathbf{\Psi}_1(x_2))^{\theta}\right)^{1/\theta}\right)$$

In Figure 12.13, we have reported the contour plot of four MEV distribution functions, whose marginals are $\mathcal{GEV}(0, 1, 1)$ and $\mathcal{GEV}(0, 1, 1.5)$. For the dependence function, we consider the Gumbel-Hougaard copula and calibrate the parameter θ with respect to the Kendall's tau.

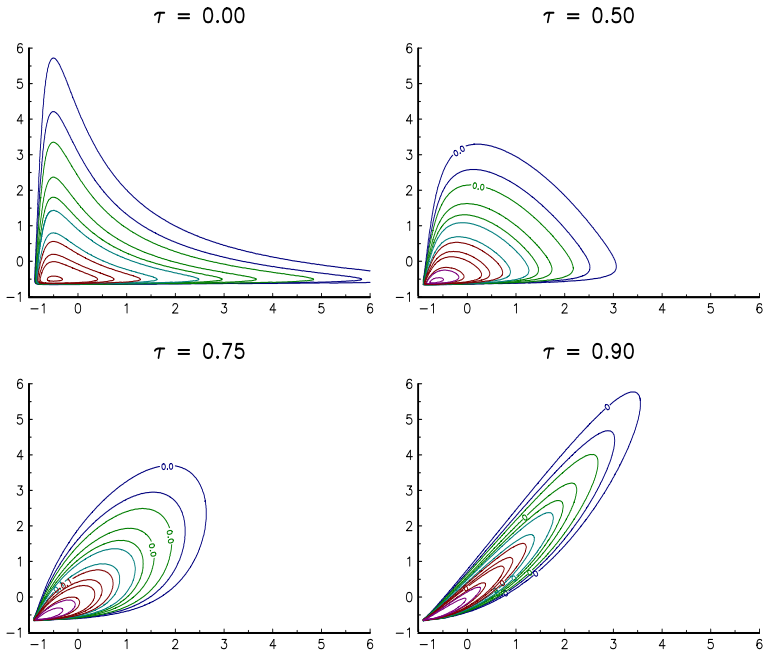


FIGURE 12.13: Multivariate extreme value distributions

12.3.1.2 Deheuvels-Pickands representation

Let $\mathbf{D}$ be a multivariate distribution function, whose survival marginals are exponential and the dependence structure is an extreme value copula. By using the relationship[15] $\mathbf{C}(u_1, \ldots, u_n) = \mathbf{C}\left(e^{-\tilde{u}_1}, \ldots, e^{-\tilde{u}_n}\right) = \mathbf{D}(\tilde{u}_1, \ldots, \tilde{u}_n)$, we have $\mathbf{D}^t(\tilde{\mathbf{u}}) = \mathbf{D}(t\tilde{\mathbf{u}})$. Therefore, $\mathbf{D}$ is a *min-stable multivariate exponential* (MSMVE) distribution.

We now introduce the Deheuvels/Pickands MSMVE representation. Let $\mathbf{D}(\tilde{\mathbf{u}})$ be a survival function with exponential marginals. $\mathbf{D}$ satisfies the relationship:

$$-\ln \mathbf{D}(t \cdot \tilde{\mathbf{u}}) = -t \cdot \ln \mathbf{D}(\tilde{\mathbf{u}}) \qquad \forall t > 0$$

[15] We recall that $\tilde{u} = -\ln u$.

if and only if the representation of $\mathbf{D}$ is:

$$-\ln \mathbf{D}(\tilde{\mathbf{u}}) = \int \cdots \int_{\mathcal{S}_n} \max_{1\leq i\leq n} (q_i \tilde{u}_i)\, \mathrm{d}S(\mathbf{q}) \qquad \forall\, \tilde{\mathbf{u}} \geq \mathbf{0}$$

where $\mathcal{S}_n$ is the n-dimensional unit simplex and S is a finite measure on $\mathcal{S}_n$. This is the formulation[16] given by Joe (1997). Sometimes, the Deheuvels/Pickands representation is presented using a dependence function $B(\mathbf{w})$ defined by:

$$\begin{aligned} \mathbf{D}(\tilde{\mathbf{u}}) &= \exp\left(-\left(\sum_{i=1}^{n} \tilde{u}_i\right) B(w_1,\ldots,w_n)\right) \\ B(\mathbf{w}) &= \int \cdots \int_{\mathcal{S}_n} \max_{1\leq i\leq n} (q_i w_i)\, \mathrm{d}S(\mathbf{q}) \end{aligned}$$

where $w_i = \left(\sum_{i=1}^{n} \tilde{u}_i\right)^{-1} \tilde{u}_i$. Tawn (1990) showed that B is a convex function and satisfies the following condition:

$$\max(w_1,\ldots,w_n) \leq B(w_1,\ldots,w_n) \leq 1 \tag{12.5}$$

We deduce that an extreme value copula satisfies the PQD property:

$$\mathbf{C}^{\perp} \prec \mathbf{C} \prec \mathbf{C}^{+}$$

In the bivariate case, the formulation can be simplified because the convexity of B and the condition (12.5) are sufficient (Tawn, 1988). We have:

$$\begin{aligned} \mathbf{C}(u_1,u_2) &= \mathbf{D}(\tilde{u}_1,\tilde{u}_2) \\ &= \exp\left(-(\tilde{u}_1+\tilde{u}_2)\, B\left(\frac{\tilde{u}_1}{\tilde{u}_1+\tilde{u}_2}, \frac{\tilde{u}_2}{\tilde{u}_1+\tilde{u}_2}\right)\right) \\ &= \exp\left(\ln(u_1u_2)\, B\left(\frac{\ln u_1}{\ln(u_1u_2)}, \frac{\ln u_2}{\ln(u_1u_2)}\right)\right) \\ &= \exp\left(\ln(u_1u_2)\, A\left(\frac{\ln u_1}{\ln(u_1u_2)}\right)\right) \end{aligned}$$

where $A(w) = B(w, 1-w)$. A is a convex function where $A(0) = A(1) = 1$ and satisfies $\max(w, 1-w) \leq A(w) \leq 1$.

Example 131 *For the Gumbel copula, we have:*

$$\begin{aligned} -\ln \mathbf{D}(\tilde{u}_1,\tilde{u}_2) &= \left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta} \\ B(w_1,w_2) &= \frac{\left(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta}\right)^{1/\theta}}{(\tilde{u}_1+\tilde{u}_2)} = \left(w_1^{\theta} + w_2^{\theta}\right)^{1/\theta} \\ A(w) &= \left(w^{\theta} + (1-w)^{\theta}\right)^{1/\theta} \end{aligned}$$

[16] Note that it is similar to Proposition 5.11 of Resnick (1987), although the author does not use copulas.

We verify that a bivariate EV copula satisfies the PQD property:

$$\begin{aligned}
& \max(w, 1-w) \leq A(w) \leq 1 \\
\Leftrightarrow \quad & \max\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right) \leq A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \leq 1 \\
\Leftrightarrow \quad & \min(\ln u_1, \ln u_2) \geq \ln(u_1 u_2) \cdot A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \geq \ln(u_1 u_2) \\
\Leftrightarrow \quad & \min(u_1, u_2) \geq \exp\left(\ln(u_1 u_2) \cdot A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right) \geq u_1 u_2 \\
\Leftrightarrow \quad & \mathbf{C}^+ \succ \mathbf{C} \succ \mathbf{C}^\perp
\end{aligned}$$

When the extreme values are independent, we have $A(w) = 1$ whereas the case of perfect dependence corresponds to $A(w) = \max(w, 1-w)$:

$$\begin{aligned}
\mathbf{C}(u_1, u_2) &= \exp\left(\ln(u_1 u_2) \cdot \max\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right)\right) \\
&= \min(u_1, u_2) \\
&= \mathbf{C}^+(u_1, u_2)
\end{aligned}$$

In Table 12.8, we have reported the dependence function $A(w)$ of the most used EV copula functions.

TABLE 12.8: List of extreme value copulas

Copula	θ	$\mathbf{C}(u_1, u_2)$	$A(w)$
$\mathbf{C}^\perp$		$u_1 u_2$	1
Gumbel	$[1, \infty)$	$\exp\left(-\left(\tilde{u}_1^\theta + \tilde{u}_2^\theta\right)^{1/\theta}\right)$	$\left(w^\theta + (1-w)^\theta\right)^{1/\theta}$
Gumbel II	$[0, 1]$	$u_1 u_2 \exp\left(\theta \frac{\tilde{u}_1 \tilde{u}_2}{\tilde{u}_1 + \tilde{u}_2}\right)$	$\theta w^2 - \theta w + 1$
Galambos	$[0, \infty)$	$u_1 u_2 \exp\left(\left(\tilde{u}_1^{-\theta} + \tilde{u}_2^{-\theta}\right)^{-1/\theta}\right)$	$1 - \left(w^{-\theta} + (1-w)^{-\theta}\right)^{-1/\theta}$
Hüsler-Reiss	$[0, \infty)$	$\exp\left(-\tilde{u}_1 \vartheta(u_1, u_2; \theta) - \tilde{u}_2 \vartheta(u_2, u_1; \theta)\right)$	$w\kappa(w; \theta) + (1-w)\kappa(1-w; \theta)$
Marshall-Olkin	$[0, 1]^2$	$u_1^{1-\theta_1} u_2^{1-\theta_2} \min\left(u_1^{\theta_1}, u_2^{\theta_2}\right)$	$\max(1 - \theta_1 w, 1 - \theta_2(1-w))$
$\mathbf{C}^+$		$\min(u_1, u_2)$	$\max(w, 1-w)$

$\vartheta(u_1, u_2; \theta) = \Phi\left(\frac{1}{\theta} + \frac{\theta}{2} \ln(\ln u_1 / \ln u_2)\right)$
$\kappa(w; \theta) = \vartheta(w, 1-w; \theta)$

Source: Ghoudi *et al.* (1998).

12.3.2 Maximum domain of attraction

Let $\mathbf{F}$ be a multivariate distribution function whose marginals are $\mathbf{F}_1, \ldots, \mathbf{F}_n$ and the copula is $\mathbf{C}\langle\mathbf{F}\rangle$. We note $\mathbf{G}$ the corresponding multivariate extreme value distribution, $\mathbf{G}_1, \ldots, \mathbf{G}_n$ the marginals of $\mathbf{G}$ and $\mathbf{C}\langle\mathbf{G}\rangle$ the associated copula function. We can show that $\mathbf{F} \in \text{MDA}(\mathbf{G})$ if and only if $\mathbf{F}_i \in \text{MDA}(\mathbf{G}_i)$ for all $i = 1, \ldots, n$ and $\mathbf{C}\langle\mathbf{F}\rangle \in \text{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$. Previously, we have seen how to characterize the max-domain of attraction in the univariate case and how to calculate the normalizing constants. These constants remains the same in the multivariate case, meaning that the only difficulty is to determine the EV copula $\mathbf{C}\langle\mathbf{G}\rangle$.

We can show that $\mathbf{C}\langle\mathbf{F}\rangle \in \text{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$ if $\mathbf{C}\langle\mathbf{F}\rangle$ satisfies the following relationship:

$$\lim_{t\to\infty} \mathbf{C}^t\langle\mathbf{F}\rangle\left(u_1^{1/t},\ldots,u_n^{1/t}\right) = \mathbf{C}\langle\mathbf{G}\rangle(u_1,\ldots,u_n)$$

Moreover, if $\mathbf{C}\langle\mathbf{F}\rangle$ is an EV copula, then $\mathbf{C}\langle\mathbf{F}\rangle \in \text{MDA}(\mathbf{C}\langle\mathbf{F}\rangle)$. This important result is equivalent to:

$$\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{w_1},\ldots,(1-u)^{w_n}\right)}{u} = B(w_1,\ldots,w_n)$$

In the bivariate case, we obtain:

$$\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{1-t},(1-u)^{t}\right)}{u} = A(t)$$

for all $t \in [0,1]$.

Example 132 *We consider the random vector (X_1, X_2) defined by the following distribution function:*

$$\mathbf{F}(x_1,x_2) = \left(\left(1-e^{-x_1}\right)^{-\theta} + x_2^{-\theta} - 1\right)^{-1/\theta}$$

on $[0,\infty]\times[0,1]$. The marginals of $\mathbf{F}(x_1,x_2)$ are $\mathbf{F}_1(x_1) = \mathbf{F}(x_1,1) = 1-e^{-x_1}$ and $\mathbf{F}_2(x_2) = \mathbf{F}(\infty,x_2) = x_2$. It follows that X_1 is an exponential random variable and X_2 is a uniform random variable. We know that:

$$\lim_{n\to\infty} \Pr\left(\frac{X_{n:n,1}-\ln n}{1} \le x_1\right) = \mathbf{\Lambda}(x_1)$$

and:

$$\lim_{n\to\infty} \Pr\left(\frac{X_{n:n,2}-1}{n^{-1}} \le x_2\right) = \mathbf{\Psi}_1(x_2)$$

Since the dependence function of $\mathbf{F}$ is the Clayton copula: $\mathbf{C}\langle\mathbf{F}\rangle(u_1,u_2) = \left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta}$, we have:

$$\begin{aligned}
\lim_{u\to 0} \frac{1-\mathbf{C}\langle\mathbf{F}\rangle\left((1-u)^{t},(1-u)^{1-t}\right)}{u} &= \lim_{u\to 0} \frac{1-(1+\theta u+o(u))^{-1/\theta}}{u} \\
&= \lim_{u\to 0} \frac{u+o(u)}{u} \\
&= 1
\end{aligned}$$

We deduce that $\mathbf{C}\langle\mathbf{G}\rangle = \mathbf{C}^{\perp}$. Finally, we obtain:

$$\begin{aligned}
\mathbf{G}(x_1,x_2) &= \lim_{n\to\infty} \Pr\{X_{n:n,1}-\ln n \le x_1, n(X_{n:n,2}-1)\le x_2\} \\
&= \mathbf{\Lambda}(x_1)\cdot\mathbf{\Psi}_1(x_2) \\
&= \exp\left(-e^{-x_1}\right)\cdot\exp(x_2)
\end{aligned}$$

If we change the copula $\mathbf{C}\langle\mathbf{F}\rangle$, only the copula $\mathbf{C}\langle\mathbf{G}\rangle$ is modified. For instance, when $\mathbf{C}\langle\mathbf{F}\rangle$ is the Normal copula with parameter $\rho<1$, then $\mathbf{G}(x_1,x_2) = \exp(-e^{-x_1})\cdot\exp(x_2)$. When the copula parameter ρ is equal to 1, we obtain $\mathbf{G}(x_1,x_2) = \min(\exp(-e^{-x_1}),\exp(x_2))$. When $\mathbf{C}\langle\mathbf{F}\rangle$ is the Gumbel copula, the MEV distribution becomes $\mathbf{G}(x_1,x_2) = \exp\left(-\left(e^{-\theta x_1}+(-x_2)^{\theta}\right)^{1/\theta}\right)$.

12.3.3 Tail dependence of extreme values

We can show that the (upper) tail dependence of $\mathbf{C}\langle\mathbf{G}\rangle$ is equal to the (upper) tail dependence of $\mathbf{C}\langle\mathbf{F}\rangle$:

$$\lambda^+\left(\mathbf{C}\langle\mathbf{G}\rangle\right) = \lambda^+\left(\mathbf{C}\langle\mathbf{F}\rangle\right)$$

This implies that extreme values are independent if the copula function $\mathbf{C}\langle\mathbf{F}\rangle$ has no (upper) tail dependence.

12.4 Exercises

12.4.1 Uniform order statistics

We assume that $X_1, \ldots, X_n$ are independent uniform random variables.

1. Show that the density function of the order statistic $X_{i:n}$ is:

$$f_{i:n}(x) = \frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)} x^{i-1}(1-x)^{n-i}$$

2. Calculate the mean $\mathbb{E}[X_{i:n}]$.

3. Show that the variance is equal to:

$$\text{var}(X_{i:n}) = \frac{i(n-i+1)}{(n+1)^2(n+2)}$$

4. We consider 10 samples of 8 independent observations from the uniform probability distribution $\mathcal{U}_{[0,1]}$:

Sample	Observation							
	1	2	3	4	5	6	7	8
1	0.24	0.45	0.72	0.14	0.04	0.34	0.94	0.55
2	0.12	0.32	0.69	0.64	0.31	0.25	0.97	0.57
3	0.69	0.50	0.26	0.17	0.50	0.85	0.11	0.17
4	0.53	0.00	0.77	0.58	0.98	0.15	0.98	0.03
5	0.89	0.25	0.15	0.62	0.74	0.85	0.65	0.46
6	0.74	0.65	0.86	0.05	0.93	0.15	0.25	0.07
7	0.16	0.12	0.63	0.33	0.55	0.61	0.34	0.95
8	0.96	0.82	0.01	0.87	0.57	0.11	0.14	0.47
9	0.68	0.83	0.73	0.78	0.27	0.85	0.55	0.57
10	0.89	0.94	0.91	0.28	0.99	0.40	0.99	0.68

   For each sample, find the order statistics. Calculate the empirical mean and standard deviation of $X_{i:8}$ for $i = 1, \ldots, 8$ and compare these values with the theoretical results.

5. We assume that n is odd, meaning that $n = 2k+1$. We consider the median statistic $X_{k+1:n}$. Show that the density function of $X_{i:n}$ is right asymmetric if $i \leq k$, symmetric about .5 if $i = k+1$ and left asymmetric otherwise.

6. We now assume that the density function of $X_1, \ldots, X_n$ is symmetric. How are impacted the results obtained in Question 5?

12.4.2 Order statistics and return period

1. Let X and $\mathbf{F}$ be the daily return of a portfolio and the associated probability distribution. We note $X_{n:n}$ the maximum of daily returns for a period of n trading days. Using the standard assumptions, define the cumulative distribution function $\mathbf{F}_{n:n}$ of $X_{n:n}$ if we suppose that $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

2. How could we test the hypothesis $\mathcal{H}_0 : X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ using $\mathbf{F}_{n:n}$?

3. Define the notion of return period. What is the return period associated to the statistics $\mathbf{F}^{-1}(99\%)$, $\mathbf{F}_{1:1}^{-1}(99\%)$, $\mathbf{F}_{5:5}^{-1}(99\%)$ and $\mathbf{F}_{21:21}^{-1}(99\%)$?

4. We consider the random variable $X_{20:20}$. Find the confidence level α which ensures that the return period associated to the quantile $\mathbf{F}_{20:20}^{-1}(\alpha)$ is equivalent to the return period of the daily value-at-risk with a 99.9% confidence level.

12.4.3 Extreme order statistics of exponential random variables

1. We note $\boldsymbol{\tau} \sim \mathcal{E}(\lambda)$. Show that:

$$\Pr\{\boldsymbol{\tau} > t \mid \boldsymbol{\tau} > s\} = \Pr\{\boldsymbol{\tau} > t - s\}$$

   where $t > s$. Comment on this result.

2. Let $\boldsymbol{\tau}_i$ be the random variable of distribution $\mathcal{E}(\lambda_i)$. Calculate the probability distribution of $\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ and $\max(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ in the independent case. Show that:

$$\Pr\{\min(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n) = \tau_i\} = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$$

3. Same question if the random variables $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ are comonotone.

12.4.4 Extreme value theory in the bivariate case

1. What is an extreme value (EV) copula $\mathbf{C}$?

2. Show that $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$ are EV copulas. Why $\mathbf{C}^{-}$ cannot be an EV copula?

3. We define the Gumbel-Hougaard copula as follows:

$$\mathbf{C}(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right]^{1/\theta}\right)$$

   with $\theta \geq 1$. Verify that it is an EV copula.

4. What is the definition of the upper tail dependence λ? What is its usefulness in multivariate extreme value theory?

5. Let $f(x)$ and $g(x)$ be two functions such that $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0$. If $g'(x_0) \neq 0$, L'Hospital's rule states that:

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

   Deduce that the upper tail dependence λ of the Gumbel-Hougaard copula is $2 - 2^{1/\theta}$. What is the correlation of two extremes when $\theta = 1$?

6. We define the Marshall-Olkin copula as follows:

$$\mathbf{C}(u_1, u_2) = u_1^{1-\theta_1} \cdot u_2^{1-\theta_2} \cdot \min\left(u_1^{\theta_1}, u_2^{\theta_2}\right)$$

where $(\theta_1, \theta_2) \in [0,1]^2$.

   (a) Verify that it is an EV copula.

   (b) Find the upper tail dependence λ of the Marshall-Olkin copula.

   (c) What is the correlation of two extremes when $\min(\theta_1, \theta_2) = 0$?

   (d) In which case are two extremes perfectly correlated?

12.4.5 Maximum domain of attraction in the bivariate case

1. We consider the following probability distributions:

Distribution		$\mathbf{F}(x)$
Exponential	$\mathcal{E}(\lambda)$	$1 - e^{-\lambda x}$
Uniform	$\mathcal{U}_{[0,1]}$	x
Pareto	$\mathcal{P}(\alpha, \theta)$	$1 - \left(\frac{\theta+x}{\theta}\right)^{-\alpha}$

For each distribution, we give the normalization parameters a_n and b_n of the Fisher-Tippet theorem and the corresponding limit probability distribution $\mathbf{G}(x)$:

Distribution	a_n	b_n	$\mathbf{G}(x)$
Exponential	λ^{-1}	$\lambda^{-1} \ln n$	$\mathbf{\Lambda}(x) = e^{-e^{-x}}$
Uniform	n^{-1}	$1 - n^{-1}$	$\mathbf{\Psi}_1(x-1) = e^{x-1}$
Pareto	$\theta\alpha^{-1}n^{1/\alpha}$	$\theta n^{1/\alpha} - \theta$	$\mathbf{\Phi}_\alpha\left(1 + \frac{x}{\alpha}\right) = e^{-\left(1+\frac{x}{\alpha}\right)^{-\alpha}}$

We note $\mathbf{G}(x_1, x_2)$ the asymptotic distribution of the bivariate random vector $(X_{1,n:n}, X_{2,n:n})$ where $X_{1,i}$ (resp. $X_{2,i}$) are *iid* random variables.

   (a) What is the expression of $\mathbf{G}(x_1, x_2)$ when $X_{1,i}$ and $X_{2,i}$ are independent, $X_{1,i} \sim \mathcal{E}(\lambda)$ and $X_{2,i} \sim \mathcal{U}_{[0,1]}$?

   (b) Same question when $X_{1,i} \sim \mathcal{E}(\lambda)$ and $X_{2,i} \sim \mathcal{P}(\theta, \alpha)$.

   (c) Same question when $X_{1,i} \sim \mathcal{U}_{[0,1]}$ and $X_{2,i} \sim \mathcal{P}(\theta, \alpha)$.

2. What happen to the previous results when the dependence function between $X_{1,i}$ and $X_{2,i}$ is the Normal copula with parameter $\rho < 1$?

3. Same question when the parameter of the Normal copula is equal to one.

4. Find the expression of $\mathbf{G}(x_1, x_2)$ when the dependence function is the Gumbel-Hougaard copula.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 13

Monte Carlo Simulation Methods

Monte Carlo methods consist of solving mathematical problems using random numbers. The term '*Monte Carlo*' was apparently coined by physicists Ulam and von Neumann at Los Alamos in 1940 and refers to gambling casinos in Monaco[1]. Until the end of the eighties, Monte Carlo methods were principally used to calculate numerical integration[2] including mathematical expectations. More recently, the Monte Carlo method designates all numerical methods that involves stochastic simulation and consider random experiments on a computer.

This chapter is divided into three sections. In the first section, we present the different approaches to generate random numbers. Section two extends simulation methods when we manipulate stochastic processes. Finally, the last section is dedicated to Monte Carlo and quasi-Monte Carlo methods.

13.1 Random variate generation

Any Monte Carlo method is based on series of random variates that are independent and identically distributed (*iid*) according to a given probability distribution $\mathbf{F}$. As we will see later, it can be done by generating uniform random numbers. This is why numerical programming softwares already contain uniform random number generators. However, true randomness is impossible to simulate with a computer. In practice, only sequences of '*pseudorandom*' numbers can be produced with statistical properties that are close from those obtained with *iid* random variables.

13.1.1 Generating uniform random numbers

A first idea is to build a pseudorandom sequence $\mathcal{S}$ and repeat this sequence as often as necessary. For instance, for simulating 10 uniform random numbers, we can set $\mathcal{S} = \{0, 0.5, 1\}$ and repeat this sequence four times. In this case, the 10 random numbers are:

$$\{0, 0.5, 1, 0, 0.5, 1, 0, 0.5, 1, 0\}$$

We notice that the period length of this sequence is three. The quality of the pseudorandom number generator depends on the period length, which should be large in order to avoid duplication and serial correlation. If we calculate the second moment of $\mathcal{S}$, we do not obtain the variance of a uniform random variable $\mathcal{U}_{[0,1]}$. A good pseudorandom number generator should therefore pass standard adequacy tests.

[1] Monte Carlo is one of the four quarters of Monaco and houses the famous casino.
[2] In this case, we speak about Monte Carlo integration methods.

The most famous and used algorithm is the linear congruential generator (LCG):

$$\begin{aligned} x_n &= (a \cdot x_{n-1} + c) \bmod m \\ u_n &= x_n / m \end{aligned}$$

where a is the multiplicative constant, c is the additive constant and m is the modulus (or the order of the congruence). To initialize the algorithm, we have to define the initial number x_0, called the seed[3]. $\{x_1, x_2, \ldots, x_n\}$ is a sequence of pseudorandom integer numbers $(0 \leq x_n < m)$ whereas $\{u_1, u_2, \ldots, u_n\}$ is a sequence of uniform random variates. We can show that the maximum period[4] is m and can be only achieved for some specific values of a, c and m. The quality of the random number generator will then depend on the values of the parameters.

Example 133 *If we consider that* $a = 3$, $c = 0$, $m = 11$ *and* $x_0 = 1$, *we obtain the following sequence:*

$$\{1, 3, 9, 5, 4, 1, 3, 9, 5, 4, 1, 3, 9, 5, 4, \ldots\}$$

The period length is only five, meaning that only five uniform random variates can be generated: 0.09091, 0.27273, 0.81818, 0.45455 *and* 0.36364.

The minimal standard LCG proposed by Lewis *et al.* (1969) is defined by $a = 7^5$, $c = 0$ and $m = 2^{31} - 1$. In Table 13.1, we report two sequences generated with the seed values 1 and 123 456. This generator is widely used in numerical programming languages. However, its period length is equal to $m - 1 = 2^{31} - 2 \approx 2.15 \times 10^9$, which can be judged as insufficient for some modern Monte Carlo applications. For instance, if we consider the LDA model in operational risk with a Poisson distribution $\mathcal{P}(1\,000)$, we need approximately 10^{10} random numbers for drawing the severity loss if the number of Monte Carlo simulations is set to ten million. Another drawback is that LCG methods may exhibit lattice structures. For instance, Figure 13.1 shows the dependogram between u_{n-1} and u_n when $a = 10$, $c = 0$ and $m = 2^{31} - 1$.

TABLE 13.1: Simulation of 10 uniform pseudorandom numbers

n	x_n	u_n	x_n	u_n
0	1	0.000000	123 456	0.000057
1	16 807	0.000008	2 074 924 992	0.966212
2	282 475 249	0.131538	277 396 911	0.129173
3	1 622 650 073	0.755605	22 885 540	0.010657
4	984 943 658	0.458650	237 697 967	0.110687
5	1 144 108 930	0.532767	670 147 949	0.312062
6	470 211 272	0.218959	1 772 333 975	0.825307
7	101 027 544	0.047045	2 018 933 935	0.940139
8	1 457 850 878	0.678865	1 981 022 945	0.922486
9	1 458 777 923	0.679296	466 173 527	0.217079
10	2 007 237 709	0.934693	958 124 033	0.446161

Nowadays, with a 64-bit computer, the maximum period of a LCG algorithm is $2^{64} - 1 \approx 1.85 \times 10^{19}$. To obtain a larger period length, one can use more sophisticated methods. For

[3] If the seed is not specified, programming softwares generally use the clock of the computer to generate the initial value.

[4] It is equal to $m - 1$ if $c = 0$.

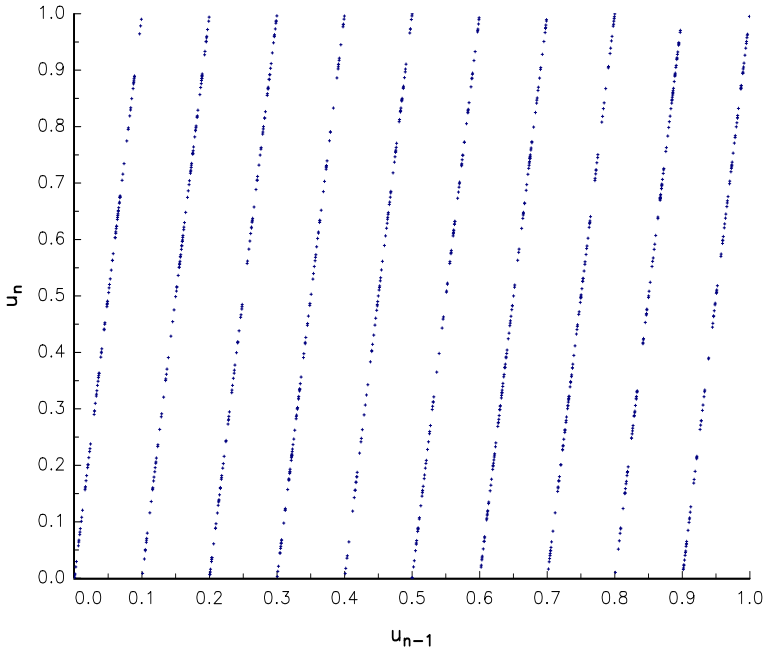


FIGURE 13.1: Lattice structure of the linear congruential generator

instance, multiple recursive generators are based on the following transition equation:

$$x_n = \left(\sum_{i=1}^{k} a_i \cdot x_{n-i} + c \right) \bmod m$$

To obtain a bigger period, we can also combine LCG algorithms with different periods. For instance, the famous MRG32k3a generator of L'Ecuyer (1999) uses two 32-bit multiple recursive generators:

$$\begin{cases} x_n = (1403580 \cdot x_{n-2} - 810728 \cdot x_{n-3}) \bmod m_1 \\ y_n = (527612 \cdot y_{n-1} - 1370589 \cdot y_{n-3}) \bmod m_2 \end{cases}$$

where $m_1 = 2^{32} - 209$ and $m_2 = 2^{32} - 22853$. The uniform random variate is then equal to:

$$u_n = \frac{x_n - y_n + \mathbb{1}\{x_n \leq y_n\} \cdot m_1}{m_1 + 1}$$

L'Ecuyer (1999) showed that the period length of this generator is equal to $2^{191} \approx 3 \times 10^{57}$.

13.1.2 Generating non-uniform random numbers

We now consider X a random variable whose distribution function is noted $\mathbf{F}$. There are many ways to simulate X, but all of them are based on uniform random variates.

13.1.2.1 Method of inversion

Continuous random variables We assume that $\mathbf{F}$ is continuous. Let $Y = \mathbf{F}(X)$ be the integral transform of X. Its cumulative distribution function $\mathbf{G}$ is equal to:

$$\begin{aligned}\mathbf{G}(y) &= \Pr\{Y \leq y\} \\ &= \Pr\{\mathbf{F}(X) \leq y\} \\ &= \Pr\left\{X \leq \mathbf{F}^{-1}(y)\right\} \\ &= \mathbf{F}\left(\mathbf{F}^{-1}(y)\right) \\ &= y\end{aligned}$$

where $\mathbf{G}(0) = 0$ and $\mathbf{G}(1) = 1$. We deduce that $\mathbf{F}(X)$ has a uniform distribution $\mathcal{U}_{[0,1]}$. It follows that if U is a uniform random variable, then $\mathbf{F}^{-1}(U)$ is a random variable whose distribution function is $\mathbf{F}$. To simulate a sequence of random variates $\{x_1, \ldots, x_n\}$, we can simulate a sequence of uniform random variates $\{u_1, \ldots, u_n\}$ and apply the transform $x_i \leftarrow \mathbf{F}^{-1}(u_i)$.

Example 134 *If we consider the generalized uniform distribution $\mathcal{U}_{[a,b]}$, we have $\mathbf{F}(x) = (x-a)/(b-a)$ and $\mathbf{F}^{-1}(u) = a + (b-a)u$. The simulation of random variates x_i is deduced from the uniform random variates u_i by using the following transform:*

$$x_i \leftarrow a + (b-a)\,u_i$$

Example 135 *In the case of the exponential distribution $\mathcal{E}(\lambda)$, we have $\mathbf{F}(x) = 1 - \exp(-\lambda x)$. We deduce that:*

$$x_i \leftarrow -\frac{\ln(1-u_i)}{\lambda}$$

Since $1-U$ is also a uniform distributed random variable, we have:

$$x_i \leftarrow -\frac{\ln(u_i)}{\lambda}$$

Example 136 *In the case of the Pareto distribution $\mathcal{P}(\alpha, x_-)$, we have $\mathbf{F}(x) = 1 - (x/x_-)^{-\alpha}$ and $\mathbf{F}^{-1}(u) = x_-(1-u)^{-1/\alpha}$. We deduce that:*

$$x_i \leftarrow \frac{x_-}{(1-u_i)^{1/\alpha}}$$

The method of inversion is easy to implement when we know the analytical expression of $\mathbf{F}^{-1}$. When it is not the case, we use the Newton-Raphson algorithm:

$$x_i^{m+1} = x_i^m + \frac{u_i - \mathbf{F}(x_i^m)}{f(x_i^m)}$$

where x_i^m is the solution of the equation $\mathbf{F}(x) = u$ at the iteration m. For instance, if we apply this algorithm to the Gaussian distribution $\mathcal{N}(0,1)$, we have:

$$x_i^{m+1} = x_i^m + \frac{u_i - \Phi(x_i^m)}{\phi(x_i^m)}$$

Discrete random variables In the case of a discrete probability distribution $\{(x_1, p_1), (x_2, p_2), \ldots, (x_n, p_n)\}$ where $x_1 < x_2 < \ldots < x_n$, we have:

$$\mathbf{F}^{-1}(u) = \begin{cases} x_1 & \text{if} \quad 0 \leq u \leq p_1 \\ x_2 & \text{if} \quad p_1 < u \leq p_1 + p_2 \\ \vdots & \\ x_n & \text{if} \quad \sum_{k=1}^{n-1} p_k < u \leq 1 \end{cases}$$

In Figure 13.2, we illustrate the method of inversion when the random variable is discrete. We assume that:

x_i	1	2	4	6	7	9	10
p_i	10%	20%	10%	5%	20%	30%	5%
$\mathbf{F}(x_i)$	10%	30%	40%	45%	65%	95%	100%

Because the cumulative distribution function is not continuous, the inverse function is a step function. If we suppose that the uniform random number is 0.5517, we deduce that the corresponding random number for the variable X is equal to 7.

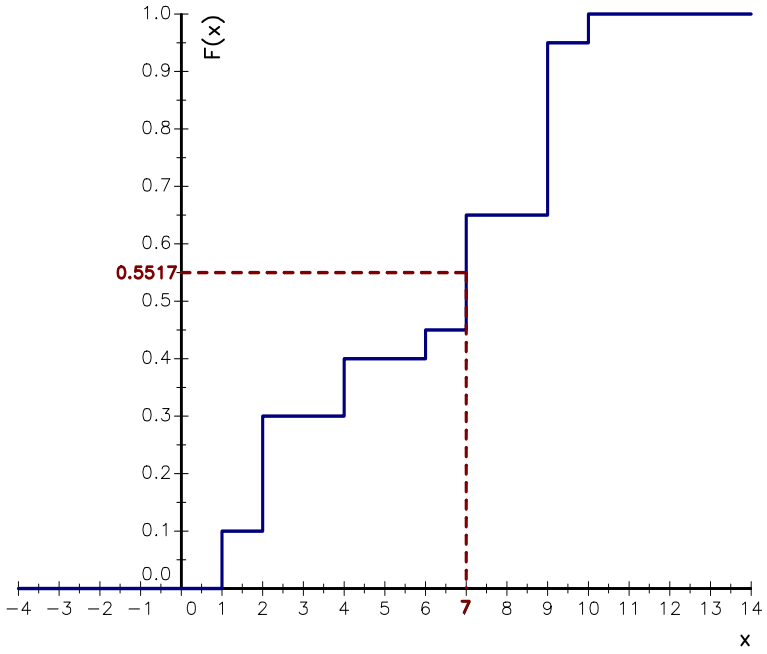


FIGURE 13.2: Inversion method when X is a discrete random variable

Example 137 *If we apply the method of inversion to the Bernoulli distribution* $\mathcal{B}(p)$*, we have:*

$$x \leftarrow \begin{cases} 0 & \textit{if} \quad 0 \leq u \leq 1-p \\ 1 & \textit{if} \quad 1-p < u \leq 1 \end{cases}$$

or:

$$x \leftarrow \begin{cases} 1 & \textit{if} \quad u \leq p \\ 0 & \textit{if} \quad u > p \end{cases}$$

Piecewise distribution functions A piecewise distribution function is defined as follows:

$$\mathbf{F}(x) = \mathbf{F}_m(x) \qquad \text{if } x \in \left]x_{m-1}^\star, x_m^\star\right]$$

where $x_m^\star$ are the knots of the piecewise function and:

$$\mathbf{F}_{m+1}(x_m^\star) = \mathbf{F}_m(x_m^\star)$$

In this case, the simulated value x_i is obtained using a search algorithm:

$$x_i \leftarrow \mathbf{F}_m^{-1}(u_i) \qquad \text{if } \mathbf{F}\left(x_{m-1}^\star\right) < u_i \leq \mathbf{F}(x_m^\star)$$

This means that we have first to calculate the value of $\mathbf{F}(x)$ for all the knots in order to determine which inverse function $\mathbf{F}_m^{-1}$ will be apply.

Let us consider the piecewise exponential model described on page 202. We reiterate that the survival function has the following expression:

$$\mathbf{S}(t) = \mathbf{S}\left(t_{m-1}^\star\right) e^{-\lambda_m\left(t - t_{m-1}^\star\right)} \qquad \text{if } t \in \left]t_{m-1}^\star, t_m^\star\right]$$

We know that $\mathbf{S}(\boldsymbol{\tau}) \sim U$. It follows that:

$$t_i \leftarrow t_{m-1}^\star + \frac{1}{\lambda_m} \ln \frac{\mathbf{S}\left(t_{m-1}^\star\right)}{u_i} \qquad \text{if } \mathbf{S}\left(t_m^\star\right) < u_i \leq \mathbf{S}\left(t_{m-1}^\star\right)$$

Example 138 *We model the default time $\boldsymbol{\tau}$ with the piecewise exponential model and the following parameters:*

$$\lambda = \begin{cases} 5\% & \textit{if } t \textit{ is less or equal than one year} \\ 8\% & \textit{if } t \textit{ is between one and five years} \\ 12\% & \textit{if } t \textit{ is larger than five years} \end{cases}$$

We have $\mathbf{S}(0) = 1$, $\mathbf{S}(1) = 0.9512$ and $\mathbf{S}(5) = 0.6907$. We deduce that:

$$t_i \leftarrow \begin{cases} 0 + (1/0.05) \cdot \ln(1/u_i) & \text{if } u_i \in [0.9512, 1] \\ 1 + (1/0.08) \cdot \ln(0.9512/u_i) & \text{if } u_i \in [0.6907, 0.9512[\\ 5 + (1/0.12) \cdot \ln(0.6907/u_i) & \text{if } u_i \in [0, 0.6907[\end{cases}$$

In Table 13.2, we have reported five simulations t_i of the default time $\boldsymbol{\tau}$. For each simulation, we indicate the values taken by $t_{m-1}^\star$, $\mathbf{S}\left(t_{m-1}^\star\right)$ and λ_m.

TABLE 13.2: Simulation of the piecewise exponential model

u_i	$t_{m-1}^\star$	$S\left(t_{m-1}^\star\right)$	λ_m	t_i
0.9950	0	1.0000	0.05	0.1003
0.3035	5	0.6907	0.12	11.8531
0.5429	5	0.6907	0.12	7.0069
0.9140	1	0.9512	0.08	1.4991
0.7127	1	0.9512	0.08	4.6087

13.1.2.2 Method of transformation

Let $\{Y_1, Y_2, \ldots\}$ be a vector of independent random variables. The simulation of the random variable $X = g(Y_1, Y_2, \ldots)$ is straightforward if we know how to easily simulate the random variables Y_i. We notice that the inversion method is a particular case of the transform method, because we have:

$$X = g(U) = \mathbf{F}^{-1}(U)$$

Example 139 *The Binomial random variable is the sum of n iid Bernoulli random variables:*

$$\mathcal{B}(n, p) = \sum_{i=1}^{n} \mathcal{B}_i(p)$$

We can therefore simulate the Binomial random variate x using n uniform random numbers:

$$x = \sum_{i=1}^{n} \mathbb{1}\{u_i \leq p\}$$

If we would like to simulate the chi-squared random variable $\chi^2(\nu)$, we can use the following relationship:

$$\chi^2(\nu) = \sum_{i=1}^{\nu} \chi_i^2(1) = \sum_{i=1}^{\nu} (\mathcal{N}_i(0, 1))^2$$

We can therefore simulate the $\chi^2(\nu)$ random variate with ν independent Gaussian random numbers $\mathcal{N}(0, 1)$. For that, we generally use the Box-Muller algorithm, which states that if U_1 and U_2 are two independent uniform random variables, then X_1 and X_2 defined by:

$$\begin{cases} X_1 = \sqrt{-2 \ln U_1} \cdot \cos(2\pi U_2) \\ X_2 = \sqrt{-2 \ln U_1} \cdot \sin(2\pi U_2) \end{cases}$$

are independent and follow the Gaussian distribution $\mathcal{N}(0, 1)$.

Remark 149 *To simulate a Student's t random variate x with ν degrees of freedom, we need $\nu + 1$ normal independent random variables n_i:*

$$x \longleftarrow \frac{n_{\nu+1}}{\sqrt{\nu^{-1} \sum_{i=1}^{\nu} n_i^2}}$$

However, this method is not efficient and we generally prefer to use the Bailey algorithm based on the polar transformation[5]*.*

On page 339, we have seen that if N_t is a Poisson process with intensity λ, the duration T between two consecutive events is an exponential distributed random variable. We have:

$$\Pr(T \leq t) = 1 - e^{-\lambda t}$$

Since the durations are independent, we have:

$$T_1 + T_2 + \ldots + T_n = \sum_{i=1}^{n} E_i$$

[5] This method is presented on page 887.

where $E_i \sim \mathcal{E}(\lambda)$. Because the Poisson random variable is the number of events that occur in the unit interval of time, we also have:

$$\begin{aligned} X &= \max\{n : T_1 + T_2 + \ldots + T_n \leq 1\} \\ &= \max\left\{n : \sum_{i=1}^{n} E_i \leq 1\right\} \end{aligned}$$

We notice that:

$$\begin{aligned} \sum_{i=1}^{n} E_i &= -\frac{1}{\lambda}\sum_{i=1}^{n} \ln U_i \\ &= -\frac{1}{\lambda}\ln\prod_{i=1}^{n} U_i \end{aligned}$$

where U_i are *iid* uniform random variables. We deduce that:

$$\begin{aligned} X &= \max\left\{n : -\frac{1}{\lambda}\ln\prod_{i=1}^{n} U_i \leq 1\right\} \\ &= \max\left\{n : \prod_{i=1}^{n} U_i \geq e^{-\lambda}\right\} \end{aligned}$$

We can then simulate the Poisson random variable with the following algorithm:

1. set $n = 0$ and $p = 1$;
2. calculate $n = n + 1$ and $p = p \cdot u_i$ where u_i is a uniform random variate;
3. if $p \geq e^{-\lambda}$, go back to step 2; otherwise, return $X = n - 1$.

13.1.2.3 Rejection sampling

Following Devroye (1986), $\mathbf{F}(x)$ and $\mathbf{G}(x)$ are two distribution functions such that $f(x) \leq cg(x)$ for all x with $c > 1$. We note $X \sim \mathbf{G}$ and consider an independent uniform random variable $U \sim \mathcal{U}_{[0,1]}$. Then, the conditional distribution function of X given that $U \leq f(X)/(cg(X))$ is $\mathbf{F}(x)$.

Let us introduce the random variables B and Z:

$$\begin{aligned} B &= \mathbb{1}\left\{U \leq \frac{f(X)}{cg(X)}\right\} \\ Z &= X \left| U \leq \frac{f(X)}{cg(X)}\right. \end{aligned}$$

We have:

$$\begin{aligned} \Pr\{B = 1\} &= \Pr\left\{U \leq \frac{f(X)}{cg(X)}\right\} \\ &= \mathbb{E}\left[\frac{f(X)}{cg(X)}\right] \\ &= \int_{-\infty}^{+\infty} \frac{f(x)}{cg(x)} g(x)\,\mathrm{d}x \\ &= \frac{1}{c}\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x \\ &= \frac{1}{c} \end{aligned}$$

The distribution function of Z is defined by:

$$\Pr\{Z \leq x\} = \Pr\left\{X \leq x \middle| U \leq \frac{f(X)}{cg(X)}\right\}$$

We deduce that:

$$\begin{aligned}
\Pr\{Z \leq x\} &= \frac{\Pr\left\{X \leq x, U \leq \dfrac{f(X)}{cg(X)}\right\}}{\Pr\left\{U \leq \dfrac{f(X)}{cg(X)}\right\}} \\
&= c\int_{-\infty}^{x}\int_{0}^{f(x)/(cg(x))} g(x)\,\mathrm{d}u\,\mathrm{d}x \\
&= c\int_{-\infty}^{x}\frac{f(x)}{cg(x)}g(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{x} f(x)\,\mathrm{d}x \\
&= \mathbf{F}(x)
\end{aligned}$$

This proves that $Z \sim \mathbf{F}$. From this theorem, we deduce the following acceptance-rejection algorithm:

1. generate two independent random variates x and u from $\mathbf{G}$ and $\mathcal{U}_{[0,1]}$;
2. calculate v as follows:
$$v = \frac{f(x)}{cg(x)}$$
3. if $u \leq v$, return x ('*accept*'); otherwise, go back to step 1 ('*reject*').

The underlying idea of this algorithm is then to simulate the distribution function $\mathbf{F}$ by assuming that it is easier to generate random numbers from $\mathbf{G}$, which is called the proposal distribution. However, some of these random numbers must be '*rejected*', because the function $c \cdot g(x)$ '*dominates*' the density function $f(x)$.

Remark 150 *We notice that the number of iterations N needed to successfully generate Z has a geometric distribution $\mathcal{G}(p)$, where $p = \Pr\{B = 1\} = c^{-1}$ is the acceptance ratio. We deduce that the average number of iterations is equal to $\mathbb{E}[N] = 1/p = c$. In order to maximize the efficiency (or the acceptance ratio) of the algorithm, we have to choose the constant c such that:*

$$c = \sup_x \frac{f(x)}{g(x)}$$

Let us consider the normal distribution $\mathcal{N}(0,1)$. We use the Cauchy distribution function as the proposal distribution, whose probability density function is given by:

$$g(x) = \frac{1}{\pi(1+x^2)}$$

We can show that:

$$\phi(x) \leq \frac{\sqrt{2\pi}}{e^{0.5}}g(x)$$

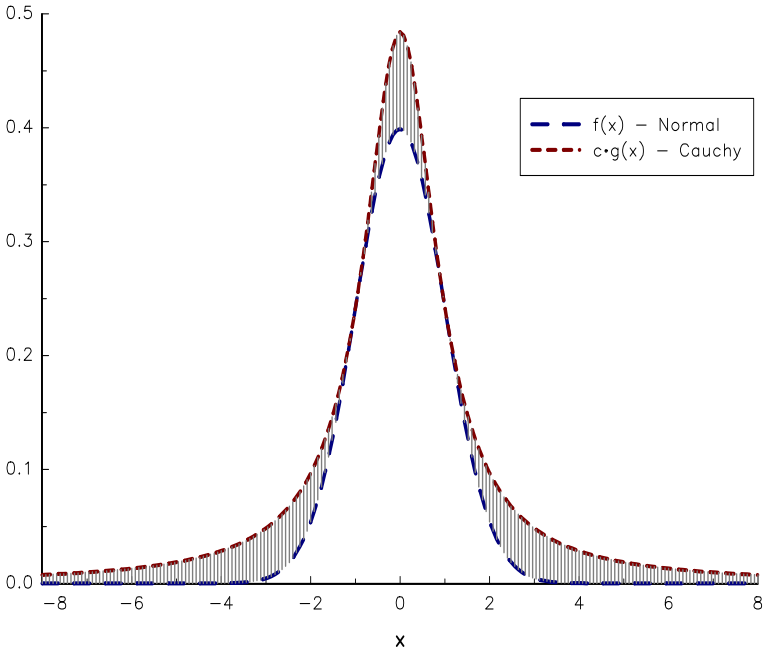


FIGURE 13.3: Rejection sampling applied to the normal distribution

meaning that $c \approx 1.52$. In Figure 13.3, we report the functions $f(x) = \phi(x)$ and $c \cdot g(x)$. The goal of the acceptance-rejection algorithm is to '*eliminate*' the random numbers, which are located in the cross-hatched region. Concerning the Cauchy distribution, we have:

$$\mathbf{G}(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$$

and:

$$\mathbf{G}^{-1}(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$$

Therefore, we deduce that the following algorithm for simulating the distribution function $\mathcal{N}(0,1)$:

1. generate two independent uniform random variates u_1 and u_2 and set:

$$x \leftarrow \tan\left(\pi\left(u_1 - \frac{1}{2}\right)\right)$$

2. calculate v as follows:

$$\begin{aligned} v &= \frac{e^{0.5}\phi(x)}{\sqrt{2\pi}g(x)} \\ &= \frac{\left(1+x^2\right)}{2e^{(x^2-1)/2}} \end{aligned}$$

3. if $u_2 \leq v$, accept x; otherwise, go back to step 1.

To illustrate this algorithm, we have simulated six Gaussian distributed random variates in Table 13.3. We notice that four simulations have been rejected. Using 1 000 simulations of

Cauchy random variates, we obtained the density given in Figure 13.4, which is very close to the exact probability density function. In our case, we accept 683 simulations, meaning that the acceptance ratio[6] is 68.3%.

TABLE 13.3: Simulation of the standard Gaussian distribution using the acceptance-rejection algorithm

u_1	u_2	x	v	test	z
0.9662	0.1291	9.3820	0.0000	reject	
0.0106	0.1106	−30.0181	0.0000	reject	
0.3120	0.8253	−0.6705	0.9544	accept	−0.6705
0.9401	0.9224	5.2511	0.0000	reject	
0.2170	0.4461	−1.2323	0.9717	accept	−1.2323
0.6324	0.0676	0.4417	0.8936	accept	0.4417
0.6577	0.1344	0.5404	0.9204	accept	0.5404
0.1596	0.6670	−1.8244	0.6756	accept	−1.8244
0.4183	0.3872	−0.2625	0.8513	accept	−0.2625
0.9625	0.0752	8.4490	0.0000	reject	

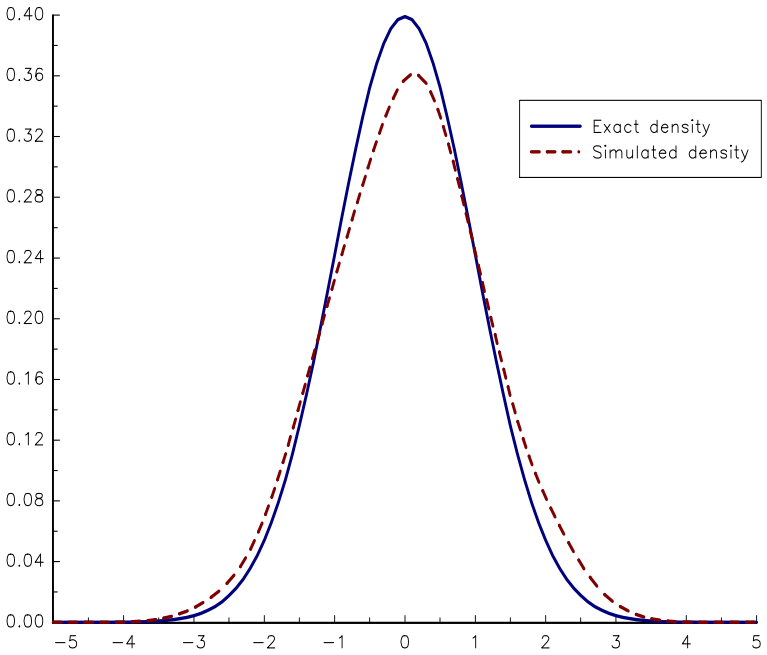


FIGURE 13.4: Comparison of the exact and simulated densities

Remark 151 *The discrete case is analogous to the continuous case. Let* $p(k)$ *and* $q(k)$ *be the probability mass function of* Z *and* X *such that* $p(k) \leq cq(k)$ *for all* k *with* $c \geq 1$. *We consider an independent uniform random variable* $U \sim \mathcal{U}_{[0,1]}$. *Then, the conditional pmf of* X *given that* $U \leq p(X) / (cq(X))$ *is the pmf* $p(k)$ *of* Z.

[6]The theoretical acceptance ratio is equal to $1/1.52 \approx 65.8\%$.

13.1.2.4 Method of mixtures

A finite mixture can be decomposed as a weighted sum of distribution functions. We have:

$$\mathbf{F}(x) = \sum_{k=1}^{n} \pi_k \cdot \mathbf{G}_k(x)$$

where $\pi_k \geq 0$ and $\sum_{k=1}^{n} \pi_k = 1$. We deduce that the probability density function is:

$$f(x) = \sum_{k=1}^{n} \pi_k \cdot g_k(x)$$

To simulate the probability distribution $\mathbf{F}$, we introduce the random variable B, whose probability mass function is defined by:

$$p(k) = \Pr\{B = k\} = \pi_k$$

It follows that:

$$\mathbf{F}(x) = \sum_{k=1}^{n} \Pr\{B = k\} \cdot \mathbf{G}_k(x)$$

We deduce the following algorithm:

1. generate the random variate b from the probability mass function $p(k)$;
2. generate the random variate x from the probability distribution $\mathbf{G}_b(x)$.

Example 140 *We assume that the default time $\boldsymbol{\tau}$ follows the hyper-exponential model:*

$$f(t) = \pi \cdot \lambda_1 e^{-\lambda_1 t} + (1-\pi) \cdot \lambda_2 e^{-\lambda_2 t}$$

To simulate this model, we consider the following algorithm:

1. we generate u and v two independent uniform random numbers;
2. we have:
$$b \leftarrow \begin{cases} 1 & \text{if } u \leq \pi \\ 2 & \text{otherwise} \end{cases}$$
3. the simulated value of $\boldsymbol{\tau}$ is:
$$t \leftarrow \begin{cases} -\lambda_1 \ln v & \text{if } b = 1 \\ -\lambda_2 \ln v & \text{if } b = 2 \end{cases}$$

Remark 152 *The previous approach can be easily extended to continuous mixtures:*

$$f(x) = \int_{\Omega} \pi(\omega)\, g(x;\omega)\, \mathrm{d}\omega$$

where $\omega \in \Omega$ is a parameter of the distribution $\mathbf{G}$. For instance, we have seen that the negative binomial distribution is a gamma-Poisson mixture distribution:

$$\begin{cases} \mathcal{NB}(r,p) \sim \mathcal{P}(\Lambda) \\ \Lambda \sim \mathcal{G}(r,(1-p)/p) \end{cases}$$

To simulate the negative binomial distribution, we first simulate the gamma random variate $g \sim \mathcal{G}(r,(1-p)/p)$, and then simulate the Poisson random variable, whose parameter[7] λ is equal to g.

[7] This means that the parameter λ changes at each simulation.

13.1.3 Generating random vectors

In this section, we consider algorithms for simulating a random vector $X = (X_1, \ldots, X_n)$ from a given distribution function $\mathbf{F}(x) = \mathbf{F}(x_1, \ldots, x_n)$. In fact, the previous methods used to generate a random variable are still valid in the multidimensional case.

13.1.3.1 Method of conditional distributions

The method of inversion cannot be applied in the multivariate case, because $U = \mathbf{F}(X_1, \ldots, X_n)$ is not any longer a uniform random variable. However, if $X_1, \ldots, X_n$ are independent, we have:

$$\mathbf{F}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathbf{F}_i(x_i)$$

To simulate X, we can then generate each component $X_i \sim \mathbf{F}_i$ individually, for example by applying the method of inversion. When $X_1, \ldots, X_n$ are dependent, we have:

$$\begin{aligned} \mathbf{F}(x_1, \ldots, x_n) &= \mathbf{F}_1(x_1)\, \mathbf{F}_{2|1}(x_2 \mid x_1)\, \mathbf{F}_{3|1,2}(x_3 \mid x_1, x_2) \times \cdots \times \\ &\quad \mathbf{F}_{n|1,\ldots,n-1}(x_n \mid x_1, \ldots, x_{n-1}) \\ &= \prod_{i=1}^{n} \mathbf{F}_{i|1,\ldots,i-1}(x_i \mid x_1, \ldots, x_{i-1}) \end{aligned}$$

where $\mathbf{F}_{i|1,\ldots,i-1}(x_i \mid x_1, \ldots, x_{i-1})$ is the conditional distribution of X_i given $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$. Let us denote this '*conditional*' random variable Y_i. We notice that the random variables $(Y_1, \ldots, Y_n)$ are independent. Therefore, the underlying idea of the method of conditional distributions is to transform the random vector X by a vector Y of independent random variables. We obtain the following algorithm:

1. generate x_1 from $\mathbf{F}_1(x)$ and set $i = 2$;
2. generate x_i from $\mathbf{F}_{i|1,\ldots,i-1}(x \mid x_1, \ldots, x_{i-1})$ given $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$ and set $i = i + 1$;
3. repeat step 2 until $i = n$.

$\mathbf{F}_{i|1,\ldots,i-1}(x \mid x_1, \ldots, x_{i-1})$ is a univariate distribution function, which depends on the argument x and parameters $x_1, \ldots, x_{i-1}$. To simulate it, we can therefore use the method of inversion:

$$x_i \leftarrow \mathbf{F}^{-1}_{i|1,\ldots,i-1}(u_i \mid x_1, \ldots, x_{i-1})$$

where $\mathbf{F}^{-1}_{i|1,\ldots,i-1}$ is the inverse of the conditional distribution function and u_i is a uniform random variate.

Example 141 *We consider the bivariate logistic distribution defined as:*

$$\mathbf{F}(x_1, x_2) = \left(1 + e^{-x_1} + e^{-x_2}\right)^{-1}$$

We have $\mathbf{F}_1(x_1) = \mathbf{F}(x_1, +\infty) = (1 + e^{-x_1})^{-1}$. We deduce that the conditional distribution of X_2 given $X_1 = x_1$ is:

$$\begin{aligned} \mathbf{F}_{2|1}(x_2 \mid x_1) &= \frac{\mathbf{F}(x_1, x_2)}{\mathbf{F}_1(x_1)} \\ &= \frac{1 + e^{-x_1}}{1 + e^{-x_1} + e^{-x_2}} \end{aligned}$$

We obtain:

$$\mathbf{F}_1^{-1}(u) = \ln u - \ln(1-u)$$

and:

$$\mathbf{F}_{2|1}^{-1}(u \mid x_1) = \ln u - \ln(1-u) - \ln\left(1 + e^{-x_1}\right)$$

We deduce the following algorithm:

1. generate two independent uniform random variates u_1 and u_2;
2. generate x_1 from u_1:
$$x_1 \leftarrow \ln u_1 - \ln(1-u_1)$$
3. generate x_2 from u_2 and x_1:
$$x_2 \leftarrow \ln u_2 - \ln(1-u_2) - \ln\left(1 + e^{-x_1}\right)$$

Because we have $(1 + e^{-x_1})^{-1} = u_1$, the last step can be replaced by:

3'. generate x_2 from u_2 and u_1:
$$x_2 \leftarrow \ln\left(\frac{u_1 u_2}{1-u_2}\right)$$

The method of conditional distributions can be used for simulating uniform random vectors $(U_1, \ldots, U_n)$ generated by copula functions. In this case, we have:

$$\begin{aligned}\mathbf{C}(u_1, \ldots, u_n) &= \mathbf{C}_1(u_1)\,\mathbf{C}_{2|1}(u_2 \mid u_1)\,\mathbf{C}_{3|1,2}(u_3 \mid u_1, u_2) \times \cdots \times \\ &\quad \mathbf{C}_{n|1,\ldots,n-1}(u_n \mid u_1, \ldots, u_{n-1}) \\ &= \prod_{i=1}^{n} \mathbf{C}_{i|1,\ldots,i-1}(u_i \mid u_1, \ldots, u_{i-1})\end{aligned}$$

where $\mathbf{C}_{i|1,\ldots,i-1}(u_i \mid u_1, \ldots, u_{i-1})$ is the conditional distribution of U_i given $U_1 = u_1, \ldots, U_{i-1} = u_{i-1}$ By definition, we have $\mathbf{C}_1(u_1) = u_1$. We obtain the following algorithm:

1. generate n independent uniform random variates $v_1, \ldots, v_n$;
2. generate $u_1 \leftarrow v_1$ and set $i = 2$;
3. generate u_i by finding the root of the equation:
$$\mathbf{C}_{i|1,\ldots,i-1}(u_i \mid u_1, \ldots, u_{i-1}) = v_i$$
and set $i = i + 1$;
4. repeat step 3 until $i = n$.

For some copula functions, there exists an analytical expression of the inverse of the conditional copula. In this case, the third step is replaced by:

3'. generate u_i by the inversion method:
$$u_i \leftarrow \mathbf{C}_{i|1,\ldots,i-1}^{-1}(v_i \mid u_1, \ldots, u_{i-1})$$

Remark 153 *For any probability distribution, the conditional distribution can be calculated as follows:*

$$\mathbf{F}_{i|1,\ldots,i-1}(x_i \mid x_1,\ldots,x_{i-1}) = \frac{\mathbf{F}(x_1,\ldots,x_{i-1},x_i)}{\mathbf{F}(x_1,\ldots,x_{i-1})}$$

In particular, we have:

$$\begin{aligned}\partial_1 \mathbf{F}(x_1,x_2) &= \partial_1\left(\mathbf{F}_1(x_1)\cdot\mathbf{F}_{2|1}(x_2 \mid x_1)\right)\\ &= f_1(x_1)\cdot\mathbf{F}_{2|1}(x_2 \mid x_1)\end{aligned}$$

For copula functions, the density $f_1(x_1)$ is equal to 1, meaning that:

$$\mathbf{C}_{2|1}(u_2 \mid u_1) = \partial_1 \mathbf{C}(u_1,u_2)$$

We can generalize this result and show that the conditional copula given some random variables U_i for $i \in \Omega$ is equal to the cross-derivative of the copula function $\mathbf{C}$ with respect to the arguments u_i for $i \in \Omega$.

We recall that Archimedean copulas are defined as:

$$\mathbf{C}(u_1,u_2) = \varphi^{-1}(\varphi(u_1)+\varphi(u_2))$$

where $\varphi(u)$ is the generator function. We have:

$$\varphi(\mathbf{C}(u_1,u_2)) = \varphi(u_1)+\varphi(u_2)$$

and:

$$\varphi'(\mathbf{C}(u_1,u_2))\cdot\frac{\partial\,\mathbf{C}(u_1,u_2)}{\partial\,u_1} = \varphi'(u_1)$$

We deduce the following expression of the conditional copula:

$$\begin{aligned}\mathbf{C}_{2|1}(u_2 \mid u_1) &= \frac{\partial\,\mathbf{C}(u_1,u_2)}{\partial\,u_1}\\ &= \frac{\varphi'(u_1)}{\varphi'(\varphi^{-1}(\varphi(u_1)+\varphi(u_2)))}\end{aligned}$$

The calculation of the inverse function gives:

$$\mathbf{C}_{2|1}^{-1}(v \mid u_1) = \varphi^{-1}\left(\varphi\left(\varphi'^{-1}\left(\frac{\varphi'(u_1)}{v}\right)\right)-\varphi(u_1)\right)$$

We obtain the following algorithm for simulating Archimedean copulas:

1. generate two independent uniform random variates v_1 and v_2;
2. generate $u_1 \leftarrow v_1$;
3. generate u_2 by the inversion method:

$$u_2 \leftarrow \varphi^{-1}\left(\varphi\left(\varphi'^{-1}\left(\frac{\varphi'(u_1)}{v_2}\right)\right)-\varphi(u_1)\right)$$

Example 142 *We consider the Clayton copula:*

$$\mathbf{C}(u_1,u_2) = \left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta}$$

The Clayton copula is an Archimedean copula, whose generator function is:

$$\varphi(u) = u^{-\theta} - 1$$

We deduce that:

$$\begin{aligned}
\varphi^{-1}(u) &= (1+u)^{-1/\theta} \\
\varphi'(u) &= -\theta u^{-(\theta+1)} \\
\varphi'^{-1}(u) &= (-u/\theta)^{-1/(\theta+1)}
\end{aligned}$$

After some calculations, we obtain:

$$\mathbf{C}_{2|1}^{-1}(v \mid u_1) = \left(1 + u_1^{-\theta}\left(v^{-\theta/(\theta+1)} - 1\right)\right)^{-1//\theta}$$

In Table 13.4, we simulate five realizations of the Clayton copula using the inverse function of the conditional copula. In the case $\theta = 0.01$, u_2 is close to v_2 because the Clayton copula is the product copula $\mathbf{C}^{\perp}$ when θ tends to 0. In the case $\theta = 1.5$, we note the impact of the conditional copula on the simulation of u_2.

TABLE 13.4: Simulation of the Clayton copula

Random uniform variates		Clayton copula $\theta = 0.01$		Clayton copula $\theta = 1.5$	
v_1	v_2	u_1	u_2	u_1	u_2
0.2837	0.4351	0.2837	0.4342	0.2837	0.3296
0.0386	0.2208	0.0386	0.2134	0.0386	0.0297
0.3594	0.5902	0.3594	0.5901	0.3594	0.5123
0.3612	0.3268	0.3612	0.3267	0.3612	0.3247
0.0797	0.6479	0.0797	0.6436	0.0797	0.1704

13.1.3.2 Method of transformation

To simulate a Gaussian random vector $X \sim \mathcal{N}(\mu, \Sigma)$, we consider the following transformation:

$$X = \mu + A \cdot N$$

where $AA^{\top} = \Sigma$ and $N \sim \mathcal{N}(\mathbf{0}, I)$. Therefore, we can simulate a correlated Gaussian random vector by using n independent Gaussian random variates $\mathcal{N}(0,1)$ and finding a square matrix A such that $AA^{\top} = \Sigma$. Since we know that Σ is a positive definite symmetric matrix, it has a unique Cholesky decomposition:

$$\Sigma = PP^{\top}$$

where P is a lower triangular matrix.

Remark 154 *The decomposition $AA^{\top} = \Sigma$ is not unique. For instance, if we use the eigendecomposition:*

$$\Sigma = U\Lambda U^{\top}$$

we can set $A = U\Lambda^{1/2}$. Indeed, we have:

$$\begin{aligned}
AA^{\top} &= U\Lambda^{1/2}\Lambda^{1/2}U^{\top} \\
&= U\Lambda U^{\top} \\
&= \Sigma
\end{aligned}$$

To simulate a multivariate Student's t distribution $Y = (Y_1, \ldots, Y_n) \sim \mathbf{T}_n(\Sigma, \nu)$, we use the relationship:

$$Y_i = \frac{X_i}{\sqrt{Z/\nu}}$$

where the random vector $X = (X_1, \ldots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and the random variable $Z \sim \chi^2(\nu)$ are independent.

The transformation method is particularly useful for simulating copula functions. Indeed, if $X = (X_1, \ldots, X_n) \sim \mathbf{F}$, then the probability distribution of the random vector $U = (U_1, \ldots, U_n)$ defined by:

$$U_i = \mathbf{F}_i(X)$$

is the copula function $\mathbf{C}$ associated to $\mathbf{F}$.

Example 143 *To simulate the Normal copula with the matrix of parameters ρ, we simulate $N \sim \mathcal{N}(\mathbf{0}, I)$ and apply the transformation:*

$$U = \Phi(P \cdot N)$$

where P is the Cholesky decomposition of the correlation matrix ρ.

Example 144 *To simulate the Student's t copula with the matrix of parameters ρ and ν degrees of freedom, we simulate $T \sim \mathbf{T}_n(\rho, \nu)$ and apply the transformation:*

$$U_i = \mathbf{T}_v(T_i)$$

In Figures 13.5 and 13.6, we draw 1 024 simulations of Normal and t_1 copulas for different values of ρ. We notice that the Student's t copula correlates the extreme values more than the Normal copula.

On page 735, frailty copulas have been defined as:

$$\mathbf{C}(u_1, \ldots, u_n) = \psi\left(\psi^{-1}(u_1) + \ldots + \psi^{-1}(u_n)\right)$$

where $\psi(x)$ is the Laplace transform of a random variable X. Using the mixture representation of frailty copulas, Marshall and Olkin (1988) showed that they can be generated using the following algorithm:

1. simulate n independent uniform random variates $v_1, \ldots, v_n$;
2. simulate the frailty random variate x with the Laplace transform ψ;
3. apply the transformation:

$$(u_1, \ldots, u_n) \leftarrow \left(\psi\left(-\frac{\ln u_1}{x}\right), \ldots, \psi\left(-\frac{\ln u_n}{x}\right)\right)$$

For instance, the Clayton copula is a frailty copula where $\psi(x) = (1+x)^{-1/\theta}$ is the Laplace transform of the gamma random variable $\mathcal{G}(1/\theta, 1)$. Therefore, the algorithm to simulate the Clayton copula is:

$$\begin{cases} x \leftarrow \mathcal{G}(1/\theta, 1) \\ (u_1, \ldots, u_n) \leftarrow \left(\left(1 - \frac{\ln u_1}{x}\right)^{-1/\theta}, \ldots, \left(1 - \frac{\ln u_n}{x}\right)^{-1/\theta}\right) \end{cases}$$

Examples of simulating the Clayton copula using this algorithm is given in Figure 13.7.

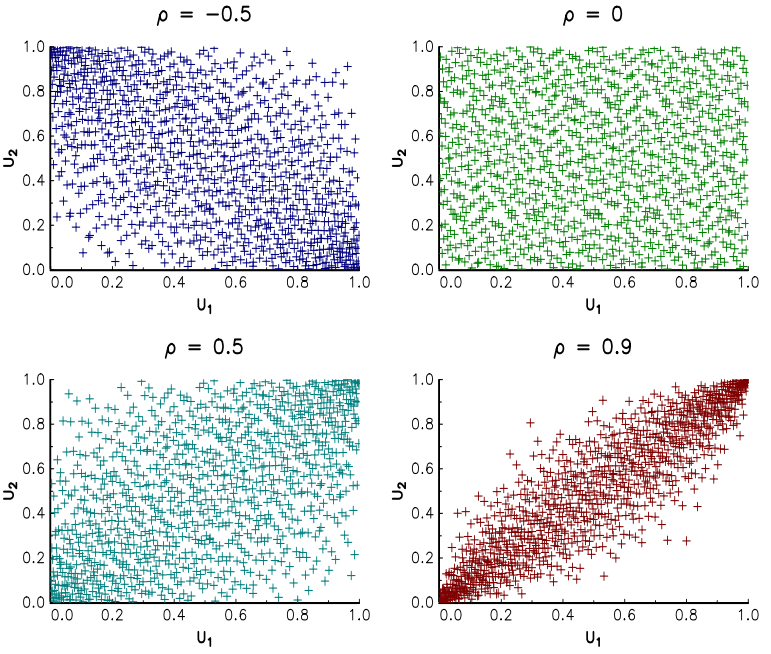


FIGURE 13.5: Simulation of the Normal copula

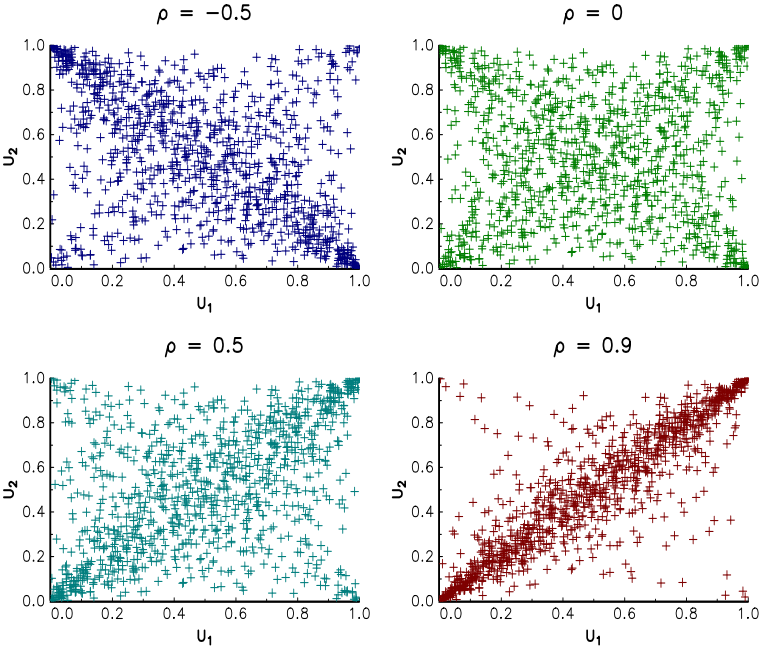


FIGURE 13.6: Simulation of the t_1 copula

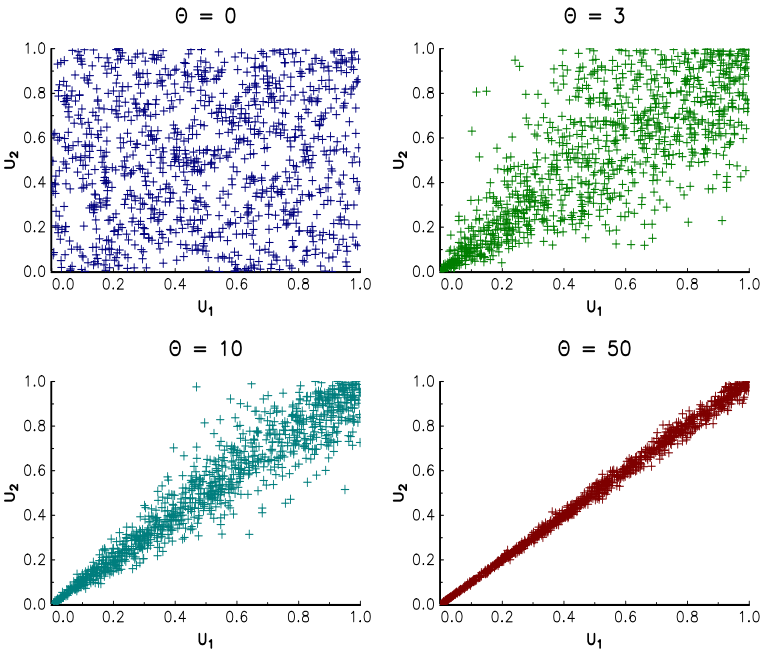


FIGURE 13.7: Simulation of the Clayton copula

Remark 155 *For other frailty copulas, the reader can refer to the survey of McNeil (2008) for the list of Laplace transforms and corresponding algorithms to simulate the frailty random variable.*

We now consider the multivariate distribution $\mathbf{F}(x_1, \ldots, x_n)$, whose canonical decomposition is defined as:

$$\mathbf{F}(x_1, \ldots, x_n) = \mathbf{C}(\mathbf{F}_1(x_1), \ldots, \mathbf{F}_n(x_n))$$

We recall that if $(U_1, \ldots, U_n) \sim \mathbf{C}$, the random vector $(X_1, \ldots, X_n) = \left(\mathbf{F}_1^{-1}(U_1), \ldots, \mathbf{F}_n^{-1}(U_n)\right)$ follows the distribution function $\mathbf{F}$. We deduce the following algorithm:

$$\begin{cases} (u_1, \ldots, u_n) \leftarrow \mathbf{C} \\ (x_1, \ldots, x_n) \leftarrow \left(\mathbf{F}_1^{-1}(u_1), \ldots, \mathbf{F}_n^{-1}(u_n)\right) \end{cases}$$

Let us consider that the default time $\boldsymbol{\tau}$ and the loss given default LGD of one counterparty are distributed according to the exponential distribution $\mathcal{E}(5\%)$ and the beta distribution $\mathcal{B}(2, 2)$. We also assume that the default time and the loss given default are correlated and the dependence function is a Clayton copula. In Figure 13.8, we use the Clayton random variates generated in Figure 13.7 and apply exponential and beta inverse transforms to them. For the beta distribution, we use the Newton-Raphson algorithm to generate the LGD random variable.

The previous algorithms suppose that we know the analytical expression $\mathbf{F}_i$ of the univariate probability distributions in order to calculate the quantile function $\mathbf{F}_i^{-1}$. This is not always the case. For instance, in the operational risk, the loss of the bank is equal to the sum of aggregate losses:

$$L = \sum_{k=1}^{K} S_k$$

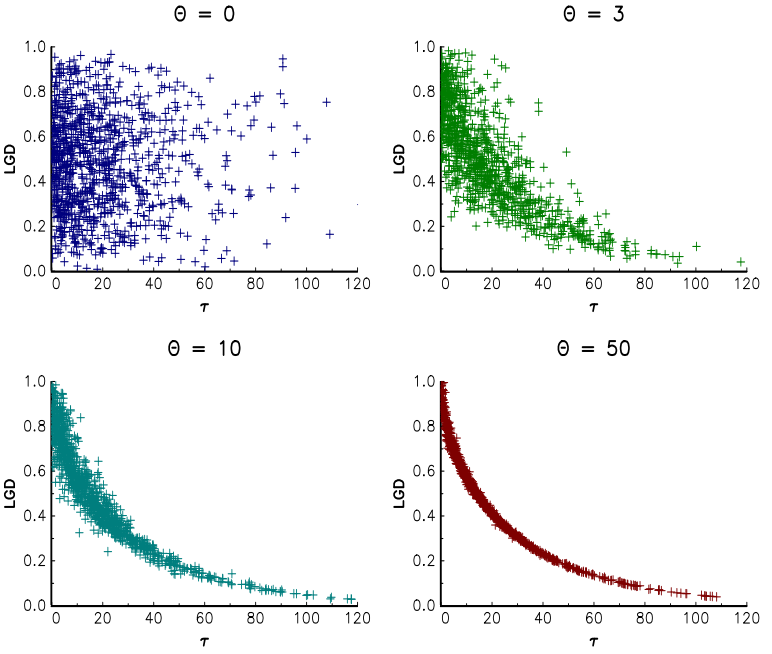


FIGURE 13.8: Simulation of the correlated random vector $(\boldsymbol{\tau}, \text{LGD})$

where S_k is also the sum of individual losses for the k^{th} cell of the mapping matrix. In practice, the probability distribution of S_k is estimated by the method of simulations. In this case, we have to use the method of the empirical quantile function. Let $\mathbb{F}_{i,m}$ be the empirical process of X_i. We know that:

$$\sup_x |\mathbb{F}_{i,m}(x) - \mathbf{F}_i(x)| \to 0 \qquad \text{when } m \to \infty$$

We note $\mathbb{U}_m$ and $\mathbb{F}_m$ the empirical processes corresponding to the distribution functions $\mathbf{C}(u_1, \ldots, u_n)$ and $\mathbf{F}(x_1, \ldots, x_n)$. The Glivenko-Cantelli theorem tells us that:

$$\sup_{x_1,\ldots,x_n} |\mathbb{F}_m(x_1, \ldots, x_n) - \mathbf{F}(x_1, \ldots, x_n)| \to 0 \qquad \text{when } m \to \infty$$

We deduce that:

$$\sup_{u_1,\ldots,u_n} \left|\mathbb{U}_{m_2}\left(\mathbb{F}_{1,m_1}^{-1}(u_1), \ldots, \mathbb{F}_{n,m_1}^{-1}(u_n)\right) - \mathbf{C}\left(\mathbf{F}_1^{-1}(u_1), \ldots, \mathbf{F}_n^{-1}(u_n)\right)\right| \to 0$$

when both m_1 and m_2 tend to ∞. It follows that the method of the empirical quantile function is implemented as follows:

1. for each random variable X_i, simulate m_1 random variates $x_{i,m}^\star$ and estimate the empirical distribution $\hat{\mathbf{F}}_i$;
2. simulate a random vector $(u_1, \ldots, u_n)$ from the copula function $\mathbf{C}(u_1, \ldots, u_n)$;
3. simulate the random vector $(x_1, \ldots, x_n)$ by inverting the empirical distributions $\hat{\mathbf{F}}_i$:
$$x_i \leftarrow \hat{\mathbf{F}}_i^{-1}(u_i)$$
we also have:
$$x_i \leftarrow \inf\left\{x \,\middle|\, \frac{1}{m_1}\sum_{m=1}^{m_1} \mathbf{1}\left\{x \le x_{i,m}^\star\right\} \ge u_i\right\}$$
4. repeat steps 2 and 3 m_2 times.

In Figure 13.9, we illustrate this algorithm by assuming that $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{N}(0,1)$ and the dependence function of (X_1, X_2) is the Clayton copula with parameter $\theta = 3$. If we use $m_1 = 50$ simulations to estimate the quantile function of X_1 and X_2, the approximation is not good. However, when we consider a large number of simulations ($m_1 = 5000$), we obtain simulated values of the random vector (X_1, X_2) that are close to the simulated values calculated with the analytical quantile function $\Phi^{-1}(u)$. We now consider a more complex example. We assume that $X_1 \sim \mathcal{N}(-1,2)$, $X_2 \sim \mathcal{N}(0,1)$, $Y_1 \sim \mathcal{G}(0.5)$ and $Y_2 \sim \mathcal{G}(1,2)$ are four independent random variables. Let $(Z_1 = X_1 + Y_1, Z_2 = X_2 \cdot Y_2)$ be the random vector, whose dependence function is the t copula with parameters $\nu = 2$ and $\rho = -70\%$. It is obvious that it is not possible to find an analytical expression of the marginal distributions of Z_1 and Z_2. However, the random variables Z_1 and Z_2 are easy to simulate (Figure 13.10). This is why we can use the method of the empirical quantile function to simulate the random vector (Z_1, Z_2). A sample of 4 000 simulated values of the vector (Z_1, Z_2) is reported in Figure 13.11.

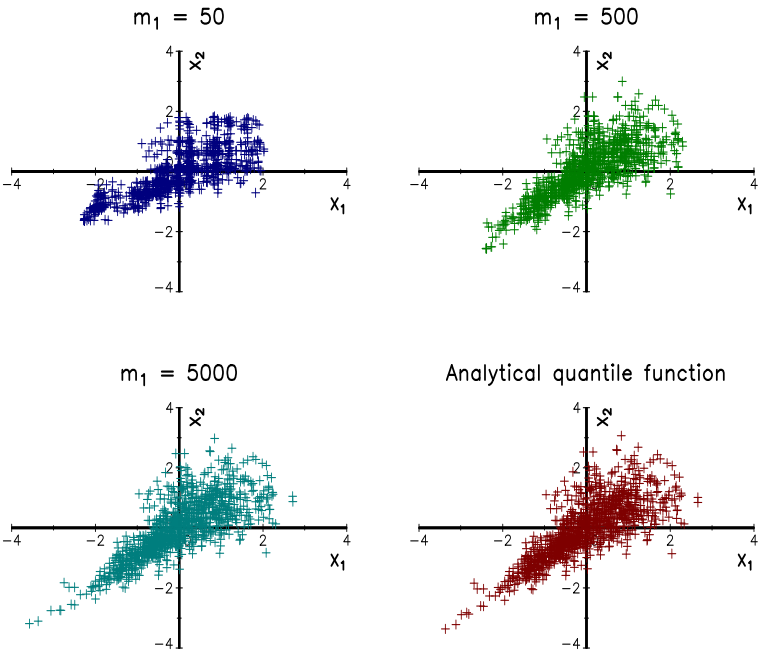


FIGURE 13.9: Convergence of the method of the empirical quantile function

13.1.4 Generating random matrices

The simulation of random matrices is a specialized topic, which is generally not covered by textbooks. However, the tools presented in this section are very useful in finance. This is particularly true when we would like to measure the correlation risk.

13.1.4.1 Orthogonal and covariance matrices

An orthogonal matrix Q is a square $n \times n$ matrix, whose columns and rows are orthonormal vectors:

$$Q^\top Q = QQ^\top = I_n$$

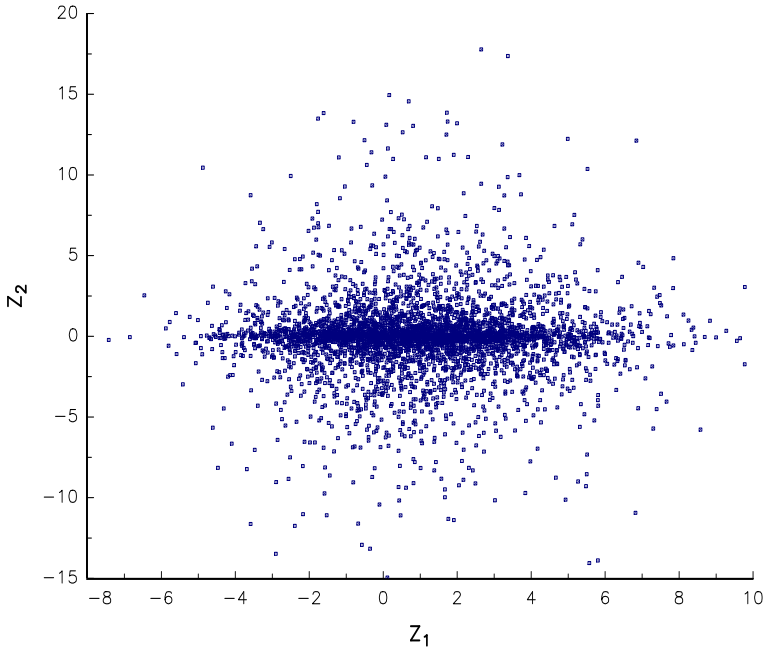


FIGURE 13.10: Simulation of the random variables Z_1 and Z_2

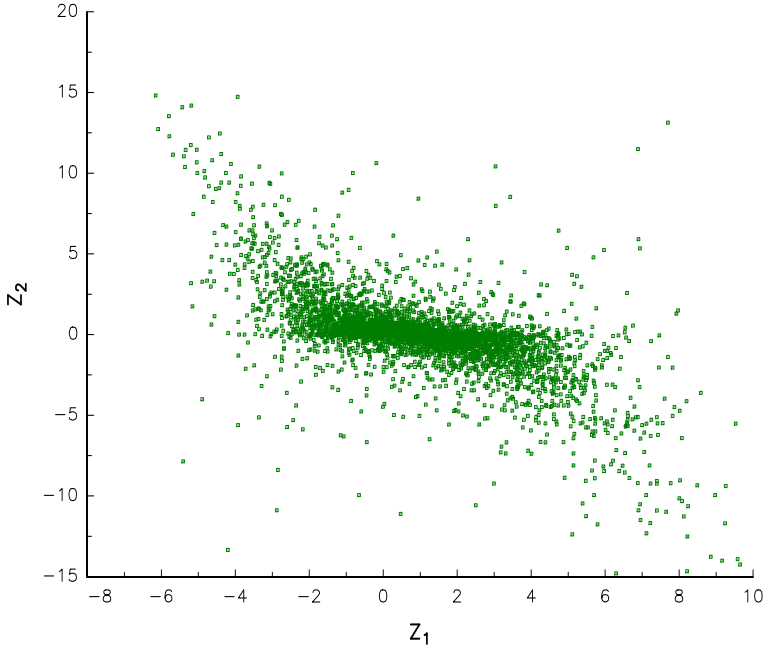


FIGURE 13.11: Simulation of the random vector (Z_1, Z_2)

It follows that $Q^{-1} = Q$. Generally, Monte Carlo methods require generation of random orthogonal matrices distributed according to the Haar measure[8]. Anderson *et al.* (1987) proposed two simple algorithms to generate Q:

1. Let X be a $n \times n$ matrix of independent standard Gaussian random variables. Q is the unitary matrix of the QR factorization of $X = QR$ where R is an upper triangular factorization.

2. Let X be a $n \times p$ matrix of independent standard Gaussian random variables with $p \geq n$. Q corresponds to the matrix V of the eigendecomposition $X^\top X = V \Lambda V^\top$ or the matrix U of the singular value decomposition $X^\top X = U \Sigma V^*$.

Stewart (1980) proposed another popular algorithm based on the Household transformation. Let H_x be the symmetric orthogonal matrix defined as:

$$H_x x = \pm \|x\| \cdot \mathbf{e}_1$$

We consider a series of independent Gaussian random vectors: $x_1 \sim \mathcal{N}_n(0, I_n)$, $x_2 \sim \mathcal{N}_{n-1}(0, I_{n-1})$, etc. We form the matrix $\tilde{H}_k = \text{diag}(I_{k-1}, H_{x_k})$. The random orthogonal matrix Q is then generated by the product:

$$Q = \left(\prod_{k=1}^{n-1} \tilde{H}_k \right) D$$

where D is the diagonal matrix with entries ± 1. To illustrate this algorithm, we simulate random orthogonal matrices Q for different values of n, and we report the distribution of the eigenvalues of Q in Figure 13.12. We verify that they are almost uniformly distributed on the unit sphere.

Remark 156 *To simulate a random covariance matrix Σ with specified eigenvalues $\lambda_1, \ldots, \lambda_n$, we generate a random orthogonal matrix Q and consider the transformation:*

$$\Sigma = Q \Lambda Q^\top$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

13.1.4.2 Correlation matrices

A correlation matrix $\mathbb{C}$ is a symmetric positive definite matrix, whose diagonal elements are equal to 1. It follows that the sum of the eigenvalues is exactly equal to n. The previous algorithm can be used to simulate a random correlation matrix. Indeed, we only need to transform Σ into $\mathbb{C}$:

$$\mathbb{C}_{i,j} = \frac{\Sigma_{i,j}}{\sqrt{\Sigma_{i,i} \cdot \Sigma_{j,j}}}$$

However, this method is not always interesting, because it does not preserve the specified eigenvalues $\lambda_1, \ldots, \lambda_n$. Let us consider an example with $\lambda_1 = 0.5$, $\lambda_2 = 1.00$ and $\lambda_3 = 1$. A simulation of Σ gives:

$$\Sigma = \begin{pmatrix} 1.28570 & -0.12868 & 0.37952 \\ -0.12868 & 0.89418 & 0.16377 \\ 0.37952 & 0.16377 & 0.82012 \end{pmatrix}$$

[8] Any column or any row of Q has a uniform distribution over the n-dimensional unit sphere.

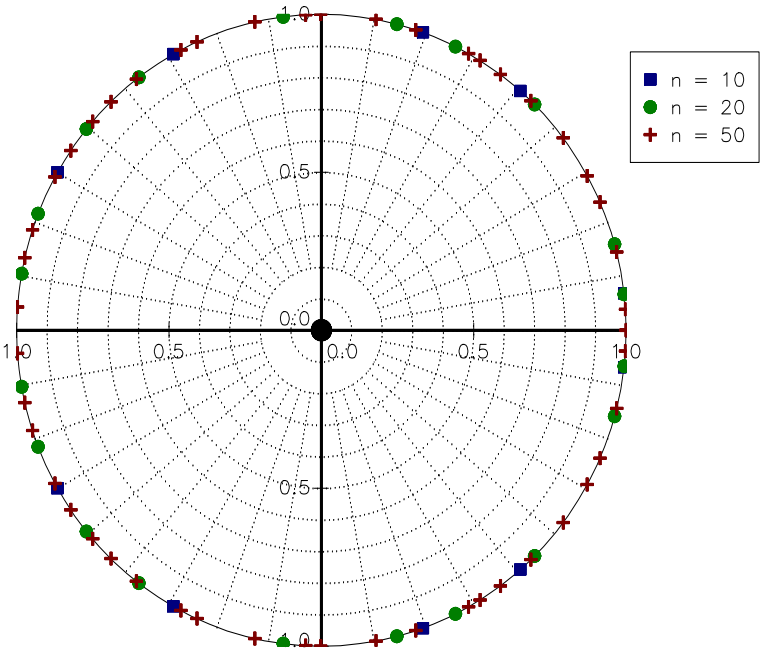


FIGURE 13.12: Distribution of the eigenvalues of simulated random orthogonal matrices

We deduce the following random correlation matrix:

$$\mathbb{C} = \begin{pmatrix} 1.00000 & -0.12001 & 0.36959 \\ -0.12001 & 1.00000 & 0.19124 \\ 0.36959 & 0.19124 & 1.00000 \end{pmatrix}$$

If we calculate the eigenvalues of $\mathbb{C}$, we obtain $\lambda_1 = 1.378$, $\lambda_2 = 1.095$ and $\lambda_3 = 0.527$. The problem comes from the fact that $Q\Lambda Q^\top$ generates a covariance matrix with specified eigenvalues, but never a correlation matrix even if the sum of the eigenvalues is equal to n.

Bendel and Mickey (1978) proposed an algorithm to transform the matrix Σ into a correlation matrix $\mathbb{C}$ with specified eigenvalues $\lambda_1, \ldots, \lambda_n$. The main idea is to perform Givens rotations[9]. Let $G_{c,s}(i,j)$ be the Givens matrix:

$$G_{c,s}(i,j) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & & & & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & & & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

such that the (i,j) element[10] of $G_{c,s}(i,j)^\top \Sigma G_{c,s}(i,j)$ is equal to 1. By performing n successive Givens transformations $\Sigma \leftarrow G_{c,s}(i,j)^\top \Sigma G_{c,s}(i,j)$, we obtain a correlation matrix $\mathbb{C}$

[9] A Givens rotation is a rotation in the plane spanned by two coordinates axes (Golub and Van Loan, 2013). Because Givens matrices are orthogonal, eigenvalues are not changed.

[10] We have $i < j$ and $\Sigma_{i,i} < 1 < \Sigma_{j,j}$ (or $\Sigma_{i,i} > 1 > \Sigma_{j,j}$).

with eigenvalues $\lambda_1, \ldots, \lambda_n$. The previous algorithm has been extensively studied by Davies and Higham (2000), who showed that:

$$c = \frac{1}{\sqrt{1+t^2}} \text{ and } s = c \cdot t$$

where:

$$t = \frac{\Sigma_{i,j} + \sqrt{\Sigma_{i,j}^2 - (\Sigma_{i,i} - 1)(\Sigma_{j,j} - 1)}}{(\Sigma_{j,j} - 1)}$$

To show the difference between the Bendel-Mickey algorithm and the previous covariance algorithm, we simulate a correlation matrix of dimension 20 with specified eigenvalues and the two algorithms. In Figure 13.13, we compare the eigenvalues calculated with the simulated correlation matrices and compare them with the specified eigenvalues. We verify that the Bendel-Mickey algorithm preserves the spectrum, which is not the case of the covariance algorithm[11].

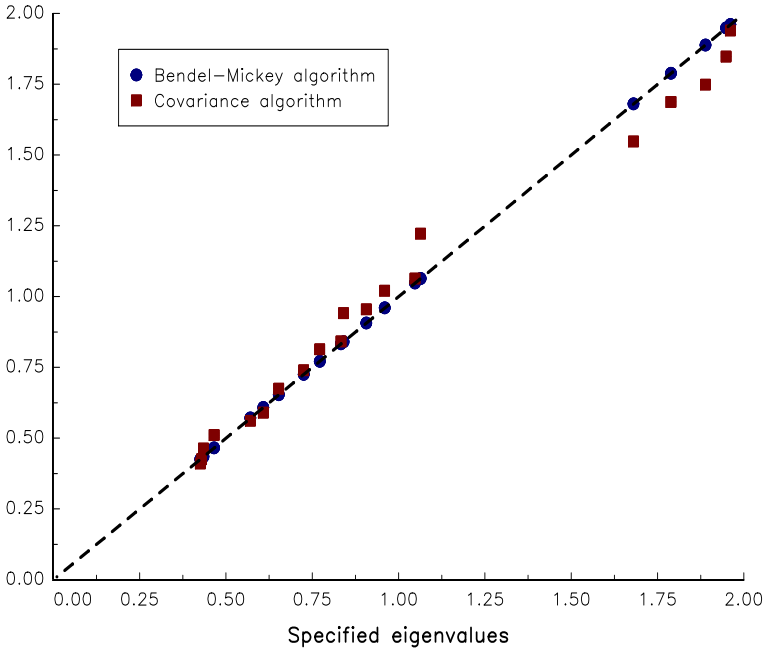


FIGURE 13.13: Comparison of the Bendel-Mickey and covariance algorithms

We study the correlation risk of the basket option, whose payoff is equal to:

$$G = (S_1(T) - S_2(T) + S_3(T) - S_4(T) - K)_+ \cdot \mathbb{1}\{S_5(T) > L\}$$

where $S_i(T)$ is the price of the i^{th} asset at the maturity T. We assume that the dynamics of the asset prices follows a Black-Scholes model:

$$S_i(T) = S_i(0) \cdot \exp\left(\left(r - \frac{1}{2}\sigma_i^2\right)T + \sigma_i(W_i(T) - W_i(0))\right)$$

[11] However, we notice that the eigenvalues are close.

where r is the risk-free rate, σ_i is the asset volatility and W_i is a Brownian process. We also assume that the Brownian processes are correlated:

$$\mathbb{E}\left[W_i\left(t\right)W_j\left(t\right)\right] = \rho_{i,j}\, t$$

To calculate the price of the basket option, we simulate the terminal value of $S_i\left(T\right)$ and average the simulated payoff G_s:

$$P = \mathbb{E}\left[e^{-rT}G\right] \approx \frac{1}{n_S}\sum_{s=1}^{n_S} e^{-rT}G_s$$

where n_S is the number of simulations. We use the following values: $S_i\left(0\right) = 100$, $r = 5\%$, $\sigma_i = 20\%$, $T = 0.25$, $K = 5$ and $L = 105$. We consider that it is difficult to estimate the correlation matrix and assume that it is unstable. In this case, we have to find an upper bound for P in order to take into account this correlation risk. Generally, we price the option by using a constant correlation matrix $\mathbb{C}_5\left(\rho\right)$ and takes the supremum:

$$P^+ = \sup_{\rho} P\left(\mathbb{C}_5\left(\rho\right)\right)$$

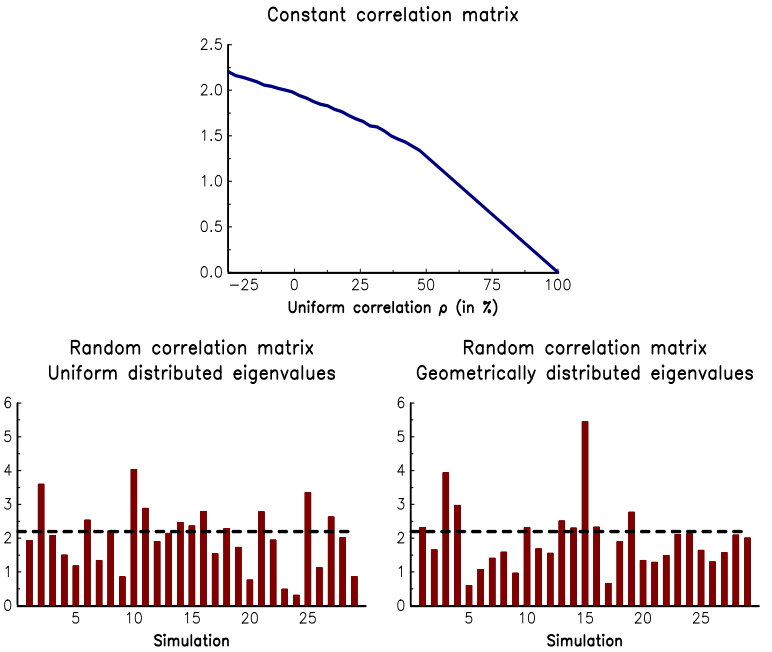


FIGURE 13.14: Price of the basket option

In the top panel in Figure 13.14, we report the price of the basket option with respect to the uniform correlation ρ. We notice that the price is a decreasing function of ρ and reaches its maximum when the uniform correlation is -25%. Therefore, we could suppose that the upper bound is equal to \$2.20. However, if we consider random correlation matrices, we observe that this price is not conservative (see bottom panels in Figure 13.14). For instance,

we obtain a price equal to \$5.45 with the following correlation matrix:

$$\mathbb{C} = \begin{pmatrix} 1.0000 & -0.4682 & -0.3034 & -0.1774 & 0.1602 \\ -0.4682 & 1.0000 & -0.3297 & 0.1381 & -0.7272 \\ -0.3034 & -0.3297 & 1.0000 & -0.3273 & 0.6106 \\ -0.1774 & 0.1381 & -0.3273 & 1.0000 & -0.1442 \\ 0.1602 & -0.7272 & 0.6106 & -0.1442 & 1.0000 \end{pmatrix}$$

This matrix indicates the type of correlation risks we face when we want to hedge this basket option. Indeed, the correlation risk is maximum when the fifth asset (which activates the barrier) is positively correlated with the first and third assets and negatively correlated with the second and fourth assets.

In the Bendel-Mickey algorithm, we control the structure of the random correlation matrix by specifying the eigenvalues. In Finance, it can be not sufficient. For instance, we may want to simulate the matrix $\mathbb{C}$ such that is expected value is equal to a given correlation matrix $\mathbb{C}^\star$:

$$\mathbb{E}[\mathbb{C}] = \mathbb{C}^\star$$

Let A be a random symmetric matrix with zeros on the diagonal and mean $\mathbb{E}[A] = \mathbf{0}$. Marsaglia and Olkin (1984) showed that $\mathbb{C} = A + \mathbb{C}^\star$ is a random correlation matrix with $\mathbb{E}[A + \mathbb{C}^\star] = \mathbb{C}^\star$ if the 2-norm of A is less than the smallest eigenvalue $\lambda_{\min}$ of $\mathbb{C}^\star$. There is a variety of algorithms that uses this result. For instance, Marsaglia and Olkin (1984) proposed to generate a random correlation matrix R with specified eigenvalues in the interval $[1 - \lambda_{\min}, 1 + \lambda_{\min}]$ and to take $\mathbb{C} = (R - I_n) + \mathbb{C}^\star$.

13.1.4.3 Wishart matrices

To generate a random Wishart matrix S, we simulate n independent Gaussian random vectors $X_i \sim \mathcal{N}_p(0, \Sigma)$ and form the $n \times p$ matrix X by concatenating the random vectors in the following way:

$$X = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix}$$

Then, we have $S = X^\top X$. The simulation of an inverse Wishart matrix T is straightforward by applying the transformation method:

$$T = S^{-1}$$

13.2 Simulation of stochastic processes

We distinguish two types of time series models, those based on discrete-time stochastic processes and those based on continuous-time stochastic processes. Discrete-time models are easier to simulate, in particular when we consider time-homogeneous Markov processes. This is not the case of continuous-time models, which are generally approximated by a time-discretized process. In this case, the convergence of the discrete simulation to the continuous solution depends on the approximation scheme.

13.2.1 Discrete-time stochastic processes

13.2.1.1 Correlated Markov chains

We consider a vector $\mathfrak{R} = (\mathfrak{R}_1, \ldots, \mathfrak{R}_M)$ of time-homogeneous Markov chains, whose transition probability matrix is P. The simulation of the Markov chain $\mathfrak{R}_m$ is given by the following algorithm:

1. we assume the initial position of the Markov chain:

$$\mathfrak{R}_m(0) = i_0$$

2. let u be a random number; we simulate the new position of $\mathfrak{R}_m$ by inverting the conditional probability distribution, whose elements are:

$$\Pr(\mathfrak{R}_m(n+1) = i_{n+1} \mid \mathfrak{R}_m(n) = i_n) = p_{i_n,i_{n+1}} = e_{i_n}^\top P e_{i_{n+1}}$$

   we have:

$$i_{n+1} = \left\{ k : \sum_{j=1}^{k-1} p_{i_n,j} < u \leq \sum_{j=1}^{k} p_{i_n,j} \right\}$$

3. we go back to step 2.

We now assume that the dependence between the Markov chains $(\mathfrak{R}_1, \ldots, \mathfrak{R}_M)$ is given by a copula function $\mathbf{C}$, implying that the Markov chains are correlated. The algorithm becomes:

1. we assume the initial position of the Markov chains:

$$\mathfrak{R}_m(0) = i_{m,0}$$

2. let $(u_1, \ldots, u_M)$ be a vector of correlated uniform random numbers such that:

$$(u_1, \ldots, u_M) \sim \mathbf{C}$$

3. for each Markov chain m, we simulate the new position of $\mathfrak{R}_m$ by inverting the conditional probability distribution; we have:

$$i_{m,n+1} = \left\{ k : \sum_{j=1}^{k-1} p_{i_{m,n},j} < u_m \leq \sum_{j=1}^{k} p_{i_{m,n},j} \right\}$$

   and $\mathfrak{R}_m(n+1) = i_{m,n+1}$.

4. we go back to step 2.

We consider four corporate firms, whose initial credit rating is AAA, BBB, B and CCC. We assume that the rating of each company is a Markov chain $\mathfrak{R}_m$ described by the credit migration matrix given on page 208. We also assume that the dependence of the credit ratings $(\mathfrak{R}_1, \mathfrak{R}_2, \mathfrak{R}_3, \mathfrak{R}_4)$ is a Normal copula with the following matrix of parameters:

$$\rho_1 = \begin{pmatrix} 1.00 & & & \\ 0.25 & 1.00 & & \\ 0.75 & 0.50 & 1.00 & \\ 0.50 & 0.25 & 0.75 & 1.00 \end{pmatrix}$$

In Figures 13.15 and 13.17, we report 10 simulated paths of the ratings for the next 30 years. We verify that the default rating is an absorbing state. Suppose now that the parameter matrix of the Normal copula is equal to:

$$\rho_2 = \begin{pmatrix} 1.00 & & & \\ -0.25 & 1.00 & & \\ -0.75 & 0.50 & 1.00 & \\ -0.50 & 0.25 & 0.75 & 1.00 \end{pmatrix}$$

Using this correlation matrix ρ_2 instead of the previous matrix ρ_1, we obtain the results given in Figures 13.16 and 13.18. If we compare Figures 13.15 and 13.16 (or Figures 13.17 and 13.18), which are based on the same uniform random numbers, we notice that the simulated paths are not the same. The reason comes from the negative correlation between the credit rating of the first company and the other credit ratings.

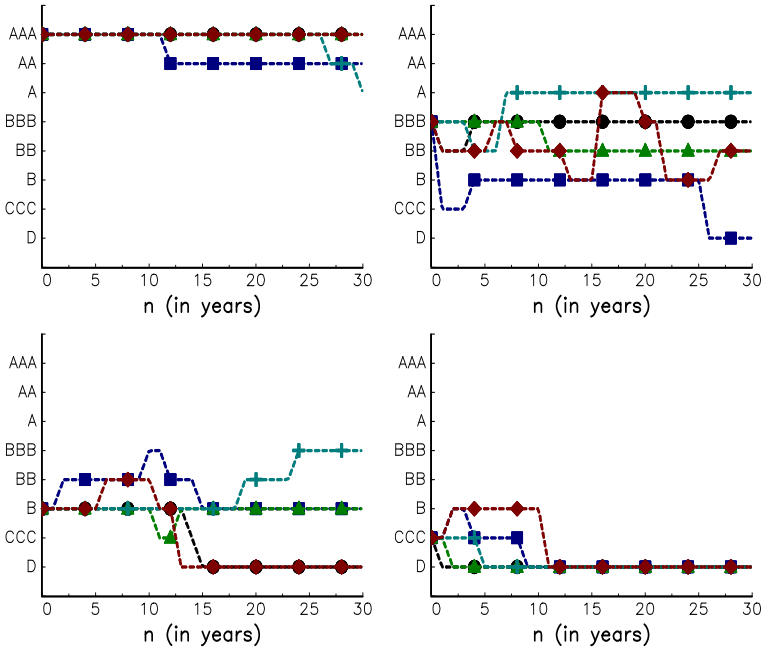


FIGURE 13.15: Simulation of rating dynamics (correlation matrix ρ_1)

13.2.1.2 Time series

A state space model (SSM) is defined by a measurement equation and a transition equation. The measurement equation describes the relationship between the observed variables y_t and the state vector α_t:

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t$$

whereas the transition equation gives the dynamics of the state variables:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t$$

The dimension of the vectors y_t and α_t is respectively $n \times 1$ and $m \times 1$. Z_t is a $n \times m$ matrix, d_t is a $n \times 1$ vector, T_t is a $m \times m$ matrix, c_t is a $m \times 1$ vector and R_t is a $m \times p$ matrix. $\varepsilon_t \sim \mathcal{N}_n(\mathbf{0}, H_t)$ and $\eta_t \sim \mathcal{N}_p(\mathbf{0}, Q_t)$ are two independent white noise processes. By

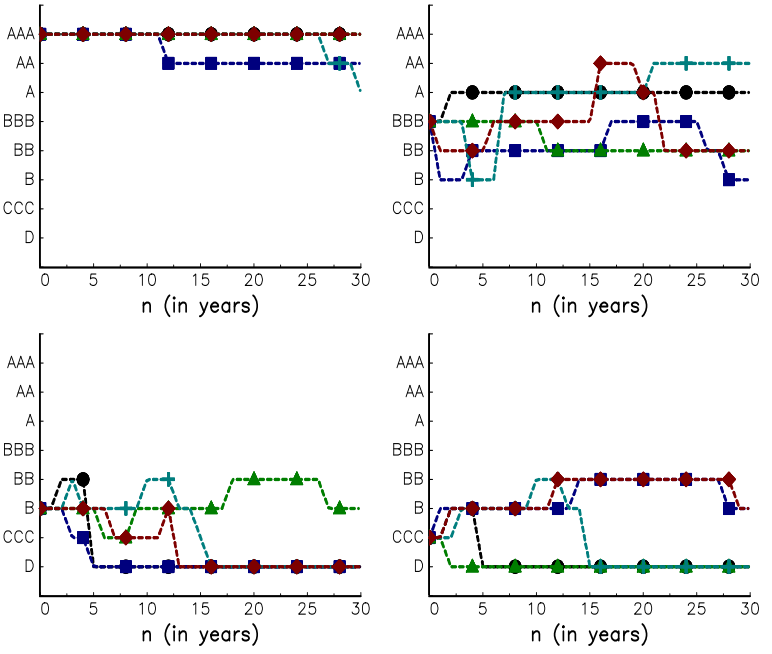


FIGURE 13.16: Simulation of rating dynamics (correlation matrix ρ_2)

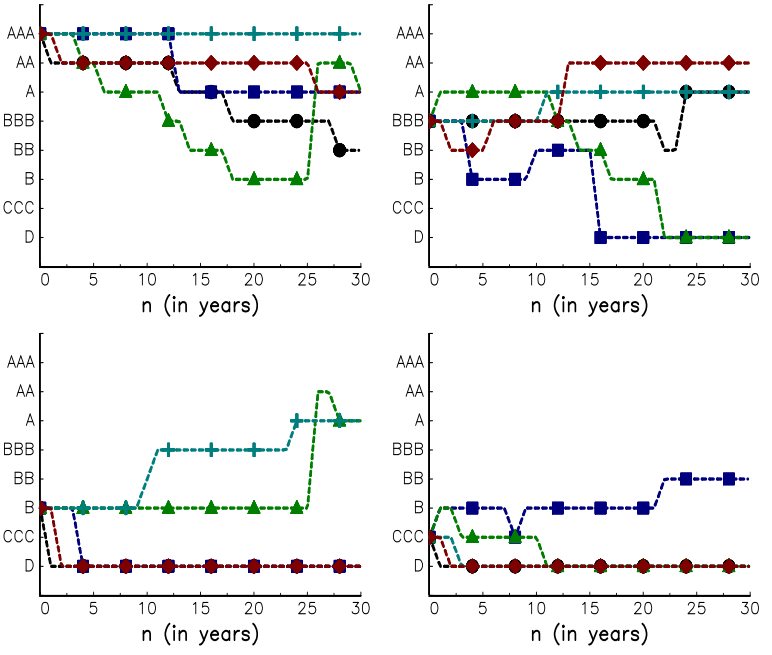


FIGURE 13.17: Simulation of rating dynamics (correlation matrix ρ_1)

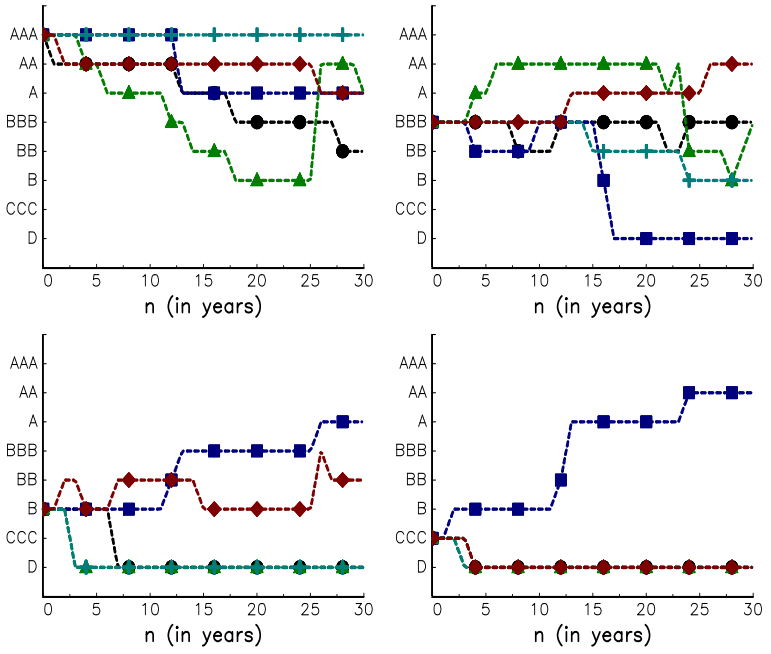


FIGURE 13.18: Simulation of rating dynamics (correlation matrix ρ_2)

construction, there is no special issue to simulate the Markov process α_t if we assume that the initial position is $\alpha_0 \sim \mathcal{N}_m(a_0, P_0)$. Indeed, we obtain the following algorithm:

1. we simulate the initial position:
$$\alpha_0 \sim \mathcal{N}_m(a_0, P_0)$$
2. we simulate the position of the state variable at time t:
$$\alpha_t \sim \mathcal{N}_m\left(T_t\alpha_{t-1} + c_t, R_tQ_tR_t^\top\right)$$
3. we simulate the space variable at time t:
$$y_t \sim \mathcal{N}_n(Z_t\alpha_t + d_t, H_t)$$
4. we go back to step 2.

Most of discrete-time stochastic processes are homogeneous, meaning that the parameters of the state space model are time-independent:

$$\begin{cases} y_t = Z\alpha_t + d + \varepsilon_t \\ \alpha_t = T\alpha_{t-1} + c + R\eta_t \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}_n(\mathbf{0}, H)$ and $\eta_t \sim \mathcal{N}_p(\mathbf{0}, Q)$. In this case, the stationary solution of the transition equation is $\alpha^\star \sim \mathcal{N}_m(a^\star, P^\star)$ where $a^\star$ is equal to $(I_m - T)^{-1} c$ and $P^\star$ satisfies the matrix equation[12]:

$$P^\star = TP^\star T^\top + RQR^\top$$

[12]We also have (Harvey, 1990):

$$\operatorname{vec}(P^\star) = (I_{m^2} - T \otimes T)^{-1} \operatorname{vec}\left(RQR^\top\right)$$

In practice, we generally use the stationary solution to initialize the state space model: $\alpha_0 \sim \mathcal{N}_m(a^\star, P^\star)$.

State space models can be used to simulate structural models, AR and MA processes, vector error correction models, VAR processes, etc. For instance, a VARMA(p,q) model with K endogenous variables is defined by:

$$\left(I - \sum_{i=1}^{p} \Phi_i L^i\right) y_t = \left(I - \sum_{i=1}^{q} \Theta_i L^i\right) u_t$$

where u_t is a multidimensional white noise process. Let α_t be the vector process $(y_t, \ldots, y_{t-p+1}, u_t, \ldots, u_{t-q+1})$, whose dimension is $K(p+q)$. We have:

$$\alpha_t = T\alpha_{t-1} + Ru_t$$

where R is the $K(p+q) \times K$ matrix $\begin{bmatrix} I_K & \mathbf{0} & \cdots & \mathbf{0} & I_K & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}^\top$ and T is the $K(p+q) \times K(p+q)$ matrix:

$$T = \begin{bmatrix}
\Phi_1 & \cdots & \Phi_{p-1} & \Phi_p & \Theta_1 & \cdots & \Theta_{q-1} & \Theta_q \\
I_K & & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \\
 & \ddots & & \vdots & \vdots & & \mathbf{0} & \\
\mathbf{0} & \cdots & I_K & \mathbf{0} & \mathbf{0} & & & \\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
 & & & \mathbf{0} & I_K & & \mathbf{0} & \mathbf{0} \\
 & \mathbf{0} & & \vdots & & \ddots & & \vdots \\
 & & & \mathbf{0} & \mathbf{0} & & I_K & \mathbf{0}
\end{bmatrix}$$

We notice that:

$$y_t = Z\alpha_t$$

where Z is the $K \times K(p+q)$ matrix $\begin{bmatrix} I_K & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$. We finally obtain the following SSM representation[13]:

$$\begin{cases} y_t = Z\alpha_t \\ \alpha_t = T\alpha_{t-1} + Ru_t \end{cases}$$

13.2.2 Univariate continuous-time processes

13.2.2.1 Brownian motion

A Brownian motion (or a Wiener process) is a stochastic process $W(t)$, whose increments are stationary and independent:

$$W(t) - W(s) \sim \mathcal{N}(0, t-s)$$

Therefore, we have:

$$\begin{cases} W(0) = 0 \\ W(t) = W(s) + \epsilon(s,t) \end{cases}$$

where $\epsilon(s,t) \sim \mathcal{N}(0, t-s)$ are *iid* random variables. This representation is helpful to simulate $W(t)$ at different dates $t_1, t_2, \ldots$ If we note W_m the numerical realization of $W(t_m)$, we have:

$$W_{m+1} = W_m + \sqrt{t_{m+1} - t_m} \cdot \varepsilon_m$$

where $\varepsilon_m \sim \mathcal{N}(0,1)$ are *iid* random variables. In the case of fixed-interval times $t_{m+1} - t_m = h$, we obtain the recursion:

$$W_{m+1} = W_m + \sqrt{h} \cdot \varepsilon_m$$

[13] We have $H = \mathbf{0}$, $Q = \text{var}(u_t)$, $d = \mathbf{0}$ and $c = \mathbf{0}$.

13.2.2.2 Geometric Brownian motion

The geometric Brownian motion is described by the following stochastic differential equation:

$$\begin{cases} \mathrm{d}X(t) = \mu X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

Its solution is given by:

$$\begin{aligned} X(t) &= x_0 \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right) \\ &= g(W(t)) \end{aligned}$$

Therefore, simulating the geometric Brownian motion $X(t)$ can be done by applying the transform method to the process $W(t)$.

Another approach to simulate $X(t)$ consists in using the following formula:

$$X(t) = X(s) \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)(t-s) + \sigma\left(W(t) - W(s)\right)\right)$$

We have:

$$X_{m+1} = X_m \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)(t_{m+1} - t_m) + \sigma\sqrt{t_{m+1} - t_m} \cdot \varepsilon_m\right)$$

where $X_m = X(t_m)$ and $\varepsilon_m \sim \mathcal{N}(0,1)$ are *iid* random variables. If we consider fixed-interval times, the numerical realization becomes:

$$X_{m+1} = X_m \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)h + \sigma\sqrt{h} \cdot \varepsilon_m\right) \tag{13.1}$$

Example 145 *In Figure 13.19, we simulate* 10 *paths of the geometric Brownian motion when* μ *and* σ *are equal to* 10% *and* 20%*. We consider a period of one year with a financial calendar of 260 trading days. This means that we use a fixed-interval time with* $h = 1/260$*. In finance, we use the convention that* $t = 1$ *corresponds to one year, which implies that* μ *and* σ *are respectively the annual expected return and volatility.*

13.2.2.3 Ornstein-Uhlenbeck process

The stochastic differential equation of the Ornstein-Uhlenbeck process is:

$$\begin{cases} \mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

We can show that the solution of the SDE is:

$$X(t) = x_0 e^{-at} + b\left(1 - e^{-at}\right) + \sigma\int_0^t e^{a(\theta - t)}\,\mathrm{d}W(\theta)$$

We also have:

$$X(t) = X(s)\,e^{-a(t-s)} + b\left(1 - e^{-a(t-s)}\right) + \sigma\int_s^t e^{a(\theta - t)}\,\mathrm{d}W(\theta)$$

where $\int_s^t e^{a(\theta-t)}\,\mathrm{d}W(\theta)$ is a Gaussian white noise process with variance $\left(1 - e^{-2a(t-s)}\right)/(2a)$. If we consider fixed-interval times, we obtain the following simulation scheme:

$$X_{m+1} = X_m e^{-ah} + b\left(1 - e^{-ah}\right) + \sigma\sqrt{\frac{1 - e^{-2ah}}{2a}} \cdot \varepsilon_m$$

where $\varepsilon_m \sim \mathcal{N}(0,1)$ are *iid* random variables.

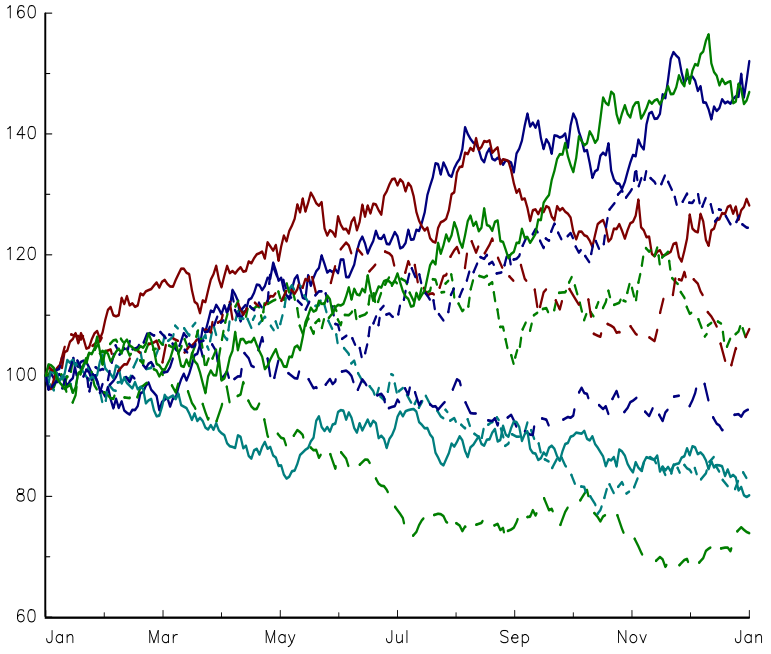


FIGURE 13.19: Simulation of the geometric Brownian motion

Example 146 *We assume that* $a = 2$, $b = 10\%$ *and* $\sigma = 1.5\%$. *The initial position of the process is* $x_0 = 5\%$. *We simulate* X_t *for two years and report the generated paths in Figure 13.20.*

13.2.2.4 Stochastic differential equations without an explicit solution

In the case of the geometric Brownian motion or the Ornstein-Uhlenbeck process, we obtain an exact scheme for simulating these processes, because we know the analytical solution. In many cases, the solution is not known and can only be simulated using approximation schemes. Let $X(t)$ be the solution of the following SDE:

$$\begin{cases} \mathrm{d}X(t) = \mu(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

The simplest numerical method for simulating $X(t)$ is the Euler-Maruyama scheme, which uses the following approximation:

$$X(t) - X(s) \approx \mu(t, X(s)) \cdot (t - s) + \sigma(t, X(s)) \cdot (W(t) - W(s))$$

If we consider fixed-interval times, the Euler-Maruyama scheme becomes:

$$X_{m+1} = X_m + \mu(t_m, X_m)\, h + \sigma(t_m, X_m)\, \sqrt{h} \cdot \varepsilon_m$$

where $\varepsilon_m \sim \mathcal{N}(0, 1)$ are *iid* random variables.

Remark 157 *The accuracy of numerical approximations is evaluated with the strong order of convergence. Let* $X_m^{(h)}$ *be the numerical solution of* $X(t_m)$ *computed with the constant stepwise* h. *A numerical scheme is said to converge strongly to the exact solution if we have:*

$$\lim_{h \to 0} \mathbb{E}\left[\left|X_m^{(h)} - X(t_m)\right|\right] = 0$$

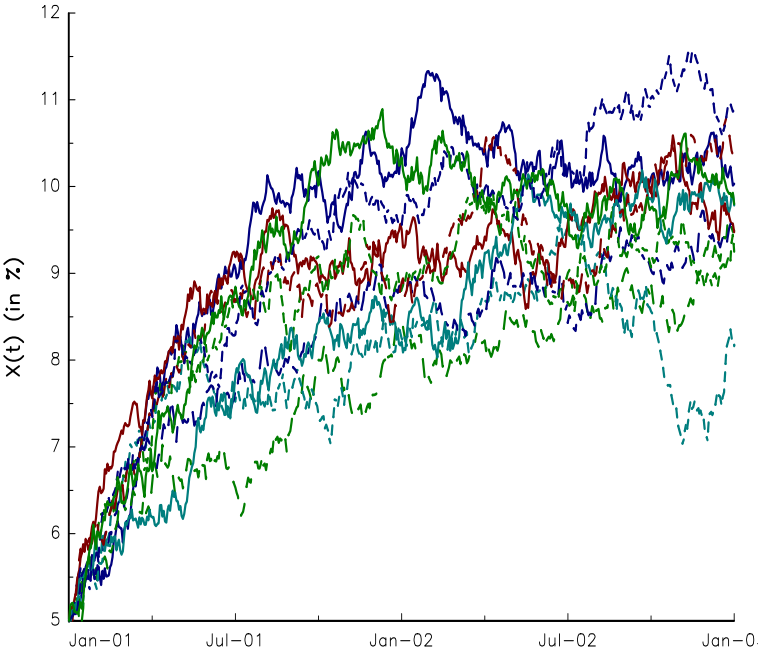


FIGURE 13.20: Simulation of the Ornstein-Uhlenbeck process

for a time t_m. The order of convergence is given by the convergence rate p:

$$\mathbb{E}\left[\left|X_m^{(h)} - X\left(t_m\right)\right|\right] \leq C \cdot h^p$$

where C is a constant and h is sufficiently small ($h \leq h_0$). In the case of the Euler-Maruyama method, the strong order of convergence is 0.5.

Example 147 *For modeling short-term interest rates, Chan et al. (1992) consider the following SDE:*

$$\mathrm{d}X\left(t\right) = \left(\alpha + \beta X\left(t\right)\right)\mathrm{d}t + \sigma X\left(t\right)^{\gamma}\mathrm{d}W\left(t\right) \tag{13.2}$$

We deduce that the fixed-interval Euler-Maruyama scheme is:

$$X_{m+1} = X_m + \left(\alpha + \beta X_m\right)h + \sigma X_m^{\gamma}\sqrt{h}\cdot\varepsilon_m$$

Kloeden and Platen (1992) provided many other approximation schemes, based on Itô-Taylor expansions of the SDE. For instance, the fixed-interval Milstein scheme is:

$$\begin{aligned} X_{m+1} &= X_m + \mu\left(t_m, X_m\right)h + \sigma\left(t_m, X_m\right)\sqrt{h}\cdot\varepsilon_m + \\ &\quad \frac{1}{2}\sigma\left(t_m, X_m\right)\partial_x\sigma\left(t_m, X_m\right)h\left(\varepsilon_m^2 - 1\right) \end{aligned} \tag{13.3}$$

The strong order for the Milstein method is equal to 1, which is better than the Euler-Maruyama method. In terms of implementation, these two approximation schemes remain simple, compared to other Taylor schemes that converge more quickly, but generally use correlated random variables and high order derivatives of the functions $\mu\left(t,x\right)$ and $\sigma\left(t,x\right)$. This is why Euler-Maruyama and Milstein schemes are the most frequent methods used in practice[14].

[14] For instance, one of the most famous methods is the strong order 1.5 Taylor scheme proposed by Platen and Wagner (1982). It requires the second derivatives $\partial_x^2\mu\left(t,x\right)$ and $\partial_x^2\sigma\left(t,x\right)$, and an additional random

If we consider the geometric Brownian motion, the Euler-Maruyama scheme is:

$$X_{m+1} = X_m + \mu X_m h + \sigma X_m \sqrt{h} \cdot \varepsilon_m$$

whereas the Milstein scheme is:

$$\begin{aligned} X_{m+1} &= X_m + \mu X_m h + \sigma X_m \sqrt{h} \cdot \varepsilon_m + \frac{1}{2}\sigma^2 X_m h \left(\varepsilon_m^2 - 1\right) \\ &= X_m + \left(\mu - \frac{1}{2}\sigma^2\right) X_m h + \sigma X_m \sqrt{h} \left(1 + \frac{1}{2}\sigma\sqrt{h}\varepsilon_m\right) \varepsilon_m \end{aligned}$$

It follows that the Milstein scheme operates two corrections for simulating the GBM process:

- the first correction concerns the drift, which is now correct;
- the second correction applies to the diffusion term, which increases if it is positive and decreases if it is negative.

In order to illustrate the differences between these two schemes, we compare them using the same random numbers. A simulation is provided in Figure 13.21 in the case where $\mu = 10\%$ and $\sigma = 50\%$. With a monthly discretization, we notice that the Milstein scheme produces a better solution than the Euler-Maruyama scheme.

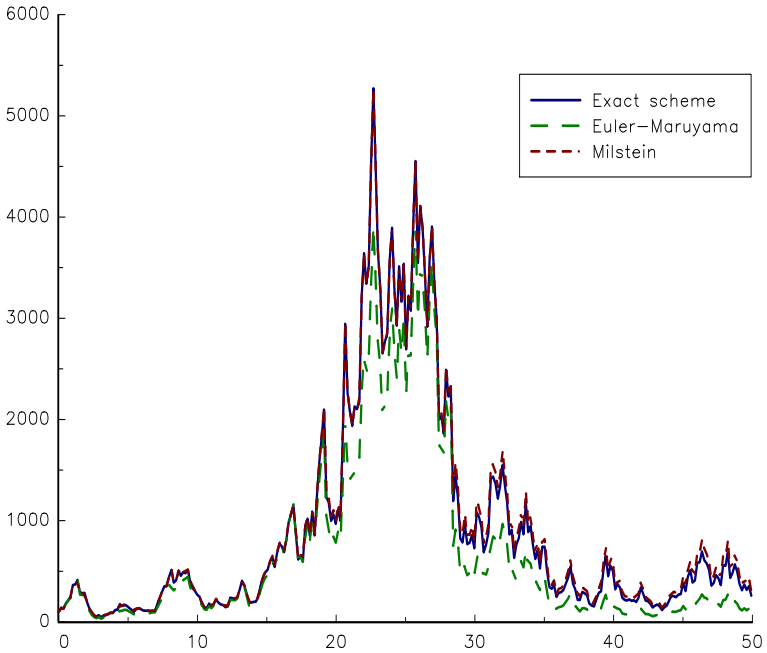


FIGURE 13.21: Comparison of exact, Euler-Maruyama and Milstein schemes (monthly discretization)

When we don't know the analytical solution of $X(t)$, it is natural to simulate the numerical solution of $X(t)$ using Euler-Maruyama and Milstein schemes. However, it may

variable correlated with the increments of the Brownian motion. Even if this scheme is interesting to study from a theoretical point of view, it is never used by practitioners because it is time-consuming.

be sometimes more efficient to find the numerical solution of $Y(t) = f(t, X(t))$ instead of $X(t)$ itself, in particular when $Y(t)$ is more regular than $X(t)$. By Itô's lemma, we have:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\partial_t f(t,X) + \mu(t,X)\,\partial_x f(t,X) + \frac{1}{2}\sigma^2(t,X)\,\partial_x^2 f(t,X)\right)\mathrm{d}t + \\ &\quad \sigma(t,X)\,\partial_x f(t,X)\,\mathrm{d}W(t) \end{aligned}$$

By using the inverse function $X(t) = f^{-1}(t, Y(t))$, we obtain:

$$\mathrm{d}Y(t) = \mu'(t,Y)\,\mathrm{d}t + \sigma'(t,Y)\,\mathrm{d}W(t)$$

where $\mu'(t,Y)$ and $\sigma'(t,Y)$ are functions of $\mu(t,X)$, $\sigma(t,X)$ and $f(t,X)$. We can then simulate the solution of $Y(t)$ using an approximation scheme and deduce the numerical solution of $X(t)$ by applying the transformation method:

$$X_m = f^{-1}(t_m, Y_m)$$

Let us consider the geometric Brownian motion $X(t)$. The solution of $Y(t) = \ln X(t)$ is equal to:

$$\mathrm{d}Y(t) = \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma\,\mathrm{d}W(t)$$

We deduce that the Euler-Maruyama (or Milstein[15]) scheme with fixed-interval times is:

$$Y_{m+1} = Y_m + \left(\mu - \frac{1}{2}\sigma^2\right)h + \sigma\sqrt{h}\cdot\varepsilon_m$$

It follows that:

$$\ln X_{m+1} = \ln X_m + \left(\mu - \frac{1}{2}\sigma^2\right)h + \sigma\sqrt{h}\cdot\varepsilon_m \tag{13.4}$$

We conclude that this numerical solution is the exact solution (13.1) of the geometric Brownian motion.

The previous application is not interesting, because we know the analytical solution. The approach is more adapted for stochastic differential equations without explicit solutions, for example the Cox-Ingersoll-Ross process:

$$\mathrm{d}X(t) = (\alpha + \beta X(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\,\mathrm{d}W(t) \tag{13.5}$$

This process is a special case of the CKLS process (13.2) with $\gamma = 1/2$ and can be viewed as an Ornstein-Uhlenbeck process[16] with a reflection at $X(t) = 0$. Using the transformation $Y(t) = \sqrt{X(t)}$, we obtain the following SDE[17]:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{1}{2}\frac{(\alpha + \beta X(t))}{\sqrt{X(t)}} - \frac{1}{8}\frac{\sigma^2 X(t)}{X(t)^{3/2}}\right)\mathrm{d}t + \frac{1}{2}\frac{\sigma\sqrt{X(t)}}{\sqrt{X(t)}}\,\mathrm{d}W(t) \\ &= \frac{1}{2Y(t)}\left(\alpha + \beta Y^2(t) - \frac{1}{4}\sigma^2\right)\mathrm{d}t + \frac{1}{2}\sigma\,\mathrm{d}W(t) \end{aligned}$$

[15] Because $\partial_y \sigma'(t,Y) = 0$.

[16] The drift can be written as $\alpha + \beta X(t) = -\beta\left(-\alpha\beta^{-1} - X(t)\right)$. We deduce that $a = -\beta$ and $b = -\alpha\beta^{-1}$.

[17] We have $\partial_x f(t,x) = \frac{1}{2}x^{-1/2}$ and $\partial_x^2 f(t,x) = -\frac{1}{4}x^{-3/2}$.

We deduce that the Euler-Maruyama scheme of $Y(t)$ is:

$$Y_{m+1} = Y_m + \frac{1}{2Y_m}\left(\alpha + \beta Y_m^2 - \frac{1}{4}\sigma^2\right)h + \frac{1}{2}\sigma\sqrt{h}\cdot\varepsilon_m$$

It follows that:

$$X_{m+1} = \left(\sqrt{X_m} + \frac{1}{2\sqrt{X_m}}\left(\alpha + \beta X_m - \frac{1}{4}\sigma^2\right)h + \frac{1}{2}\sigma\sqrt{h}\cdot\varepsilon_m\right)^2$$

We can show that this approximation is better than the Euler-Maruyama or Milstein approximation directly applied to the SDE (13.5).

Remark 158 *Generally, we choose the Lamperti transform $Y(t) = f(X(t))$ in order to obtain a constant diffusion term ($\partial_y \sigma'(t,y) = 0$). This implies that:*

$$f(x) = c\int_a^x \frac{1}{\sigma(t,u)}\,\mathrm{d}u$$

Because we have $\partial_x f(t,x) = c/\sigma(t,x)$ and $\partial_x^2 f(t,x) = -c\partial_x\sigma(t,x)/\sigma^2(t,x)$, we obtain:

$$\mathrm{d}Y(t) = c\left(\frac{\mu(t,X)}{\sigma(t,X)} - \frac{\partial_x\sigma(t,X)}{2}\right)\mathrm{d}t + c\,\mathrm{d}W(t)$$

In this case, the Euler-Maruyama scheme coincides with the Milstein scheme. Most of the time, the approximation $X_m = f^{-1}(Y_m)$ gives better results than those obtained with a Milstein method applied to the process $X(t)$.

13.2.2.5 Poisson processes

We have seen that simulating a Poisson process $N(t)$ with constant intensity λ is straightforward, because the inter-arrival times are independent and exponentially distributed with parameter λ. Let t_m be the time when the m^{th} event occurs. The numerical algorithm is then:

1. we set $t_0 = 0$ and $N(t_0) = 0$;
2. we generate a uniform random variate u and calculate the random variate $e \sim \mathcal{E}(\lambda)$ with the formula:
$$e = -\frac{\ln u}{\lambda}$$
3. we update the Poisson process with:
$$t_{m+1} \leftarrow t_m + e \qquad \text{and} \qquad N(t_{m+1}) \leftarrow N(t_m) + 1$$
4. we go back to step 2.

We can also use this algorithm to simulate mixed Poisson process (MPP), which are defined as Poisson process with stochastic intensity Λ. In this case, the algorithm is initialized with a realization λ of the random intensity Λ. On the contrary, this method is not valid in the case of non-homogenous Poisson process (NHPP), where the intensity $\lambda(t)$ varies with time[18]. However, we can show that the inter-arrival times remain independent and exponentially distributed with:

$$\Pr\{T_1 > t\} = \exp(-\Lambda(t))$$

[18]Indeed, we don't know the value of the intensity to use at the second step of the algorithm.

where T_1 is the duration of the first event and $\Lambda(t)$ is the integrated intensity function:

$$\Lambda(t) = \int_0^t \lambda(s)\, \mathrm{d}s$$

It follows that:

$$\Pr\left\{T_1 > \Lambda^{-1}(t)\right\} = \exp(-t) \Leftrightarrow \Pr\left\{\Lambda(T_1) > t\right\} = \exp(-t)$$

We deduce that if $\{t_1, t_2, \ldots, t_M\}$ are the occurrence times of the NHPP of intensity $\lambda(t)$, then $\{\Lambda(t_1), \Lambda(t_2), \ldots, \Lambda(t_M)\}$ are the occurrence times of the homogeneous Poisson process (HPP) of intensity one. Therefore, the algorithm is:

1. we simulate t'_m the time arrivals of the homogeneous Poisson process with intensity $\lambda = 1$;
2. we apply the transform $t_m = \Lambda^{-1}(t'_m)$.

To implement this algorithm, we need to compute the inverse function $\Lambda^{-1}(t)$. When there is no analytical expression, this algorithm may be time-consuming, in particular when $\Lambda(t)$ is calculated with a method of numerical integration. Another approach consists in using the acceptance-rejection algorithm for simulating the NHPP over the period $[0, T]$:

1. we set $\lambda^+ = \max_{t \leq T} \lambda(t)$, $t = 0$, $t_0 = 0$ and $N(t_0) = 0$;
2. we generate a uniform random variate u and calculate the random variate $e \sim \mathcal{E}(\lambda^+)$ with the formula:
$$e = -\frac{\ln u}{\lambda^+}$$
3. we calculate $t = t + e$;
4. if $t > T$, we stop the algorithm;
5. we generate a uniform random variable v; if $v \leq \lambda(t)/\lambda^+$, then we accept the arrival time:
$$t_{m+1} \leftarrow t \qquad \text{and} \qquad N(t_{m+1}) \leftarrow N(t_m) + 1$$
else we reject it;
6. we go back to step 2.

In Figure 13.22, we simulate a non-homogenous Poisson process with the following intensity function:

$$\lambda(t) = 90 + 80 \cdot \sin\left(\frac{6\pi}{5} \cdot t\right)$$

Since $\lambda(t)$ is a cyclical function and $\lambda(t) \in [10, 170]$, the intensity function can vary very quickly. In the bottom/right panel, we draw the histogram of arrival process for the interval $[t, t + \mathrm{d}t]$ and compare it with the expected arrival frequency, which is equal to:

$$\mathbb{E}\left[N(t + \mathrm{d}t) - N(t)\right] = \int_t^{t+\mathrm{d}t} \lambda(s)\, \mathrm{d}s \approx \lambda(t)\, \mathrm{d}t$$

The compound Poisson process $Y(t)$ is defined by:

$$Y(t) = \sum_{i=1}^{N(t)} X_i$$

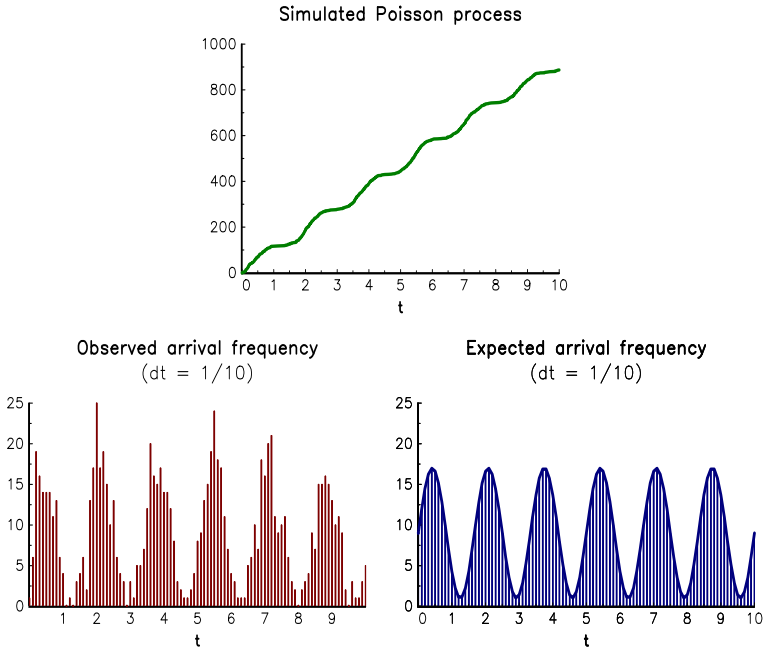


FIGURE 13.22: Simulation of a non-homogenous Poisson process with cyclical intensity

where $N(t)$ is a Poisson process of intensity λ and $\{X_i\}_{i\geq 1}$ is a sequence of *iid* random variables with distribution function $\mathbf{F}$. This process is a generalization of the Poisson process by assuming that jump sizes are not equal to one, but are random. Let $\{t_1, t_2, \ldots, t_M\}$ be the arrival times of the Poisson process. We have:

$$Y(t) = Y(t_m) \qquad \text{if } t \in [t_m, t_{m+1}[$$

and:

$$Y(t_{m+1}) = Y(t_m) + X_{m+1}$$

where X_{m+1} is generated from the distribution function $\mathbf{F}$. Another method to simulate $Y(t)$ is to use the following property: conditionally to $N(T) = n$, the arrival times $\{t_1, t_2, \ldots, t_n\}$ of the Poisson process on the interval $[0, T]$ are distributed as n independent ordered uniform random variables. We deduce this algorithm:

1. we simulate the number n of jumps on the time interval $[0, T]$ by generating a Poisson random variable with parameter λT;
2. we simulate n uniform random variates $(u_1, \ldots, u_n)$ and sort them[19]:

$$u_{1:n} \leq u_{2:n} \leq \ldots \leq u_{n:n}$$

3. we simulate n random variates $(x_1, \ldots, x_n)$ from the probability distribution $\mathbf{F}$;
4. we finally generate the compound Poisson process by:

$$Y(t) = \sum_{i=1}^{n} \mathbb{1}\left\{u_{i:n} \leq \frac{t}{T}\right\} \cdot x_i$$

[19]The arrival times are given by the formula: $t_m = T \cdot u_{m:n}$.

13.2.2.6 Jump-diffusion processes

A jump-diffusion model is a process, which is generally defined by:

$$\mathrm{d}X(t) = \mu(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W(t) + \eta\left(t^-, X^-\right)\mathrm{d}J(t) \tag{13.6}$$

where $J(t)$ is a jump process. Between two jumps, $\mathrm{d}J(t)$ is equal to zero and the process $X(t)$ is continuous and evolves according to the SDE:

$$\mathrm{d}X(t) = \mu(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W(t)$$

When a jump occurs, the process is discontinuous and we have:

$$X(t) = X\left(t^-\right) + \eta\left(t^-, X\left(t^-\right)\right)\mathrm{d}J(t)$$

The jump process may be a Poisson process $N(t)$ with intensity λ or a compound Poisson process $Y(t) = \sum_{i=1}^{N(t)} Z_i$ where $\{Z_i\}_{i\geq 1}$ is a sequence of *iid* random variables with distribution $\mathbf{F}$.

In the case $J(t) = N(t)$, the Euler scheme is:

$$\begin{aligned} X(t_{m+1}) &= X(t_m) + \mu(t_m, X(t_m)) \cdot (t_{m+1} - t_m) + \\ &\quad \sigma(t_m, X(t_m)) \cdot (W(t_{m+1}) - W(t_m)) + \\ &\quad \eta(t_m, X(t_m)) \cdot (N(t_{m+1}) - N(t_m)) \end{aligned}$$

We finally obtain:

$$\begin{aligned} X_{m+1} &= X_m + \mu(t_m, X_m) \cdot (t_{m+1} - t_m) + \\ &\quad \sigma(t_m, X_m) \cdot \sqrt{t_{m+1} - t_m} \cdot \varepsilon_m + \eta(t_m, X_m) \cdot \xi_m \end{aligned} \tag{13.7}$$

where $\varepsilon_m \sim \mathcal{N}(0,1)$ and $\xi_m \sim \mathcal{P}(\lambda(t_{m+1} - t_m))$. We have:

$$\begin{aligned} \Pr\{\xi_m = 0\} &= e^{-\lambda(t_{m+1}-t_m)} \\ \Pr\{\xi_m = 1\} &= e^{-\lambda(t_{m+1}-t_m)}\lambda(t_{m+1} - t_m) \\ \Pr\{\xi_m \geq 2\} &= 1 - e^{-\lambda(t_{m+1}-t_m)}(1 + \lambda(t_{m+1} - t_m)) \end{aligned}$$

If we assume that the stepsize $t_{m+1} - t_m$ is small, we obtain $\Pr\{\xi_m = 0\} \approx 1 - \lambda(t_{m+1} - t_m)$, $\Pr\{\xi_m = 1\} \approx \lambda(t_{m+1} - t_m)$ and $\Pr\{\xi_m \geq 2\} \approx 0$. Therefore, we can generate ξ_m by:

$$\xi_m = \begin{cases} 1 & \text{if } \lambda(t_{m+1} - t_m) \leq u_m \\ 0 & \text{otherwise} \end{cases}$$

where u_m is a uniform random variate. Another way to simulate $X(t)$ is to first simulate the arrival times of the Poisson process. We denote these times by $\tau_1, \tau_2, \ldots, \tau_N$, and combine this grid with the initial grid $t_1, t_2, \ldots, t_M$. We then apply the Euler scheme (13.7) on the augmented grid, but we are now sure that we cannot have more than one jump between two discretization times. We illustrate this algorithm by considering the SDE:

$$\mathrm{d}X(t) = 0.15 \cdot X(t)\,\mathrm{d}t + 0.20 \cdot X(t)\,\mathrm{d}W(t) + \left(30 - 0.30 \cdot X\left(t^-\right)\right) \cdot \mathrm{d}J(t)$$

A simulated path is given in Figure 13.23, where the jumps are indicated by a dashed line.

In the case of the compound Poisson process $J(t) = Y(t)$, we can obtain explicit solutions for some processes. For instance, the model of Merton (1976) considers that the continuous part is a geometric Brownian motion:

$$\mathrm{d}X(t) = \mu X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}W(t) + X\left(t^-\right)\mathrm{d}J(t)$$

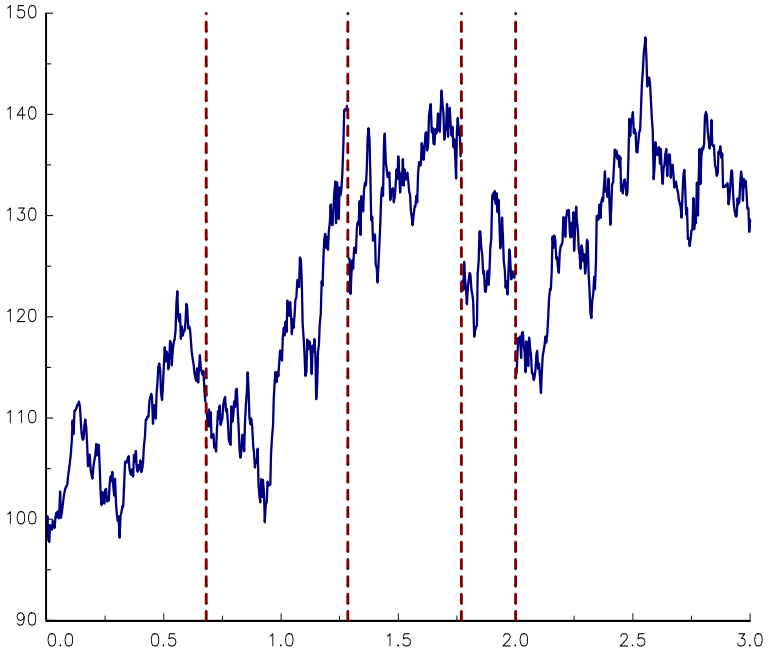


FIGURE 13.23: Simulation of a jump-diffusion process

If we assume that the i^{th} jump occurs at time t, we obtain[20]:

$$\begin{aligned} X(t) &= X(t^-) + X(t^-) Z_i \\ &= (1 + Z_i) X(t^-) \end{aligned}$$

We deduce that:

$$\begin{aligned} X(t) &= X(0) \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t) + J(t)\right) \\ &= X(0) e^{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)} \prod_{i=1}^{N(t)} (1 + Z_i) \end{aligned}$$

In the general case, the Euler scheme is:

$$\begin{aligned} X_{m+1} &= X_m + \mu(t_m, X_m) \cdot (t_{m+1} - t_m) + \\ &\quad \sigma(t_m, X_m) \cdot \sqrt{t_{m+1} - t_m} \cdot \varepsilon_m + \eta(t_m, X_m) \cdot \xi_m \end{aligned}$$

where $\varepsilon_m \sim \mathcal{N}(0,1)$ and $\xi_m = Y(t_{m+1}) - Y(t_m)$. As we have previously presented an algorithm to generate $Y(t)$, there is no difficulty to simulate $X(t)$.

13.2.2.7 Processes related to Brownian motion

We have previously shown how to simulate a stochastic differential equation by assuming the initial position of the random process. In finance, we also need to simulate stochastic processes with other constraints (Brownian bridge, Brownian meander) or statistics of the SDE (minimum, maximum, stopping time).

[20] We assume that $Z_i > -1$.

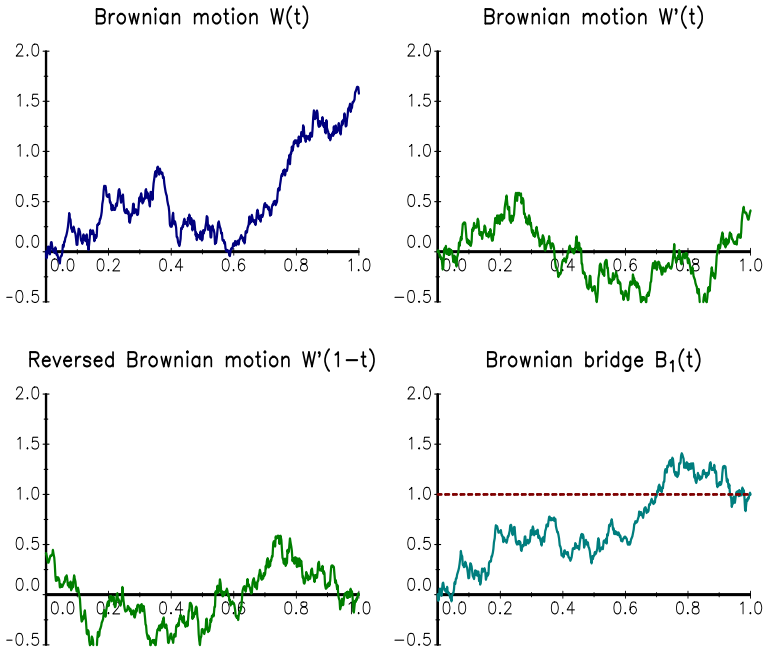


FIGURE 13.24: Simulation of the Brownian bridge $B_1(t)$ using the time reversibility property

A Brownian bridge $B_r(t)$ is a Brownian motion $W(t)$ such that $W(0) = 0$ and $W(1) = r$ (Revuz and Yor, 1999). For $t \in [0, 1]$, we have[21]:

$$B_r(t) = W(t) + (r - W(1)) \cdot t$$

Devroye (2010) noticed that:

$$W(1) = W(t) + (W(1) - W(t))$$

The time reversibility property of the Brownian motion implies that $W(1) - W(t) \overset{\mathcal{L}}{=} W(1-t)$. It follows that:

$$\begin{aligned} B_r(t) &= W(t) + (r - (W(t) + W'(1-t))) \cdot t \\ &= r \cdot t + (1-t) \cdot W(t) + t \cdot W'(1-t) \end{aligned}$$

Figure 13.24 illustrates the simulation of $B_1(t)$ by using two simulated paths $W(t)$ and $W'(t)$. We also notice that:

$$\begin{aligned} B_r(t) &= r \cdot t + (1-t) \cdot \sqrt{t} \cdot \varepsilon_1 + t \cdot \sqrt{1-t} \cdot \varepsilon_2 \\ &= r \cdot t + t \cdot \sqrt{t(1-t)} \cdot \varepsilon \end{aligned}$$

where ε_1, ε_2 and ε are standard Gaussian random variables. If we now assume that $s \leq t \leq u$, $W(s) = w_s$ and $W(u) = w_u$, the Brownian bridge becomes:

$$B(t) = \frac{(u-t) \cdot w_s + (t-s) \cdot w_u}{u-s} + \sqrt{\frac{(t-s) \cdot (u-t)}{u-s}} \cdot \varepsilon$$

[21] We verify that the increments of $B_r(t)$ are independent, $B_r(0) = 0$ and $B_r(1) = r$.

because of the scaling property of the Brownian motion[22]. If we consider the simulation of $B(t)$ for different values $t_m \in [s, u]$, we proceed by filling the path with the iterative algorithm:

1. we initialize the algorithm with $m = 1$;
2. we simulate the Brownian bridge $B(t_m)$ such that $B(s) = w_s$ and $B(u) = w_u$;
3. we set $s = t_m$ and $B(s) = B(t_m)$;
4. we go back to step 2.

In Figure 13.25, we report 5 simulations of the Brownian Bridge $B(t)$ such that $B(0) = 0$, $B(1) = 1$, $B(3) = 3$ and $B(5) = 2$.

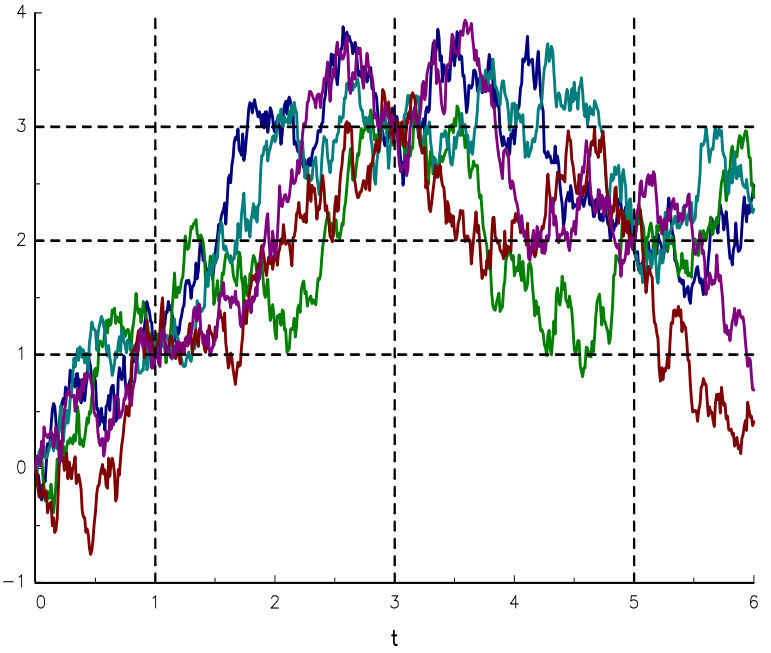


FIGURE 13.25: Simulation of the Brownian bridge $B(t)$

To simulate a process $X(t)$ with fixed values at times $\tau_1, \ldots, \tau_p$, we assume that we have an explicit solution $X(t) = g(W(t))$ implying that $W(t) = g^{-1}(X(t))$. Simulating a diffusion bridge $X(t)$ consists then in generating the Brownian bridge $B(t)$ such that $W(\tau_i) = g^{-1}(X(\tau_i))$, and applying the transformation $X(t) = g(B(t))$. For instance, if we consider the geometric Brownian motion, we have:

$$X(t) = g(W(t)) = x_0 \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right)$$

and:

$$W(t) = g^{-1}(X(t)) = \frac{\ln X(t) - \ln x_0 - \left(\mu - \frac{1}{2}\sigma^2\right)t}{\sigma}$$

[22] See also Exercise 13.4.8 on page 891 for an alternative proof (Glasserman, 2003).

We assume that $x_0 = 100$, $\mu = 0$ and $\sigma = 10\%$. The fixed values of $X(t)$ are given in the table below. Using the previous formula, we deduce the values taken by the Brownian bridge:

τ_j	$X(\tau_i)$	$W(\tau_i)$
0	100	0.0000
1	110	1.0031
3	100	0.1500
5	90	−0.8036

We have reported five simulated path of this diffusion bridge in Figure 13.26.

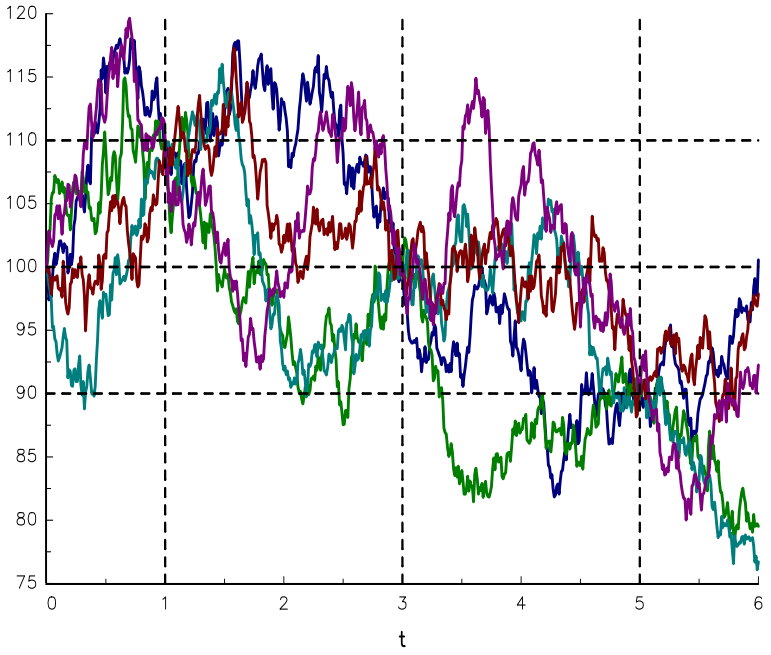


FIGURE 13.26: Simulation of the diffusion bridge $X(t)$

Diffusion bridges are important in finance when we would like to study extremes of a diffusion process. If we want to find the maximum of the stochastic process $X(t)$ over $[0, T]$, we can simulate $X(t)$ and take the maximum of the generated path:

$$\hat{M} = \max_m X_m$$

Another approach consists in locating the maximum:

$$m^\star = \arg\max_m X_m$$

and simulating the diffusion bridge $B(t)$ such that $X(t_{m^\star-1}) = X_{m^\star-1}$, $X(t_{m^\star}) = X_{m^\star}$ and $X(t_{m^\star+1}) = X_{m^\star+1}$. In this case, we can define another estimator of the maximum:

$$\tilde{M} = \max_{t \in \left[t_{m^\star-1}, t_{m^\star+1}\right]} B(t)$$

By construction, we always have $\tilde{M} \geq \hat{M}$. For instance, we report the probability density function of $\hat{M}$ and $\tilde{M}$ in Figure 13.27 when we consider the geometric Brownian motion

with $x_0 = 100$, $\mu = 0$, $\sigma = 15\%$ and $T = 1$. The GBM process has been simulated with a fixed stepsize $h = 0.1$, whereas the diffusion bridge has been simulated with $h = 0.001$. This implies that each path uses $1/0.1 = 10$ discretization points in the first case and $10 + 0.2/0.001 = 210$ discretization points in the second case. The estimation based on the diffusion bridge is then equivalent to consider a scheme with $1/0.001 = 1\,000$ discretization points.

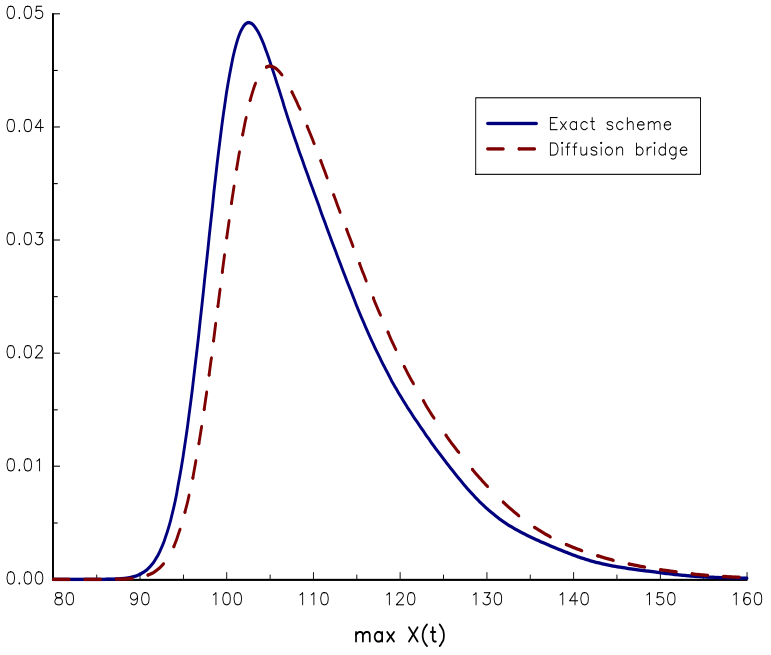


FIGURE 13.27: Density of the maximum estimators $\hat{M}$ and $\tilde{M}$

Remark 159 *In the case of the geometric Brownian motion $X(t)$, the distribution function of the maximum is known. Indeed, we have*[23]*:*

$$\Pr\{M(t) \geq x\} = \exp\left(\frac{2\eta x}{\sigma^2}\right)\Phi\left(\frac{-x-\eta t}{\sigma\sqrt{t}}\right) + \Phi\left(\frac{-x+\eta t}{\sigma\sqrt{t}}\right)$$

where $M(t)$ is the maximum of a Brownian motion with a constant drift:

$$M(t) = \max_{s\leq t} \eta t + \sigma W(t)$$

We notice that:

$$\ln\frac{X(t)}{X(0)} = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)$$

[23] In the case of the minimum, we can use the following identity:

$$m(t) = \min_{s\leq t} \eta t + \sigma W(t) = -\max_{s\leq t} -\eta t - \sigma W(t)$$

It follows that:

$$\Pr\{m(t) \leq x\} = \exp\left(\frac{2\eta x}{\sigma^2}\right)\Phi\left(\frac{x+\eta t}{\sigma\sqrt{t}}\right) + \Phi\left(\frac{x-\eta t}{\sigma\sqrt{t}}\right)$$

It follows that:

$$\begin{aligned}\Pr\left\{\max_{s\leq t} X(s) \geq x\right\} &= \Pr\left\{\max_{s\leq t} \ln X(s) \geq \ln x\right\} \\ &= \Pr\left\{M(t) \geq \ln x - \ln x_0\right\}\end{aligned}$$

where $\eta = \mu - \frac{1}{2}\sigma^2$.

Remark 160 *Diffusion bridges are extensively used when pricing look-back options by Monte Carlo, but also barrier options. Indeed, we need to locate precisely the stopping time when the process crosses the barrier. More generally, they may accelerate the convergence of Monte Carlo methods in the case of path-dependent derivatives (Glasserman, 2003).*

13.2.3 Multivariate continuous-time processes

13.2.3.1 Multidimensional Brownian motion

Let $W(t) = (W_1(t), \ldots, W_n(t))$, be a n-dimensional Brownian motion. Each component $W_i(t)$ is a Brownian motion:

$$W_i(t) - W_i(s) \sim \mathcal{N}(0, t-s)$$

Moreover, we have:

$$\mathbb{E}\left[W_i(t) W_j(s)\right] = \min(t,s) \cdot \rho_{i,j}$$

where $\rho_{i,j}$ is the correlation between the two Brownian motions W_i and W_j. We deduce that:

$$\begin{cases} W(0) = \mathbf{0} \\ W(t) = W(s) + \epsilon(s,t) \end{cases}$$

where $\epsilon(s,t) \sim \mathcal{N}_n(\mathbf{0}, (t-s)\,\rho)$ are *iid* random vectors. It follows that the numerical solution is:

$$W_{m+1} = W_m + \sqrt{t_{m+1} - t_m} \cdot P \cdot \varepsilon_m$$

where P is the Cholesky decomposition of the correlation matrix ρ and $\varepsilon_m \sim \mathcal{N}_n(0, I)$ are *iid* random vectors. In the case of fixed-interval times, the recursion becomes:

$$W_{m+1} = W_m + \sqrt{h} \cdot P \cdot \varepsilon_m$$

In Figures 13.28 and 13.29, we simulate the realization of two-dimensional Brownian motions. Since the two simulated paths use the same random numbers, the difference comes from the correlation $\rho_{1,2}$, which is equal to zero for the first case and 85% for the second case.

13.2.3.2 Multidimensional geometric Brownian motion

Let us now consider the multidimensional geometric Brownian motion[24]:

$$\begin{cases} \mathrm{d}X(t) = \mu \odot X(t)\ \mathrm{d}t + \mathrm{diag}\left(\sigma \odot X(t)\right) \mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

[24]The symbol $\odot$ is the Hadamard product.

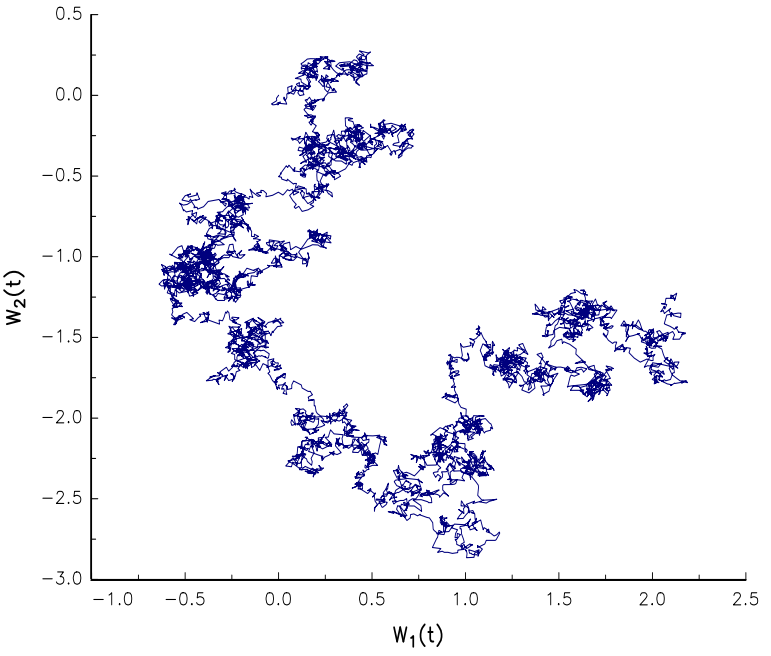


FIGURE 13.28: Brownian motion in the plane (independent case)

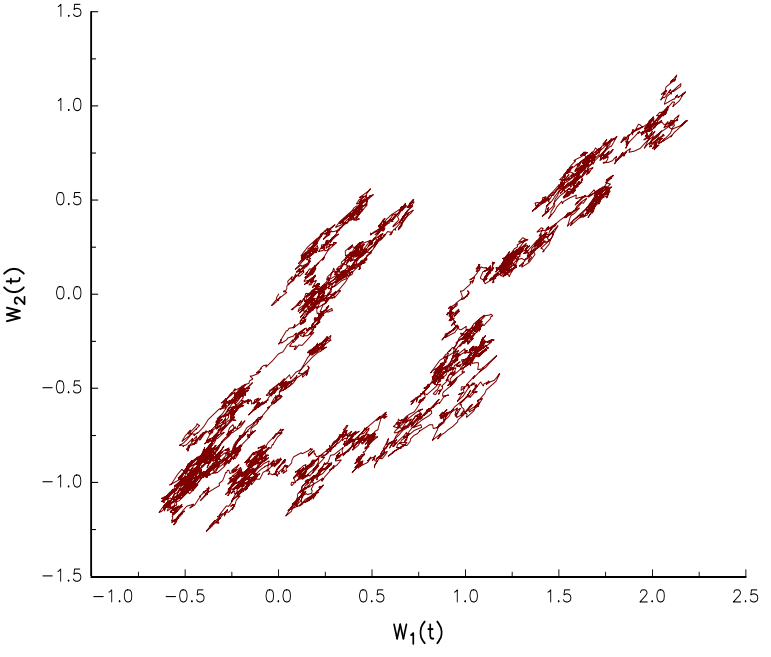


FIGURE 13.29: Brownian motion in the plane ($\rho_{1,2} = 85\%$)

where $X(t) = (X_1(t), \ldots, X_n(t))$, $\mu = (\mu_1, \ldots, \mu_n)$, $\sigma = (\sigma_1, \ldots, \sigma_n)$ and $W(t) = (W_1(t), \ldots, W_n(t))$ is a n-dimensional Brownian motion with $\mathbb{E}\left[W(t) W(t)^\top\right] = \rho t$. If we consider the j^{th} component of $X(t)$, we have:

$$\mathrm{d}X_j(t) = \mu_j X_j(t)\,\mathrm{d}t + \sigma_j X_j(t)\,\mathrm{d}W_j(t)$$

The solution of the multidimensional SDE is a multivariate log-normal process with:

$$X_j(t) = X_j(0) \cdot \exp\left(\left(\mu_j - \frac{1}{2}\sigma_j^2\right) t + \sigma_j W_j(t)\right)$$

where $W(t) \sim \mathcal{N}_n(0, \rho t)$. We deduce that the exact scheme to simulate the multivariate GBM is:

$$\begin{cases} X_{1,m+1} = X_{1,m} \cdot \exp\left(\left(\mu_1 - \frac{1}{2}\sigma_1^2\right)(t_{m+1} - t_m) + \sigma_1\sqrt{t_{m+1} - t_m} \cdot \varepsilon_{1,m}\right) \\ \vdots \\ X_{j,m+1} = X_{j,m} \cdot \exp\left(\left(\mu_j - \frac{1}{2}\sigma_j^2\right)(t_{m+1} - t_m) + \sigma_j\sqrt{t_{m+1} - t_m} \cdot \varepsilon_{j,m}\right) \\ \vdots \\ X_{n,m+1} = X_{n,m} \cdot \exp\left(\left(\mu_n - \frac{1}{2}\sigma_n^2\right)(t_{m+1} - t_m) + \sigma_n\sqrt{t_{m+1} - t_m} \cdot \varepsilon_{n,m}\right) \end{cases}$$

where $(\varepsilon_{1,m}, \ldots, \varepsilon_{n,m}) \sim \mathcal{N}_n(\mathbf{0}, \rho)$.

Remark 161 *Monte Carlo methods extensively use this scheme for calculating the price of multi-asset derivatives in the Black-Scholes model.*

13.2.3.3 Euler-Maruyama and Milstein schemes

We consider the general SDE:

$$\begin{cases} \mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

where $X(t)$ and $\mu(t, X(t))$ are $n \times 1$ vectors, $\sigma(t, X(t))$ is a $n \times p$ matrix and $W(t)$ is a $p \times 1$ vector. We assume that $\mathbb{E}\left[W(t) W(t)^\top\right] = \rho t$, where ρ is a $p \times p$ correlation matrix. The corresponding Euler-Maruyama scheme is:

$$X_{m+1} = X_m + \mu(t_m, X_m) \cdot (t_{m+1} - t_m) + \sigma(t_m, X_m)\sqrt{t_{m+1} - t_m} \cdot \varepsilon_m$$

where $\varepsilon_m \sim \mathcal{N}_p(0, \rho)$. In the case of a diagonal system[25], we retrieve the one-dimensional scheme:

$$X_{j,m+1} = X_{j,m} + \mu_j(t_m, X_{j,m}) \cdot (t_{m+1} - t_m) + \sigma_{j,j}(t_m, X_{j,m}) \cdot \sqrt{t_{m+1} - t_m}\varepsilon_{j,m}$$

However, the random variables $\varepsilon_{j,m}$ and $\varepsilon_{j',m}$ may be correlated.

Example 148 *We consider the Heston model:*

$$\begin{cases} \mathrm{d}X(t) = \mu X(t)\,\mathrm{d}t + \sqrt{v(t)}X(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}v(t) = a(b - v(t))\,\mathrm{d}t + \sigma\sqrt{v(t)}\,\mathrm{d}W_2(t) \end{cases}$$

[25] This means that $\mu_j(t, x) = \mu_j(t, x_j)$ and $\sigma(t, x)$ is a $n \times n$ diagonal matrix with $\sigma_{j,j}(t, x) = \sigma_{j,j}(t, x_j)$.

where $\mathbb{E}[W_1(t) W_2(t)] = \rho t$. *By applying the fixed-interval Euler-Maruyama scheme to* $(\ln X(t), v(t))$, *we obtain:*

$$\ln X_{m+1} = \ln X_m + \left(\mu - \frac{1}{2} v_m\right) h + \sqrt{v_m h} \cdot \varepsilon_{1,m}$$

and[26]*:*

$$v_{m+1} = v_m + a(b - v_m) h + \sigma \sqrt{v_m h} \cdot \varepsilon_{2,m}$$

Here, $\varepsilon_{1,m}$ *and* $\varepsilon_{2,m}$ *are two standard Gaussian random variables with correlation* ρ.

The multidimensional version of the Milstein scheme is[27]:

$$\begin{aligned} X_{j,m+1} &= X_{j,m} + \mu_j(t_m, X_m)(t_{m+1} - t_m) + \sum_{k=1}^{p} \sigma_{j,k}(t_m, X_m) \Delta W_{k,m} + \\ &\quad \sum_{k=1}^{p} \sum_{k'=1}^{p} \mathcal{L}^{(k)} \sigma_{j,k'}(t_m, X_m) \mathcal{I}_{(k,k')} \end{aligned}$$

where:

$$\mathcal{L}^{(k)} f(t,x) = \sum_{k''=1}^{n} \sigma_{k'',k}(t_m, X_m) \frac{\partial f(t,x)}{\partial x_{k''}}$$

and:

$$\mathcal{I}_{(k,k')} = \int_{t_m}^{t_{m+1}} \int_{t_m}^{s} \mathrm{d}W_k(t)\, \mathrm{d}W_{k'}(s)$$

In the case of a diagonal system, the Milstein scheme may be simplified as follows:

$$\begin{aligned} X_{j,m+1} &= X_{j,m} + \mu_j(t_m, X_{j,m})(t_{m+1} - t_m) + \sigma_{j,j}(t_m, X_{j,m}) \Delta W_{j,m} + \\ &\quad \mathcal{L}^{(j)} \sigma_{j,j}(t_m, X_{j,m}) \mathcal{I}_{(j,j)} \end{aligned}$$

where[28]:

$$\begin{aligned} \mathcal{I}_{(j,j)} &= \int_{t_m}^{t_{m+1}} \int_{t_m}^{s} \mathrm{d}W_j(t)\, \mathrm{d}W_j(s) \\ &= \int_{t_m}^{t_{m+1}} (W_j(s) - W_j(t_m))\, \mathrm{d}W_j(s) \\ &= \frac{1}{2}\left((\Delta W_{j,m})^2 - (t_{m+1} - t_m)\right) \end{aligned}$$

We deduce that the Milstein scheme is:

$$\begin{aligned} X_{j,m+1} &= X_{j,m} + \mu_j(t_m, X_{j,m})(t_{m+1} - t_m) + \\ &\quad \sigma_{j,j}(t_m, X_{j,m}) \sqrt{t_{m+1} - t_m} \varepsilon_{j,m} + \\ &\quad \frac{1}{2} \sigma_{j,j}(t_m, X_{j,m}) \partial_{x_j} \sigma_{j,j}(t_m, X_{j,m})(t_{m+1} - t_m)\left(\varepsilon_{j,m}^2 - 1\right) \end{aligned}$$

[26] To avoid that v_{m+1} is negative, we can use the truncation method:

$$v_{m+1} \leftarrow \max(v_{m+1}, 0)$$

or the reflection method:

$$v_{m+1} \leftarrow |v_{m+1}|$$

[27] We have $\Delta W_{k,m} = W_k(t_{m+1}) - W_k(t_m)$.

[28] By applying Itô's lemma to $Y_t = \frac{1}{2}\left(W_j(t)^2 - t\right)$, we obtain $\mathrm{d}Y(t) = W_j(t)\, \mathrm{d}W_j(t)$.

We obtain the same expression as the formula given by Equation (13.3), except that the random variables $\varepsilon_{j,m}$ and $\varepsilon_{j',m}$ may now be correlated.

Example 149 *If we apply the fixed-interval Milstein scheme to the Heston model, we obtain:*

$$\ln X_{m+1} = \ln X_m + \left(\mu - \frac{1}{2}v_m\right) h + \sqrt{v_m h} \cdot \varepsilon_{1,m}$$

and:

$$v_{m+1} = v_m + a\left(b - v_m\right) h + \sigma\sqrt{v_m h} \cdot \varepsilon_{2,m} + \frac{1}{4}\sigma^2 h\left(\varepsilon_{2,m}^2 - 1\right)$$

Here, $\varepsilon_{1,m}$ and $\varepsilon_{2,m}$ are two standard Gaussian random variables with correlation ρ.

Remark 162 *The multidimensional Milstein scheme is generally not used, because the terms $\mathcal{L}^{(k)}\sigma_{j,k'}\left(t_m, X_m\right)\mathcal{I}_{(k,k')}$ are complicated to simulate. For the Heston model, we obtain a very simple scheme, because we only apply the Milstein scheme to the process $v\left(t\right)$ and not to the vector process $\left(\ln X\left(t\right), v\left(t\right)\right)$. If we also apply the Milstein scheme to $\ln X\left(t\right)$, we obtain:*

$$\ln X_{m+1} = \ln X_m + \left(\mu - \frac{1}{2}v_m\right) h + \sqrt{v_m h} \cdot \varepsilon_{1,m} + A_m$$

where:

$$\begin{aligned} A_m &= \sum_{k=1}^{2}\sum_{k'=1}^{2}\left(\sum_{k''=1}^{2}\sigma_{k'',k}\left(t_m, X_m\right)\frac{\sigma_{1,k'}\left(t_m, X_m\right)}{\partial\, x_{k''}}\right)\mathcal{I}_{(k,k')} \\ &= \sigma\sqrt{v\left(t\right)} \cdot \frac{1}{2\sqrt{v\left(t\right)}} \cdot \mathcal{I}_{(2,1)} \\ &= \frac{\sigma}{2} \cdot \mathcal{I}_{(2,1)} \end{aligned}$$

Let $W_2\left(t\right) = \rho W_1\left(t\right) + \sqrt{1-\rho^2}W^\star\left(t\right)$ where $W^\star\left(t\right)$ is a Brownian motion independent from $W_1\left(t\right)$. It follows that:

$$\begin{aligned} \mathcal{I}_{(2,1)} &= \int_{t_m}^{t_{m+1}}\int_{t_m}^{s} \mathrm{d}W_2\left(t\right)\mathrm{d}W_1\left(s\right) \\ &= \int_{t_m}^{t_{m+1}}\left(\rho W_1\left(s\right) + \sqrt{1-\rho^2}W^\star\left(s\right)\right)\mathrm{d}W_1\left(s\right) - \\ &\quad \int_{t_m}^{t_{m+1}}\left(\rho W_1\left(t_m\right) + \sqrt{1-\rho^2}W^\star\left(t_m\right)\right)\mathrm{d}W_1\left(s\right) \\ &= \rho\int_{t_m}^{t_{m+1}}\left(W_1\left(s\right) - W_1\left(t_m\right)\right)\mathrm{d}W_1\left(s\right) + \\ &\quad \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}}\left(W^\star\left(s\right) - W^\star\left(t_m\right)\right)\mathrm{d}W_1\left(s\right) \end{aligned}$$

and:

$$\mathcal{I}_{(2,1)} = \frac{1}{2}\rho\left(\left(\Delta W_{1,m}\right)^2 - \left(t_{m+1} - t_m\right)\right) + B_m$$

We finally deduce that the multidimensional Milstein scheme of the Heston model is:

$$\ln X_{m+1} = \ln X_m + \left(\mu - \frac{1}{2}v_m\right) h + \sqrt{v_m h} \cdot \varepsilon_{1,m} + \frac{1}{4}\rho\sigma h\left(\varepsilon_{1,m}^2 - 1\right) + B_m$$

and:

$$v_{m+1} = v_m + a\left(b - v_m\right)h + \sigma\sqrt{v_m h}\cdot\varepsilon_{2,m} + \frac{1}{4}\sigma^2 h\left(\varepsilon_{2,m}^2 - 1\right)$$

where B_m is a correction term defined by:

$$B_m = \sqrt{1-\rho^2}\int_{t_m}^{t_{m+1}}\left(W^\star(s) - W^\star(t_m)\right)\mathrm{d}W_1(s)$$

We notice that B_m cannot be explicitly calculate and requires numerical integration to be simulated[29] .

13.3 Monte Carlo methods

At the beginning, the Monte Carlo method is a numerical tool for computing integrals based on the simulation of random variables (Metropolis and Ulam, 1949). By extension, it now defines all numerical methods, which use simulations.

13.3.1 Computing integrals

13.3.1.1 A basic example

One of the early uses of the Monte Carlo method was the numerical calculation of the number π by Bouffon and Laplace. Suppose we have a circle with radius r and a $2r \times 2r$ square of the same center. Since the area of the circle is equal to πr^2, the numerical calculation of π is equivalent to compute the area of the circle with $r = 1$. In this case, the area of the square is 4, and we have[30]:

$$\pi = 4\frac{\mathcal{A}(\text{circle})}{\mathcal{A}(\text{square})}$$

To determine π, we simulate n_S random vectors (u_s, v_s) of uniform random variables $\mathcal{U}_{[-1,1]}$ and we obtain:

$$\pi = \lim_{n_S\to\infty} 4\frac{n_c}{n}$$

where n_c is the number of points (u_s, v_s) in the circle:

$$n_c = \sum_{s=1}^{n_S}\mathbb{1}\left\{u_s^2 + v_s^2 \leq r^2\right\}$$

We illustrate this numerical computation in Figure 13.30 with 1 000 simulated points (u_s, v_s). We indicate by a red cross symbol (resp. by a blue square symbol) the points which are inside (resp. outside) the circle. In this experiment, we obtain $n_c = 802$ and $\pi \simeq 4 \times 802/1\,000 = 3.2080$.

[29] However, B_m is not independent from $\varepsilon_{1,m}$.

[30] In fact, this relationship holds for all values of r.

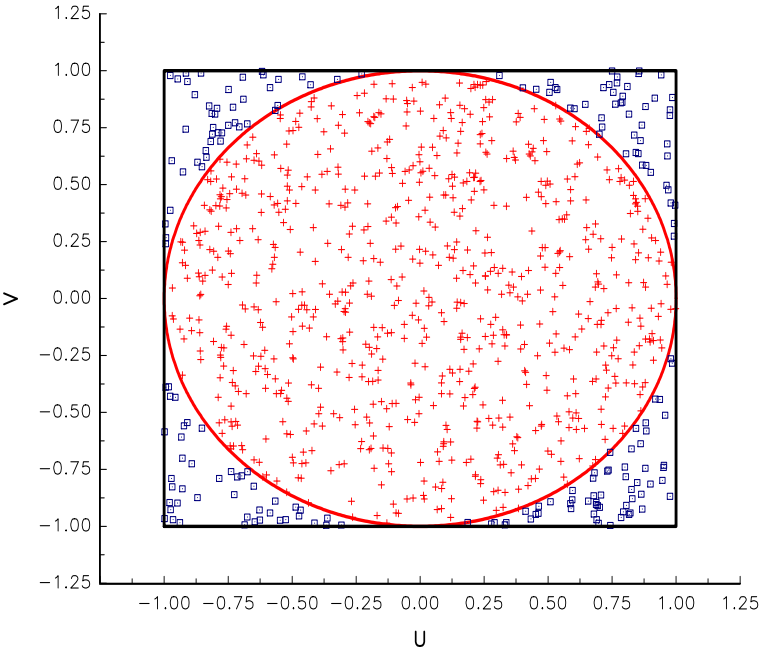


FIGURE 13.30: Computing π with 1 000 simulations

13.3.1.2 Theoretical framework

We consider the multiple integral:

$$I = \int \cdots \int_{\Omega} \varphi(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

Let $X = (X_1, \ldots, X_n)$ be a uniform random vector with probability distribution $\mathcal{U}_{[\Omega]}$, such that Ω is inscribed within the hypercube $[\Omega]$. By construction, the probability density function is:

$$f(x_1, \ldots, x_n) = 1$$

We deduce that:

$$\begin{aligned} I &= \int \cdots \int_{[\Omega]} \mathbb{1}\{(x_1, \ldots, x_n) \in \Omega\} \cdot \varphi(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \mathbb{E}\left[\mathbb{1}\{(X_1, \ldots, X_n) \in \Omega\} \cdot \varphi(X_1, \ldots, X_n)\right] \\ &= \mathbb{E}\left[h(X_1, \ldots, X_n)\right] \end{aligned}$$

where:

$$h(x_1, \ldots, x_n) = \mathbb{1}\{(x_1, \ldots, x_n) \in \Omega\} \cdot \varphi(x_1, \ldots, x_n)$$

Let $\hat{I}_{n_S}$ be the random variable defined by:

$$\hat{I}_{n_S} = \frac{1}{n_S} \sum_{s=1}^{n_S} h(X_{1,s}, \ldots, X_{n,s})$$

where $\{X_{1,s}, \ldots, X_{n,s}\}_{s \geq 1}$ is a sequence of *iid* random vectors with probability distribution $\mathcal{U}_{[\Omega]}$. Using the strong law of large numbers, we obtain:

$$\begin{aligned} \lim_{n_s \to \infty} \hat{I}_{n_s} &= \mathbb{E}\left[h\left(X_1, \ldots, X_n\right)\right] \\ &= \int \cdots \int_{\Omega} \varphi\left(x_1, \ldots, x_n\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned}$$

Moreover, the central limit theorem states that:

$$\lim_{n_s \to \infty} \sqrt{n_S} \left(\frac{\hat{I}_{n_s} - I}{\sigma\left(h\left(X_1, \ldots, X_n\right)\right)} \right) = \mathcal{N}(0,1)$$

When n_S is large, we can deduce the following confidence interval:

$$\left[\hat{I}_{n_S} - c_\alpha \cdot \frac{\hat{S}_{n_S}}{\sqrt{n_S}}, \hat{I}_{n_S} + c_\alpha \cdot \frac{\hat{S}_{n_S}}{\sqrt{n_S}} \right]$$

where α is the confidence level, $c_\alpha = \Phi^{-1}\left((1+\alpha)/2\right)$ and $\hat{S}_{n_S}$ is the usual estimate of the standard deviation:

$$\hat{S}_{n_S} = \sqrt{\frac{1}{n_S - 1} \sum_{s=1}^{n_S} h^2\left(X_{1,s}, \ldots, X_{n,s}\right) - \hat{I}_{n_s}}$$

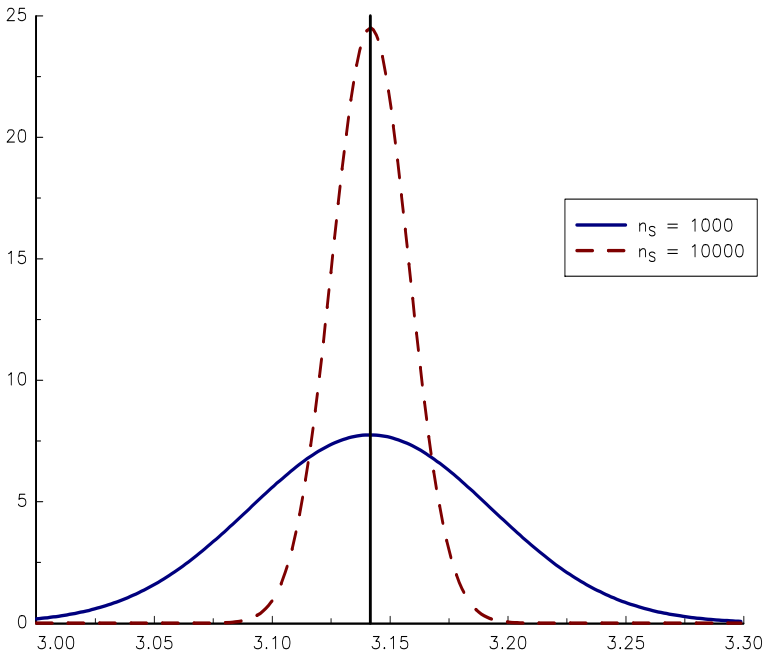


FIGURE 13.31: Density function of $\hat{\pi}_{n_S}$

We consider again the calculation of π. Previously, we obtain an estimate, which is far from the true value. In order to obtain a better precision, we can increase the number n_S

of simulations. In Figure 13.31, we report the density function of the estimator $\hat{\pi}_{n_S}$ when n_S is respectively equal to 1 000 and 10 000. We notice that the precision increases by a factor of $\sqrt{10}$ every time we multiply the number of simulations by ten. More generally, for a given precision p, we can deduce the sufficient number of simulations:

$$n_S \geq \left(c_\alpha \frac{\hat{S}_{n_S}}{p} \right)^2$$

In the case of the calculation of π, we have $\hat{S}_{n_S} \approx 1.642$. To obtain a precision of π with six digits after the decimal point at a 99% confidence level, we need about 18 trillion simulations:

$$\begin{aligned} n_S &\geq \left(\Phi^{-1}(0.995) \times \frac{1.642}{10^{-6}} \right)^2 \\ &\geq 17.9 \times 10^{12} \end{aligned}$$

Example 150 *We would like to calculate the following integral:*

$$I = \iiint_\Omega (x + y + z)^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

This integral can be easily evaluated with Gaussian quadrature methods when Ω is a cube. However, the problem is more tricky when:

$$\Omega = \left\{ (x, y, z) \in \mathbb{R}_+^3 : x^2 + y^2 + z^2 \leq 25, x + y + z \geq 2 \right\}$$

Using the Monte Carlo method, we have $I = \mathbb{E}[h(X, Y, Z)]$ where X, Y and Z are three independent uniform random variables with probability distribution $\mathcal{U}_{[0,5]}$ and:

$$h(x, y, z) = \begin{cases} (x + y + z)^2 & \text{if} \quad (x, y, z) \in \Omega \\ 0 & \text{if} \quad (x, y, z) \notin \Omega \end{cases}$$

In Figure 13.32, we report the estimate $\hat{I}_{n_S}$ and the corresponding 99% confidence interval with respect to the number of simulations n_S.

13.3.1.3 Extension to the calculation of mathematical expectations

Let $X = (X_1, \ldots, X_n)$ be a random vector with probability distribution $\mathbf{F}$. We have:

$$\begin{aligned} \mathbb{E}[\varphi(X_1, \ldots, X_n)] &= \int \cdots \int \varphi(x_1, \ldots, x_n) \, \mathrm{d}\mathbf{F}(x_1, \cdots, x_n) \\ &= \int \cdots \int \varphi(x_1, \ldots, x_n) \, f(x_1, \cdots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \int \cdots \int h(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned}$$

where f is the density function. The Monte Carlo estimator of this integral is:

$$\hat{I}_{n_S} = \frac{1}{n_S} \sum_{s=1}^{n_S} \varphi(X_{1,s}, \ldots, X_{n,s})$$

where $\{X_{1,s}, \ldots, X_{n,s}\}_{s \geq 1}$ is a sequence of *iid* random vectors with probability distribution $\mathbf{F}$. Moreover, all the previous results hold in this general case where the random variables are not uniform.

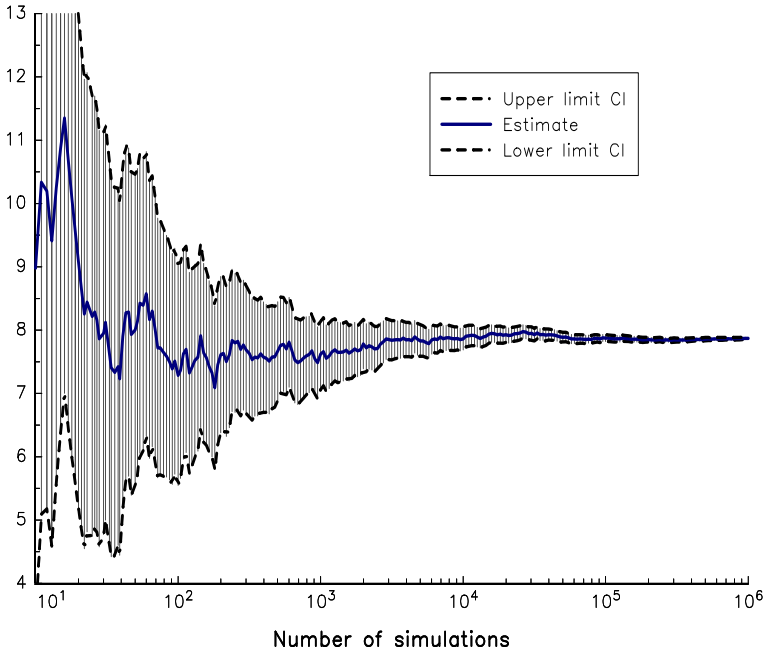


FIGURE 13.32: Convergence of the estimator $\hat{I}_{n_S}$

Example 151 *In the Black-Scholes model, the price of the look-back option with maturity T is given by:*

$$\mathcal{C} = e^{-rT}\mathbb{E}\left[\left(S(T) - \min_{0\leq t\leq T} S(t)\right)^+\right]$$

where the price $S(t)$ of the underlying asset is given by the following SDE:

$$\mathrm{d}S(t) = rS(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t)$$

where r is the interest rate and σ is the volatility of the asset. It is difficult to calculate $\mathcal{C}$ analytically, because it requires the joint distribution of $S(T)$ and $\min_{0\leq t\leq T} S(t)$. However, we can easily calculate it using the Monte Carlo method. For a given simulation s, we use the exact scheme to simulate the geometric Brownian motion:

$$S_{m+1}^{(s)} = S_m^{(s)} \cdot \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(t_{m+1} - t_m) + \sigma\sqrt{t_{m+1} - t_m}\cdot\varepsilon_m^{(s)}\right)$$

where $\varepsilon_m^{(s)} \sim \mathcal{N}(0,1)$ and $T = t_M$. The Monte Carlo estimator of the option price is then equal to:

$$\widehat{\mathcal{C}} = \frac{e^{-rT}}{n_S}\sum_{s=1}^{n_S}\left(S_M^{(s)} - \min_m S_m^{(s)}\right)^+$$

We deduce that the precision of the estimate depends on the number n_S of simulations, but also on the number M of discretization points. In Figure 13.33, the option price is calculated using these parameters: $S_0 = 100$, $r = 5\%$, $\sigma = 20\%$ and $T = 3/12$. We consider 100 000 simulations whereas the number M of discretization points varies between 5 and 100. We notice that the 99% confidence interval does not really depend on M. However, the option price increases with the number of discretization points. This is normal because

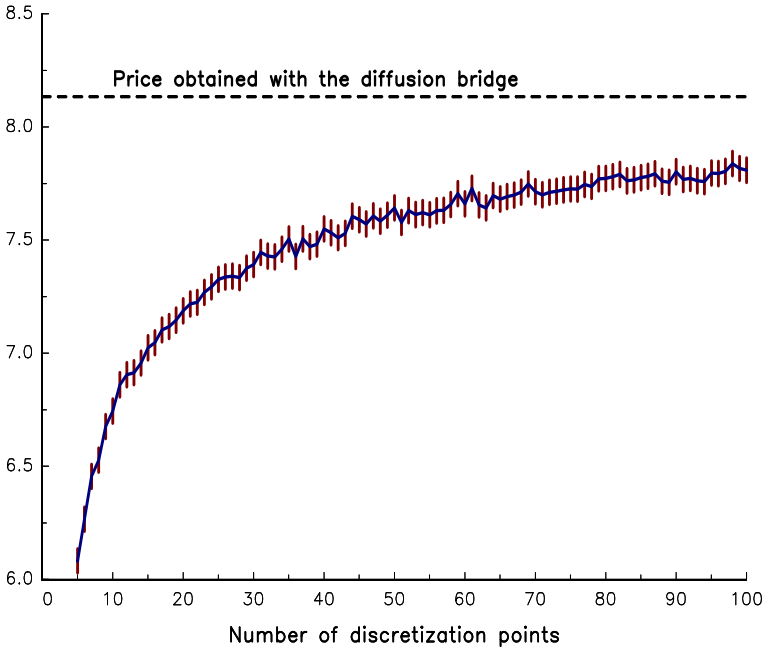


FIGURE 13.33: Computing the look-back option price

$\min_{0\leq t\leq T} S(t)$ *is always underestimated by* $\min_m S_m^{(s)}$. *This is why we have also reported the option price using the diffusion bridge approach. This example shows that the MC method does not always converge when the function* $\varphi(x_1,\ldots,x_n)$ *is approximated.*

Let us consider the following integral:

$$I = \int \cdots \int h(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

We can write it as follows:

$$I = \int \cdots \int \frac{h(x_1,\ldots,x_n)}{f(x_1,\cdots,x_n)} f(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

where $f(x_1,\cdots,x_n)$ is a multidimensional density function. We deduce that:

$$I = \mathbb{E}\left[\frac{h(X_1,\ldots,X_n)}{f(X_1,\ldots,X_n)}\right]$$

This implies that we can compute an integral with the MC method by using any multidimensional distribution function. If we apply this result to the calculation of π, we have:

$$\begin{aligned}
\pi &= \iint_{x^2+y^2\leq 1} \mathrm{d}x\,\mathrm{d}y \\
&= \iint \mathbb{1}\left\{x^2+y^2\leq 1\right\}\,\mathrm{d}x\,\mathrm{d}y \\
&= \iint \frac{\mathbb{1}\left\{x^2+y^2\leq 1\right\}}{\phi(x)\,\phi(y)}\phi(x)\,\phi(y)\,\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

We deduce that:

$$\pi = \mathbb{E}\left[\frac{\mathbb{1}\left\{X^2 + Y^2 \leq 1\right\}}{\phi(X)\,\phi(Y)}\right]$$

where X and Y are two independent standard Gaussian random variables. We can then estimate π by:

$$\hat{\pi}_{n_S} = \frac{1}{n_S}\sum_{s=1}^{n_S}\frac{\mathbb{1}\left\{x_s^2 + y_s^2 \leq 1\right\}}{\phi(x_s)\,\phi(y_s)}$$

where x_s and y_s are two independent random variates from the probability distribution $\mathcal{N}(0,1)$. For instance, we report the points (x_s, y_s) used to calculate π with 1 000 simulations in Figure 13.34.

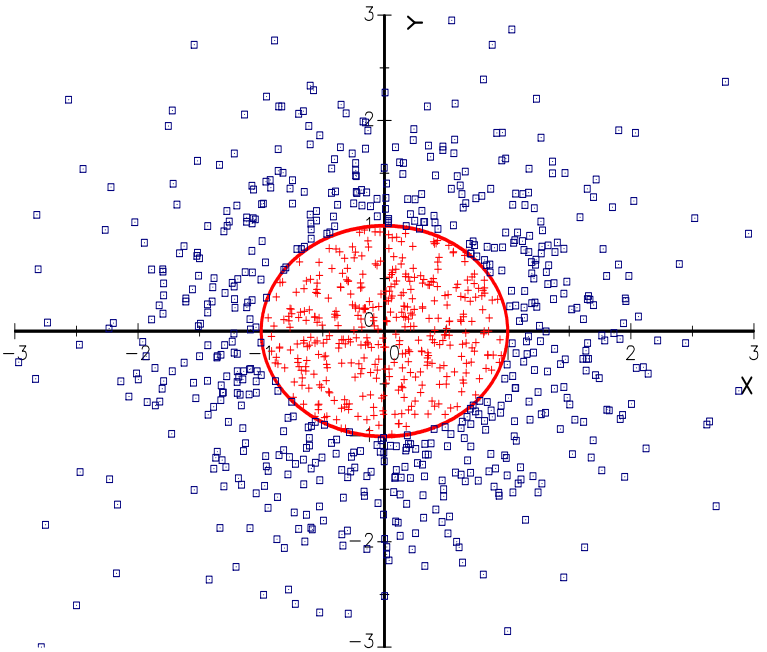


FIGURE 13.34: Computing π with normal random numbers

Remark 163 *The previous approach is particularly interesting when the set Ω is not bounded, which implies that we cannot use uniform random numbers.*

13.3.2 Variance reduction

We consider two unbiased estimators $\hat{I}_{n_S}^{(1)}$ and $\hat{I}_{n_S}^{(2)}$ of the integral I, meaning that $\mathbb{E}\left[\hat{I}_{n_S}^{(1)}\right] = \mathbb{E}\left[\hat{I}_{n_S}^{(2)}\right] = I$. We will say that $\hat{I}_{n_S}^{(1)}$ is more efficient than $\hat{I}_{n_S}^{(2)}$ if the inequality $\text{var}\left(\hat{I}_{n_S}^{(1)}\right) \leq \text{var}\left(\hat{I}_{n_S}^{(2)}\right)$ holds for all values of n_S that are larger than $n_S^\star$. Variance reduction is then the search of more efficient estimators.

13.3.2.1 Antithetic variates

Theoretical aspects We have:

$$I = \mathbb{E}\left[\varphi(X_1, \ldots, X_n)\right] = \mathbb{E}\left[Y\right]$$

where $Y = \varphi(X_1, \ldots, X_n)$ is a one-dimensional random variable. It follows that:

$$\hat{I}_{n_S} = \bar{Y}_{n_S} = \frac{1}{n_S} \sum_{s=1}^{n_S} Y_s$$

We now consider the estimators $\bar{Y}_{n_S}$ and $\bar{Y}'_{n_S}$ based on two different samples and define $\bar{Y}^\star$ as follows:

$$\bar{Y}^\star = \frac{\bar{Y}_{n_S} + \bar{Y}'_{n_S}}{2}$$

We have:

$$\begin{aligned}
\mathbb{E}\left[\bar{Y}^\star\right] &= \mathbb{E}\left[\frac{\bar{Y}_{n_S} + \bar{Y}'_{n_S}}{2}\right] \\
&= \mathbb{E}\left[\bar{Y}_{n_S}\right] \\
&= I
\end{aligned}$$

and:

$$\begin{aligned}
\text{var}\left(\bar{Y}^\star\right) &= \text{var}\left(\frac{\bar{Y}_{n_S} + \bar{Y}'_{n_S}}{2}\right) \\
&= \frac{1}{4}\text{var}\left(\bar{Y}_{n_S}\right) + \frac{1}{4}\text{var}\left(\bar{Y}'_{n_S}\right) + \frac{1}{2}\text{cov}\left(\bar{Y}_{n_S}, \bar{Y}'_{n_S}\right) \\
&= \frac{1 + \rho\left\langle \bar{Y}_{n_S}, \bar{Y}'_{n_S}\right\rangle}{2}\text{var}\left(\bar{Y}_{n_S}\right) \\
&= \frac{1 + \rho\left\langle Y_s, Y'_s\right\rangle}{2}\text{var}\left(\bar{Y}_{n_S}\right)
\end{aligned}$$

where[31] $\rho\left\langle Y_s, Y'_s\right\rangle$ is the correlation between Y_s and Y'_s. Because we have $\rho\left\langle Y_s, Y'_s\right\rangle \leq 1$, we deduce that:

$$\text{var}\left(\bar{Y}^\star\right) \leq \text{var}\left(\bar{Y}_{n_S}\right)$$

If we simulate the random variates Y_s and Y'_s independently, $\rho\left\langle Y_s, Y'_s\right\rangle$ is equal to zero and the variance of the estimator is divided by 2. However, the number of simulations have been multiplied by two. The efficiency of the estimator has then not been improved.

The underlying idea of antithetic variables is therefore to use two perfectly dependent random variables Y_s and Y'_s:

$$Y'_s = \psi(Y_s)$$

[31] We have:

$$\begin{aligned}
\text{cov}\left(\bar{Y}_{n_S}, \bar{Y}'_{n_S}\right) &= \mathbb{E}\left[\left(\frac{1}{n_S}\sum\nolimits_{s=1}^{n_S} Y_s - \mathbb{E}[Y]\right) \cdot \left(\frac{1}{n_S}\sum\nolimits_{s'=1}^{n_S} Y'_{s'} - \mathbb{E}[Y]\right)\right] \\
&= \frac{1}{n_S^2}\sum\nolimits_{s=1}^{n_S} \mathbb{E}\left[(Y_s - \mathbb{E}[Y]) \cdot \left(Y'_s - \mathbb{E}[Y]\right)\right] + \\
&\quad \frac{2}{n_S^2}\sum\nolimits_{s>s'}^{n_S} \mathbb{E}\left[(Y_s - \mathbb{E}[Y]) \cdot \left(Y'_{s'} - \mathbb{E}[Y]\right)\right] \\
&= \frac{1}{n_S} \cdot \text{cov}\left(Y_s, Y'_s\right) + \frac{2}{n_S^2} \cdot 0
\end{aligned}$$

It follows that:

$$\rho\left\langle \bar{Y}_{n_S}, \bar{Y}'_{n_S}\right\rangle = \frac{\text{cov}\left(\bar{Y}_{n_S}, \bar{Y}'_{n_S}\right)}{\sqrt{\text{var}\left(\bar{Y}_{n_S}\right) \cdot \text{var}\left(\bar{Y}'_{n_S}\right)}} = \rho\left(Y_s, Y'_s\right)$$

where ψ is a deterministic function. This implies that:

$$\bar{Y}_{n_S}^{\star} = \frac{1}{n_S}\sum_{s=1}^{n_S} Y_s^{\star}$$

where:

$$Y_s^{\star} = \frac{Y_s + Y_s'}{2} = \frac{Y_s + \psi(Y_s)}{2}$$

It follows that:

$$\rho\left\langle \bar{Y}_{n_S}, \bar{Y}_{n_S}' \right\rangle = \rho\left\langle Y, Y' \right\rangle = \rho\left\langle Y, \psi(Y) \right\rangle$$

Minimizing the variance $\text{var}\left(\bar{Y}^{\star}\right)$ is then equivalent to minimize the correlation $\rho\langle Y, \psi(Y)\rangle$. We also know that the correlation reaches its lower bound if the dependence function between Y and $\psi(Y)$ is equal to the lower Fréchet copula:

$$\mathbf{C}\langle Y, \psi(Y)\rangle = \mathbf{C}^{-}$$

However, $\rho\langle Y, \psi(Y)\rangle$ is not necessarily equal to -1 except in some special cases.

We consider the one-dimensional case with $Y = \varphi(X)$. If we assume that φ is an increasing function, it follows that:

$$\begin{aligned}\mathbf{C}\langle Y, \psi(Y)\rangle &= \mathbf{C}\langle \varphi(X), \psi(\varphi(X))\rangle \\ &= \mathbf{C}\langle X, \psi(X)\rangle\end{aligned}$$

To obtain the lower bound $\mathbf{C}^{-}$, X and $\psi(X)$ must be countermonotonic. We know that[32]:

$$\psi(X) = \mathbf{F}^{-1}(1 - \mathbf{F}(X)) \tag{13.8}$$

where $\mathbf{F}$ is the probability distribution of X. For instance, if $X \sim \mathcal{U}_{[0,1]}$, we have $X' = 1 - X$. In the case where $X \sim \mathcal{N}(0,1)$, we have:

$$\begin{aligned}X' &= \Phi^{-1}(1 - \Phi(X)) \\ &= \Phi^{-1}(\Phi(-X)) \\ &= -X\end{aligned}$$

Example 152 *We consider the following functions:*

1. $\varphi_1(x) = x^3 + x + 1$;
2. $\varphi_2(x) = x^4 + x^2 + 1$;
3. $\varphi_3(x) = x^4 + x^3 + x^2 + x + 1$;

For each function, we want to estimate $I = \mathbb{E}[\varphi(\mathcal{N}(0,1))]$ *using the antithetic estimator:*

$$\bar{Y}_{n_S}^{\star} = \frac{1}{n_S}\sum_{s=1}^{n_S} \frac{\varphi(X_s) + \varphi(-X_s)}{2}$$

where $X_s \sim \mathcal{N}(0,1)$. *We obtain the following results*[33]*:*

$\varphi(x)$	$\varphi_1(x)$	$\varphi_2(x)$	$\varphi_3(x)$
$\mathbb{E}[\varphi(X_s)]$ *or* $\mathbb{E}[\varphi(-X_s)]$	1	5	5
$\text{var}(\varphi(X_s))$ *or* $\text{var}(\varphi(-X_s))$	22	122	144
$\text{cov}(\varphi(X_s), \varphi(-X_s))$	-22	122	100
$\rho\langle\varphi(X_s), \varphi(-X_s)\rangle$	-1	1	$25/36$

[32] See Section 11.2.1 on page 722.

[33] Let $X \sim \mathcal{N}(0,1)$. We have $\mathbb{E}\left[X^2\right] = 1$, $\mathbb{E}\left[X^{2m}\right] = (2m-1)\,\mathbb{E}\left[X^{2m-2}\right]$ and $\mathbb{E}\left[X^{2m+1}\right] = 0$ for $m \in \mathbb{N}$.

We notice that the antithetic estimator is fully efficient in the first case, because its variance is equal to zero. In the second case, it is not efficient because we have $\text{var}\left(\bar{Y}_{n_S}^{\star}\right) = \text{var}\left(\bar{Y}_{n_S}\right)$. *Finally, the antithetic estimator reduces the variance by* 15.3% *in the last case.*

To understand these numerical results, we must study the relationship between $\mathbf{C}\langle X, X'\rangle$ and $\mathbf{C}\langle Y, Y'\rangle$. Indeed, we have:

$$\left\{\mathbf{C}\langle X, X'\rangle = \mathbf{C}^- \Rightarrow \mathbf{C}\langle Y, Y'\rangle = \mathbf{C}^-\right\} \Leftrightarrow \varphi'(x) \geq 0$$

We have represented the three functions $\varphi_1(x)$, $\varphi_2(x)$ and $\varphi_3(x)$ in Figure 13.35. Because $\varphi_1(x)$ is an increasing function, it follows that the copula function between Y and Y' reaches the lower Fréchet bound. The function $\varphi_2(x)$ is perfectly symmetric around $x = 0$. In this case, it is impossible to reduce the variance of the MC estimator by the use of antithetic Gaussian variates. Even if the function $\varphi_3(x)$ is not monotonous, it is however sufficiently asymmetric to obtain a low but significant reduction of the variance of the MC estimator.

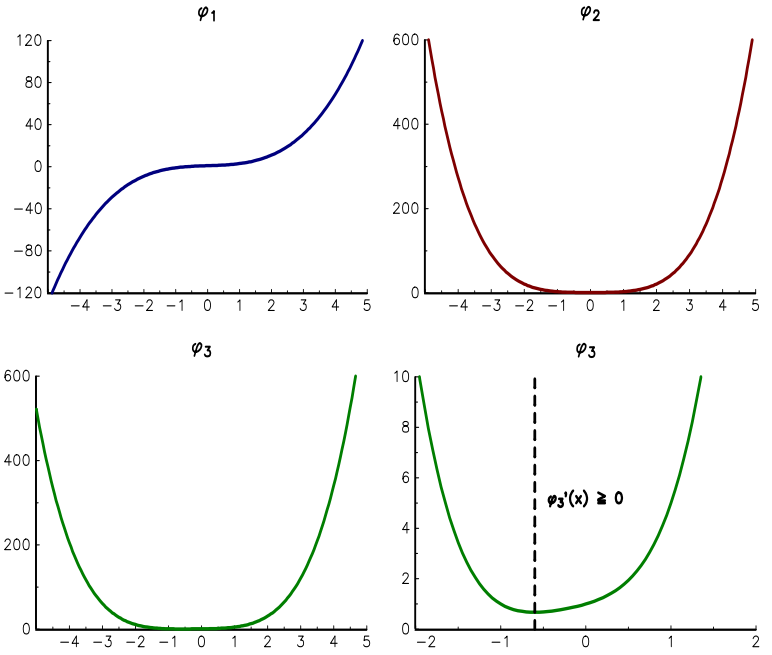


FIGURE 13.35: Functions $\varphi_1(x)$, $\varphi_2(x)$ and $\varphi_3(x)$

Remark 164 *In the case where* φ *is a decreasing function, we can show that the lower bound* $\mathbf{C}^-$ *between* Y *and* Y' *is also reached when* X *and* $\psi(X)$ *are countermonotonic (Ross, 2012).*

The extension of the previous results to the multidimensional case is not straightforward. Indeed, the copula condition between Y and Y' becomes:

$$\mathbf{C}\langle Y, Y'\rangle = \mathbf{C}^- \Leftrightarrow \mathbf{C}\langle \varphi(X_1, \ldots, X_n), \varphi(X_1', \ldots, X_n')\rangle = \mathbf{C}^-$$

where $X_1', \ldots, X_n'$ are the antithetic variates of $X_1, \ldots, X_n$. A natural generalization of the relationship (13.8) is:

$$X_i' = \mathbf{F}_i^{-1}(1 - \mathbf{F}_i(X_i))$$

where $\mathbf{F}_i$ is the probability distribution of X_i. By assuming that φ is a monotonic function of each of its arguments, Ross (2012) shows that:

$$\rho\left\langle Y, Y'\right\rangle < 0$$

where:

$$Y' = \varphi\left(\mathbf{F}_1^{-1}\left(1-\mathbf{F}_1\left(X_1\right)\right), \ldots, \mathbf{F}_n^{-1}\left(1-\mathbf{F}_n\left(X_n\right)\right)\right)$$

This means that we can reduce the variance of the MC estimator by using the antithetic variates $X_i' = \mathbf{F}_i^{-1}\left(1-\mathbf{F}_i\left(X_i\right)\right)$. However, it does not prove that this approach minimizes the correlation $\rho\left\langle Y, Y'\right\rangle$. Moreover, we have no results when φ is a general function.

Application to the geometric Brownian motion In the Gaussian case $X \sim \mathcal{N}(0,1)$, the antithetic variable is:

$$X' = -X$$

As the simulation of $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$ is obtained using the relationship $Y = \mu + \sigma X$, we deduce that the antithetic variable is:

$$\begin{aligned} Y' &= \mu - \sigma X \\ &= \mu - \sigma\frac{(Y-\mu)}{\sigma} \\ &= 2\mu - Y \end{aligned}$$

If we consider the geometric Brownian motion, the fixed-interval scheme is:

$$X_{m+1} = X_m \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)h + \sigma\sqrt{h}\cdot\varepsilon_m\right)$$

whereas the antithetic path is given by:

$$X'_{m+1} = X'_m \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)h - \sigma\sqrt{h}\cdot\varepsilon_m\right)$$

In Figure 13.36, we report 4 trajectories of the GBM process and the corresponding antithetic paths[34].

In the multidimensional case, we recall that:

$$X_{j,m+1} = X_{j,m} \cdot \exp\left(\left(\mu_j - \frac{1}{2}\sigma_j^2\right)h + \sigma_j\sqrt{h}\cdot\varepsilon_{j,m}\right)$$

where $\varepsilon_m = (\varepsilon_{1,m}, \ldots, \varepsilon_{n,m}) \sim \mathcal{N}_n(\mathbf{0}, \rho)$. We simulate ε_m by using the relationship $\varepsilon_m = P \cdot \eta_m$ where $\eta_m \sim \mathcal{N}_n(\mathbf{0}, I_n)$ and P is the Cholesky matrix satisfying $PP^\top = \rho$. The antithetic trajectory is then:

$$X'_{j,m+1} = X'_{j,m} \cdot \exp\left(\left(\mu_j - \frac{1}{2}\sigma_j^2\right)h + \sigma_j\sqrt{k}\cdot\varepsilon'_{j,m}\right)$$

where:

$$\varepsilon'_m = -P \cdot \eta_m = -\varepsilon_m$$

We verify that $\varepsilon'_m = \left(\varepsilon'_{1,m}, \ldots, \varepsilon'_{n,m}\right) \sim \mathcal{N}_n(\mathbf{0}, \rho)$.

[34] The parameter values are $X_0 = 100$, $\mu = 10\%$ and $\sigma = 20\%$.

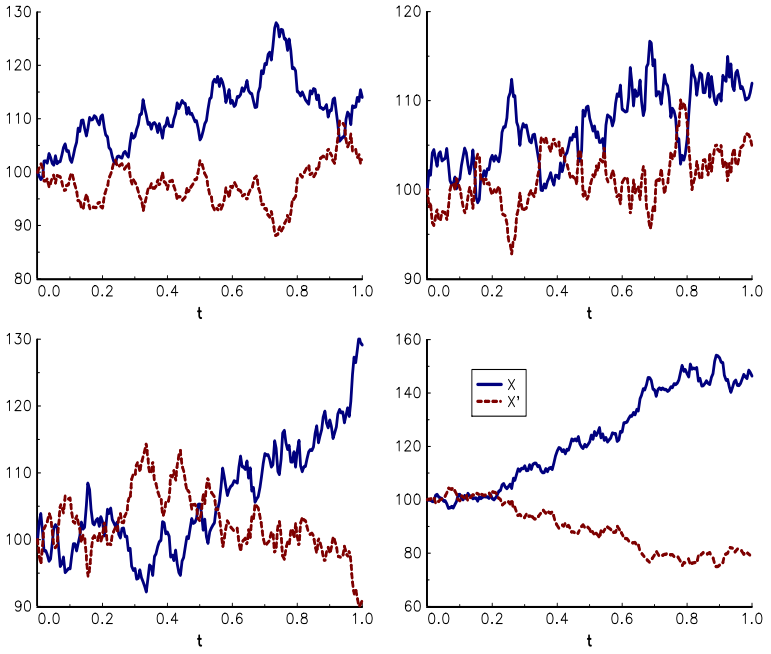


FIGURE 13.36: Antithetic simulation of the GBM process

Example 153 *In the Black-Scholes model, the price of the spread option with maturity T and strike K is given by:*

$$\mathcal{C} = e^{-rT}\mathbb{E}\left[\left(S_1(T) - S_2(T) - K\right)^+\right]$$

where the prices $S_1(t)$ and $S_2(t)$ of the underlying assets are given by the following SDE:

$$\begin{cases} \mathrm{d}S_1(t) = rS_1(t)\,\mathrm{d}t + \sigma_1 S_1(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}S_2(t) = rS_2(t)\,\mathrm{d}t + \sigma_2 S_2(t)\,\mathrm{d}W_2(t) \end{cases}$$

and $\mathbb{E}[W_1(t)\,W_2(t)] = \rho t$. To calculate the option price using Monte Carlo methods, we simulate the bivariate GBM $S_1(t)$ and $S_2(t)$ and the MC estimator is:

$$\hat{\mathcal{C}}_{\mathrm{MC}} = \frac{e^{-rT}}{n_S}\sum_{s=1}^{n_S}\left(S_1^{(s)}(T) - S_2^{(s)}(T) - K\right)^+$$

where $S_j^{(s)}(T)$ is the s^{th} simulation of the terminal value $S_j(T)$. For the AV estimator, we obtain:

$$\hat{\mathcal{C}}_{\mathrm{AV}} = \frac{e^{-rT}}{n_S}\sum_{s=1}^{n_S}\frac{\left(S_1^{(s)}(T) - S_2^{(s)}(T) - K\right)^+ + \left(S_1^{\prime(s)}(T) - S_2^{\prime(s)}(T) - K\right)^+}{2}$$

where $S_j^{\prime(s)}(T)$ is the antithetic variate of $S_j^{(s)}(T)$. In Figure 13.37, we report the probability density function of the estimators $\hat{\mathcal{C}}_{\mathrm{MC}}$ and $\hat{\mathcal{C}}_{\mathrm{AV}}$ when n_S is equal to 1 000[35]*. We observe*

[35]The parameters are $S_1(0) = S_2(0) = 100$, $r = 5\%$, $\sigma_1 = \sigma_2 = 20\%$, $\rho = 50\%$, $T = 1$ and $K = 5$.

that the variance reduction is significative and we obtain:

$$\frac{\text{var}\left(\hat{\mathcal{C}}_{\text{AV}}\right)}{\text{var}\left(\hat{\mathcal{C}}_{\text{MC}}\right)} = 34.7\%$$

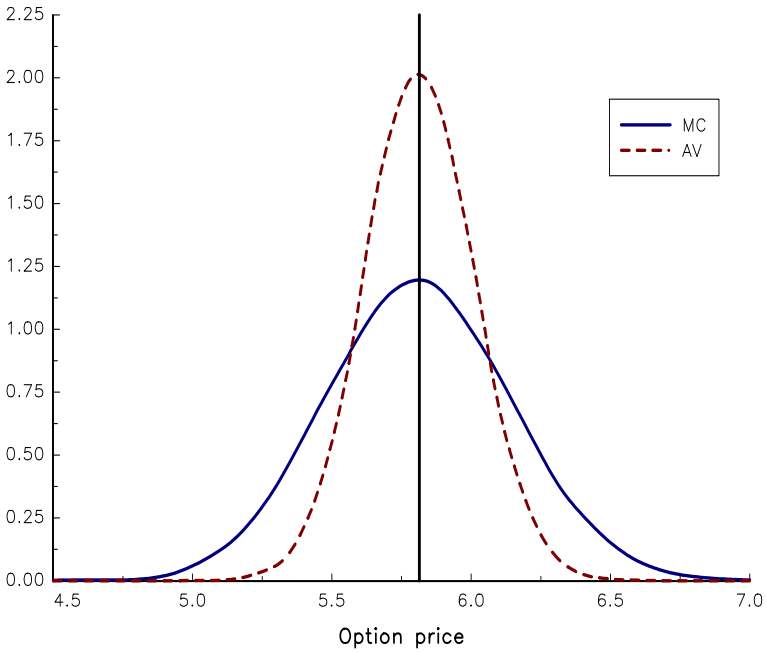


FIGURE 13.37: Probability density function of $\widehat{\mathcal{C}}_{\text{MC}}$ and $\widehat{\mathcal{C}}_{\text{AV}}$ ($n_S = 1\,000$)

13.3.2.2 Control variates

Let $Y = \varphi(X_1, \ldots, X_n)$ and V be a random variable with known mean $\mathbb{E}[V]$. We define Z as follows:

$$Z = Y + c \cdot (V - \mathbb{E}[V])$$

We deduce that:

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[Y + c \cdot (V - \mathbb{E}[V])] \\
&= \mathbb{E}[Y] + c \cdot \mathbb{E}[V - \mathbb{E}[V]] \\
&= \mathbb{E}[\varphi(X_1, \ldots, X_n)]
\end{aligned}$$

and:

$$\begin{aligned}
\text{var}(Z) &= \text{var}(Y + c \cdot (V - \mathbb{E}[V])) \\
&= \text{var}(Y) + 2 \cdot c \cdot \text{cov}(Y, V) + c^2 \cdot \text{var}(V)
\end{aligned}$$

It follows that:

$$\begin{aligned}
\text{var}(Z) \leq \text{var}(Y) &\Leftrightarrow 2 \cdot c \cdot \text{cov}(Y, V) + c^2 \cdot \text{var}(V) \leq 0 \\
&\Rightarrow c \cdot \text{cov}(Y, V) \leq 0
\end{aligned}$$

In order to obtain a lower variance, a necessary condition is that c and $\text{cov}(Y,V)$ have opposite signs. The minimum is obtained when $\partial_c \text{var}(Z) = 0$ or equivalently when:

$$c^\star = -\frac{\text{cov}(Y,V)}{\text{var}(V)} = -\beta$$

The optimal value $c^\star$ is then equal to the opposite of the beta of Y with respect to the control variate V. In this case, we have:

$$Z = Y - \frac{\text{cov}(Y,V)}{\text{var}(V)} \cdot (V - \mathbb{E}[V])$$

and:

$$\begin{aligned}\text{var}(Z) &= \text{var}(Y) - \frac{\text{cov}^2(Y,V)}{\text{var}(V)} \\ &= \left(1 - \rho^2 \langle Y,V \rangle\right) \cdot \text{var}(Y)\end{aligned}$$

This implies that we have to choose a control variate V that is highly (positively or negatively) correlated with Y in order to reduce the variance.

Example 154 *We consider that* $X \sim \mathcal{U}_{[0,1]}$ *and* $\varphi(x) = e^x$. *We would like to estimate:*

$$I = \mathbb{E}[\varphi(X)] = \int_0^1 e^x \,\mathrm{d}x$$

We set $Y = e^X$ *and* $V = X$. *We know that* $\mathbb{E}[V] = 1/2$ *and* $\text{var}(V) = 1/12$. *It follows that:*

$$\begin{aligned}\text{var}(Y) &= \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y] \\ &= \int_0^1 e^{2x} \,\mathrm{d}x - \left(\int_0^1 e^x \,\mathrm{d}x\right)^2 \\ &= \left[\frac{e^{2x}}{2}\right]_0^1 - \left(e^1 - e^0\right)^2 \\ &= \frac{4e - e^2 - 3}{2} \\ &\approx 0.2420\end{aligned}$$

and:

$$\begin{aligned}\text{cov}(Y,V) &= \mathbb{E}[VY] - \mathbb{E}[V]\,\mathbb{E}[Y] \\ &= \int_0^1 x e^x \,\mathrm{d}x - \frac{1}{2}\left(e^1 - e^0\right) \\ &= \left[x e^x\right]_0^1 - \int_0^1 e^x \,\mathrm{d}x - \frac{1}{2}\left(e^1 - e^0\right) \\ &= \frac{3 - e}{2} \\ &\approx 0.1409\end{aligned}$$

If we consider the VC estimator Z *defined by*[36]*:*

$$\begin{aligned}Z &= Y - \frac{\text{cov}(Y,V)}{\text{var}(V)} \cdot (V - \mathbb{E}[V]) \\ &= Y - (18 - 6e) \cdot \left(V - \frac{1}{2}\right)\end{aligned}$$

[36] We have $\beta \approx 1.6903$.

we obtain:

$$\begin{aligned}\text{var}(Z) &= \text{var}(Y) - \frac{\text{cov}^2(Y,V)}{\text{var}(V)} \\ &= \frac{4e - e^2 - 3}{2} - 3 \cdot (3-e)^2 \\ &\approx 0.0039\end{aligned}$$

We conclude that we have dramatically reduced the variance of the estimator, because we have:

$$\frac{\text{var}\left(\hat{I}_{\text{CV}}\right)}{\text{var}\left(\hat{I}_{\text{MC}}\right)} = \frac{\text{var}(Z)}{\text{var}(Y)} = 1.628\%$$

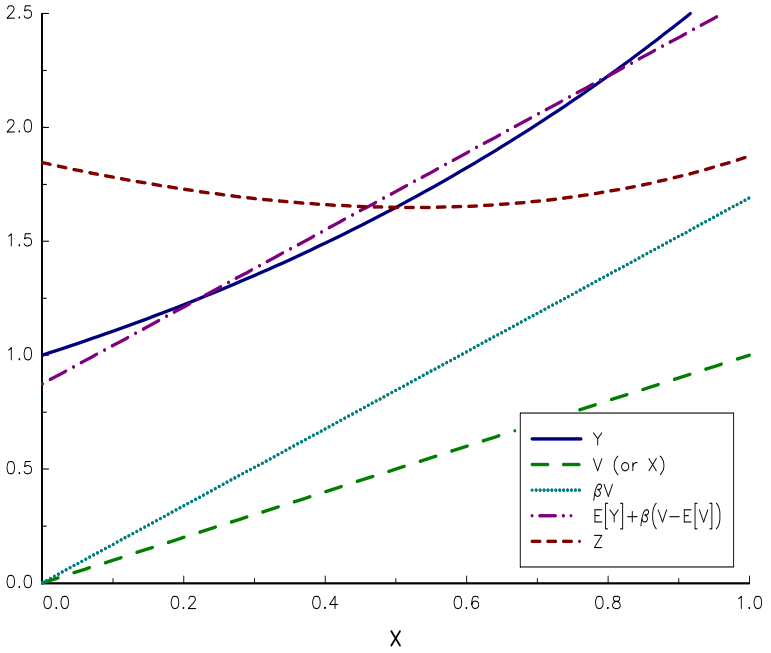


FIGURE 13.38: Understanding the variance reduction in control variates

This example may be disturbing, because the variance reduction is huge. To understand the mechanisms underlying control variates, we illustrate the previous example in Figure 13.38. For each variable, we have represented the relationship with respect to the random variable X. We have $Y = \exp(X)$ and $V = X$. To maximize the dependence between Y and the control variate, it is better to consider βV instead of V. However, the random variable βV is not well located, because it does not fit well Y. This is not the case of $\hat{Y} = \mathbb{E}[Y] + \beta(V - \mathbf{E}[V])$. Indeed, $\hat{Y}$ is the conditional expectation of Y with respect to V:

$$\mathbb{E}[Y \mid V] = \mathbb{E}[Y] + \beta(V - \mathbf{E}[V])$$

This is the best linear estimator of Y. The residual U of the linear regression is then equal to:

$$\begin{aligned}U &= Y - \hat{Y} \\ &= (Y - \mathbb{E}[Y]) - \beta(V - \mathbf{E}[V])\end{aligned}$$

The CV estimator Z is a translation of the residual in order to satisfy $\mathbb{E}[Z] = \mathbb{E}[Y]$:

$$\begin{aligned} Z &= \mathbb{E}[Y] + U \\ &= Y - \beta(V - \mathbf{E}[V]) \end{aligned}$$

By construction, the variance of the residual U is lower than the variance of the random variable Y. We conclude that:

$$\operatorname{var}(Z) = \operatorname{var}(U) \leq \operatorname{var}(Y)$$

We can therefore obtain a large variance reduction if the following conditions are satisfied:

- the control variate V largely explains the random variable Y;
- the relationship between Y and V is almost linear.

In the previous example, these conditions are largely satisfied and the residuals are very small[37].

Remark 165 *In practice, we don't know the optimal value $c^\star$. However, the previous framework helps us to estimate it. Indeed, we have:*

$$c^\star = -\hat{\beta}$$

where $\hat{\beta}$ is the OLS estimate associated to the linear regression model:

$$Y_s = \alpha + \beta V_s + u_s$$

Because Y_s and V_s are the simulated values of Y and V, this implies that $c^\star$ is calculated at the final step of the Monte Carlo method.

We recall that the price of an arithmetic Asian call option is given by:

$$\mathcal{C} = e^{-rT}\mathbb{E}\left[\left(\bar{S} - K\right)^+\right]$$

where K is the strike of the option and $\bar{S}$ denotes the average of $S(t)$ on a given number of fixing dates[38] $\{t_1, \ldots, t_{n_F}\}$:

$$\bar{S} = \frac{1}{n_F}\sum_{m=1}^{n_F} S(t_m)$$

We can estimate the option price using the Black-Scholes model. We can also reduce the variance of the MC estimator by considering the following control variates:

1. the terminal value $V_1 = S(T)$ of the underlying asset;
2. the average value $V_2 = \bar{S}$;
3. the discounted payoff of the call option $V_3 = e^{-rT}(S(T) - K)^+$;
4. the discounted payoff of the geometric Asian call option $V_4 = e^{-rT}\left(\tilde{S} - K\right)^+$ where:

$$\tilde{S} = \left(\prod_{m=1}^{n_F} S(t_m)\right)^{1/n_F}$$

[37] The variance of residuals represents 1.628% of the variance of Y.
[38] We have $t_{n_F} = T$.

For these control variates, we know the expected value. In the first and second cases, we have:

$$\mathbb{E}\left[S\left(T\right)\right] = S_0 e^{rT}$$

and:

$$\mathbb{E}\left[\bar{S}\right] = \frac{S_0}{n_F}\sum_{m=1}^{n_F} e^{rt_m}$$

The expected value of the third control variate is the Black-Scholes formula of the European call option. For the last control variate, we have:

$$\begin{aligned}\tilde{S} &= \left(\prod_{m=1}^{n_F} S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)t_m+\sigma W(t_m)}\right)^{1/n_F} \\ &= S_0 \cdot \exp\left(\left(r-\frac{1}{2}\sigma^2\right)\bar{t}+\sigma\bar{W}\right)\end{aligned}$$

where:

$$\bar{t} = \frac{1}{n_F}\sum\nolimits_{m=1}^{n_F} t_m$$

and:

$$\bar{W} = \frac{1}{n_F}\sum\nolimits_{m=1}^{n_F} W\left(t_m\right)$$

Because $\tilde{S}$ has a log-normal distribution, we deduce that the expected value of the fourth control variate is also given by a Black-Scholes formula[39]. We consider the following parameters $S_0 = 100$, $K = 104$, $r = 5\%$, $\sigma = 20\%$ and $T = 5$. The fixing dates of the Asian option are $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ and $t_5 = 5$. In top panels in Figure 13.39, we report the probability density function of the MC estimator $\hat{\boldsymbol{\mathcal{C}}}_{\text{MC}}$ and the CV estimator $\hat{\boldsymbol{\mathcal{C}}}_{\text{CV}}$ when the number of simulations is equal to 1 000. The variance ratio $\text{var}\left(\hat{\boldsymbol{\mathcal{C}}}_{\text{CV}}\right) / \text{var}\left(\hat{\boldsymbol{\mathcal{C}}}_{\text{MC}}\right)$ is respectively equal to 22.6% for $V_1 = S\left(T\right)$, 9.4% for $V_2 = \bar{S}$, 19.5% for $V_3 = e^{-rT}\left(S\left(T\right)-K\right)^+$ and 0.5% for $V_4 = e^{-rT}\left(\tilde{S}-K\right)^+$. In bottom panels in Figure 13.39, we also show the relationship between the simulated value $Y = e^{-rT}\left(\bar{S}-K\right)^+$ and the control variates V_1 and V_4. We verify that the linear regression produces lower residuals for V_4 than for V_1.

[39] We have:

$$\mathbb{E}\left[\ln\tilde{S}\right] = \ln S_0 + \left(r-\frac{1}{2}\sigma^2\right)\bar{t}$$

and:

$$\text{var}\left(\ln\tilde{S}\right) = \sigma^2\upsilon$$

where:

$$\begin{aligned}\upsilon &= \text{var}\left(\frac{1}{n_F}\sum\nolimits_{m=1}^{n_F} W\left(t_m\right)\right) \\ &= \frac{1}{n_F^2}\mathbb{E}\left[\sum\nolimits_{m=1}^{n_F} W^2\left(t_m\right) + 2\sum\nolimits_{m'>m} W\left(t_{m'}\right)W\left(t_m\right)\right] \\ &= \frac{1}{n_F^2}\left(\sum\nolimits_{m=1}^{n_F} t_m + 2\sum\nolimits_{m=1}^{n_F}\left(n_F-m\right)t_m\right)\end{aligned}$$

We deduce that:

$$\mathbb{E}\left[e^{-rT}\left(\tilde{S}-K\right)^+\right] = S_0 e^{\gamma-rT}\Phi\left(d+\sigma\sqrt{\upsilon}\right) - Ke^{-rT}\Phi\left(d\right)$$

where:

$$d = \frac{1}{\sigma\sqrt{\upsilon}}\left(\ln\frac{S_0}{K} + \left(r-\frac{1}{2}\sigma^2\right)\bar{t}\right)$$

and:

$$\gamma = r\bar{t} + \frac{1}{2}\sigma^2\left(\upsilon-\bar{t}\right)$$

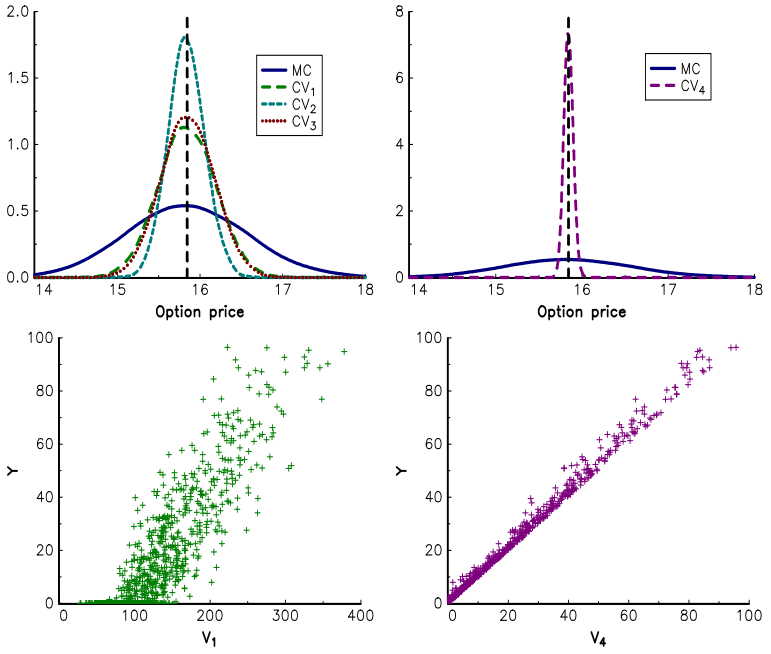


FIGURE 13.39: CV estimator of the arithmetic Asian call option

The previous approach can be extended in the case of several control variates:

$$\begin{aligned} Z &= Y + \sum_{i=1}^{n_{CV}} c_i \cdot (V_i - \mathbb{E}[V_i]) \\ &= Y + c^\top (V - \mathbb{E}[V]) \end{aligned}$$

where $c = (c_1, \ldots, c_{n_{CV}})$ and $V = (V_1, \ldots, V_{n_{CV}})$. We can show that the optimal value of c is equal to:

$$c^\star = -\operatorname{cov}(V, V)^{-1} \cdot \operatorname{cov}(V, Y)$$

By noting that minimizing the variance of Z is equivalent to minimize the variance of U where:

$$\begin{aligned} U &= Y - \hat{Y} \\ &= Y - \left(\alpha + \beta^\top V\right) \end{aligned}$$

we deduce that $c^\star = -\beta$. It follows that

$$\begin{aligned} \operatorname{var}(Z) &= \operatorname{var}(U) \\ &= \left(1 - R^2\right) \cdot \operatorname{var}(Y) \end{aligned}$$

where R^2 is the R-squared coefficient of the linear regression $Y = \alpha + \beta^\top V + U$.

Let us consider the previous example of the arithmetic Asian call option. In Table 13.5, we give the results of the linear regression by considering the combination of the four control variates. Previously, we found that the variance ratio was equal to 9.4% for the second control variate. If we combine the first three variates, this ratio becomes 3.5%. With the four control variates, the variance of the Monte Carlo estimator is divided by a factor of 500!

TABLE 13.5: Linear regression between the Asian call option and the control variates

$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2	$1-R^2$
−51.482	0.036	0.538			90.7%	9.3%
−24.025	−0.346	0.595	0.548		96.5%	3.5%
−4.141	0.069		0.410		81.1%	18.9%
−38.727		0.428	0.174		92.9%	7.1%
−1.559	−0.040	0.054	0.111	0.905	99.8%	0.2%

Remark 166 *The reader may consult the book of Lamberton and Lapeyre (2007) for other examples of control variates in option pricing. In particular, they show how to use the put-call parity formula for reducing the volatility by noting that the variance of put options are generally smaller than the variance of call options.*

13.3.2.3 Importance sampling

Let $X = (X_1, \ldots, X_n)$ be a random vector with distribution function $\mathbf{F}$. We have:

$$\begin{aligned} I &= \mathbb{E}\left[\varphi\left(X_1, \ldots, X_n\right) \mid \mathbf{F}\right] \\ &= \int \cdots \int \varphi\left(x_1, \ldots, x_n\right) f\left(x_1, \ldots, x_n\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{aligned}$$

where $f\left(x_1, \ldots, x_n\right)$ is the probability density function of X. It follows that:

$$\begin{aligned} I &= \int \cdots \int \left(\varphi\left(x_1, \ldots, x_n\right) \frac{f\left(x_1, \ldots, x_n\right)}{g\left(x_1, \ldots, x_n\right)}\right) g\left(x_1, \ldots, x_n\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \mathbb{E}\left[\varphi\left(X_1, \ldots, X_n\right) \frac{f\left(X_1, \ldots, X_n\right)}{g\left(X_1, \ldots, X_n\right)} \middle| \mathbf{G}\right] \\ &= \mathbb{E}\left[\varphi\left(X_1, \ldots, X_n\right) \mathcal{L}\left(X_1, \ldots, X_n\right) \mid \mathbf{G}\right] \end{aligned} \tag{13.9}$$

where $g\left(x_1, \ldots, x_n\right)$ is the probability density function of $\mathbf{G}$ and $\mathcal{L}$ is the likelihood ratio:

$$\mathcal{L}\left(x_1, \ldots, x_n\right) = \frac{f\left(x_1, \ldots, x_n\right)}{g\left(x_1, \ldots, x_n\right)}$$

The values taken by $\mathcal{L}\left(x_1, \ldots, x_n\right)$ are also called the importance sampling weights. Using the vector notation, the relationship (13.9) becomes:

$$\mathbb{E}\left[\varphi\left(X\right) \mid \mathbf{F}\right] = \mathbb{E}\left[\varphi\left(X\right) \mathcal{L}\left(X\right) \mid \mathbf{G}\right]$$

It follows that:

$$\mathbb{E}\left[\hat{I}_{\mathrm{MC}}\right] = \mathbb{E}\left[\hat{I}_{\mathrm{IS}}\right] = I$$

where $\hat{I}_{\mathrm{MC}}$ and $\hat{I}_{\mathrm{IS}}$ are the Monte Carlo and importance sampling estimators of I. We also deduce that[40]:

$$\mathrm{var}\left(\hat{I}_{\mathrm{IS}}\right) = \mathrm{var}\left(\varphi\left(X\right) \mathcal{L}\left(X\right) \mid \mathbf{G}\right)$$

[40] Recall that we use the vector notation, meaning that $x = (x_1, \ldots, x_n)$.

It follows that:

$$\begin{aligned}
\operatorname{var}\left(\hat{I}_{\mathrm{IS}}\right) &= \mathbb{E}\left[\varphi^2(X)\mathcal{L}^2(X) \mid \mathbf{G}\right] - \mathbb{E}^2\left[\varphi(X)\mathcal{L}(X) \mid \mathbf{G}\right] \\
&= \int \varphi^2(x)\mathcal{L}^2(x)g(x)\,\mathrm{d}x - I^2 \\
&= \int \varphi^2(x)\frac{f^2(x)}{g^2(x)}g(x)\,\mathrm{d}x - I^2 \\
&= \int \varphi^2(x)\frac{f^2(x)}{g(x)}\,\mathrm{d}x - I^2
\end{aligned} \tag{13.10}$$

If we compare the variance of the two estimators $\hat{I}_{\mathrm{MC}}$ and $\hat{I}_{\mathrm{IS}}$, we obtain:

$$\begin{aligned}
\operatorname{var}\left(\hat{I}_{\mathrm{IS}}\right) - \operatorname{var}\left(\hat{I}_{\mathrm{MC}}\right) &= \int \varphi^2(x)\frac{f^2(x)}{g(x)}\,\mathrm{d}x - \int \varphi^2(x)f(x)\,\mathrm{d}x \\
&= \int \varphi^2(x)\left(\frac{f(x)}{g(x)} - 1\right)f(x)\,\mathrm{d}x \\
&= \int \varphi^2(x)\left(\mathcal{L}(x) - 1\right)f(x)\,\mathrm{d}x
\end{aligned}$$

The difference may be negative if the weights $\mathcal{L}(x)$ are small ($\mathcal{L}(x) \ll 1$) because the values of $\varphi^2(x)f(x)$ are positive. The importance sampling approach changes then the importance of some values x by transforming the original probability distribution $\mathbf{F}$ into another probability distribution $\mathbf{G}$. Equation (13.10) is also interesting because it gives us some insights about the optimal IS distribution[41]:

$$\begin{aligned}
g^\star(x) &= \arg\min \operatorname{var}\left(\hat{I}_{\mathrm{IS}}\right) \\
&= \arg\min \int \varphi^2(x)\frac{f^2(x)}{g(x)}\,\mathrm{d}x \\
&= c \cdot |\varphi(x)| \cdot f(x)
\end{aligned}$$

where c is the normalizing constant such that $\int g^\star(x)\,\mathrm{d}x = 1$. A good choice of the IS density $g(x)$ is then an approximation of $|\varphi(x)| \cdot f(x)$ such that $g(x)$ can easily be simulated.

Remark 167 *In order to simplify the notation and avoid confusions, we consider that* $X \sim \mathbf{F}$ *and* $Z \sim \mathbf{G}$ *in the sequel. This means that* $\hat{I}_{\mathrm{MC}} = \varphi(X)$ *and* $\hat{I}_{\mathrm{IS}} = \varphi(Z)\mathcal{L}(Z)$.

We consider the estimation of the probability $p = \Pr\{X \geq 3\}$ when $X \sim \mathcal{N}(0,1)$. We have:

$$\varphi(x) = \mathbb{1}\{x \geq 3\}$$

Because the probability p is low ($\Pr\{X \geq 3\} \approx 0.1350\%$), the MC estimator will not be efficient. Indeed, it will be rare to simulate a random variate greater than 3. To reduce the variance of the MC estimator, we can use important sampling with $Z \sim \mathcal{N}\left(\mu_z, \sigma_z^2\right)$. For $\mu_z = 3$ and $\sigma_z = 1$, we report in Figure 13.40 the histogram of the estimators[42] $\hat{p}_{\mathrm{MC}}$ and

[41] The first-order condition is:

$$-\varphi^2(x) \cdot \frac{f^2(x)}{g^2(x)} = \lambda$$

where λ is a constant.

[42] We have:

$$\hat{p}_{\mathrm{MC}} = \frac{1}{n_S}\sum_{s=1}^{n_S} \mathbb{1}\{X_s \geq 3\}$$

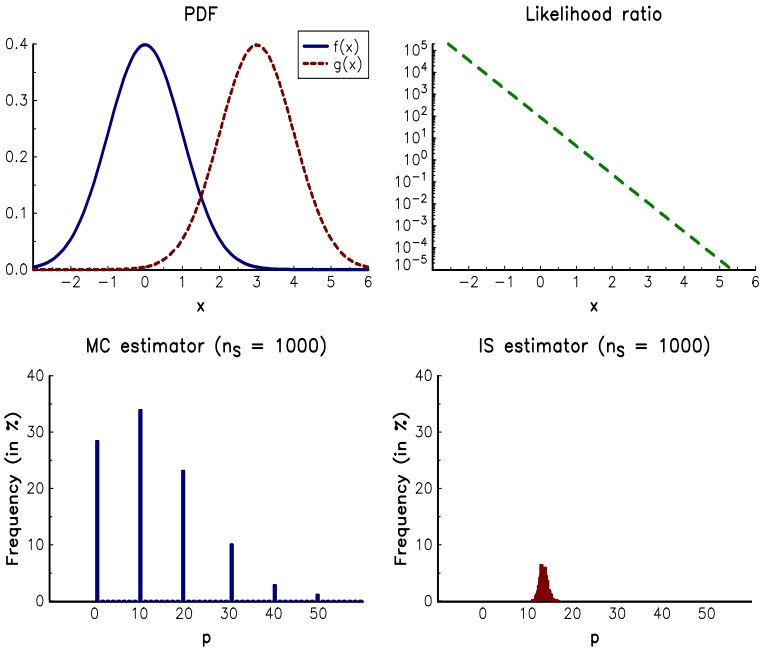


FIGURE 13.40: Histogram of the MC and IS estimators ($n_S = 1\,000$)

$\hat{p}_{\text{IS}}$ when the number of simulations is equal to 1 000. It is obvious that the IS estimator is better than the MC estimator. To explain that, we report the probability density function of X and Z in the top/left panel in Figure 13.40. Whereas $\Pr\{X \geq 3\}$ is close to zero, the probability $\Pr\{Z \geq 3\}$ is equal to 50%. Therefore, it is easier to simulate $Z \geq 3$, but we have to apply a correction to obtain the right probability. This correction is given by the likelihood ratio, which is represented in the top/right panel. In Figure 13.41, we show the standard deviation $\sigma(\hat{p}_{\text{IS}})$ for different values of μ_z and σ_z. When $\sigma_z = 1$ and $\mu_z \in [0, 5]$, it is lower than the standard deviation of $\hat{p}_{\text{MC}}$. For $\mu_z = 3$, the variance ratio is approximately equal to 1% meaning that the variance of $\hat{p}_{\text{MC}}$ is divided by a factor of 100. We also notice that we reduce the variance by using a higher value of σ_z. In fact, we can anticipate that the IS estimator is more efficient than the MC estimator if the following condition holds:

$$\Pr\{Z \geq 3\} \geq \Pr\{X \geq 3\}$$

The calculation of the optimal values of μ_z and σ_z is derived in Exercise 13.4.9 on page 891.

---

and:

$$\hat{p}_{\text{IS}} = \frac{1}{n_S} \sum_{s=1}^{n_S} \mathbb{1}\{Z_s \geq 3\} \cdot \mathcal{L}(Z_s)$$

where:

$$\mathcal{L}(z) = \sigma_z \exp\left(\frac{1}{2}\left(\frac{z-\mu_z}{\sigma_z}\right)^2 - \frac{1}{2}z^2\right)$$

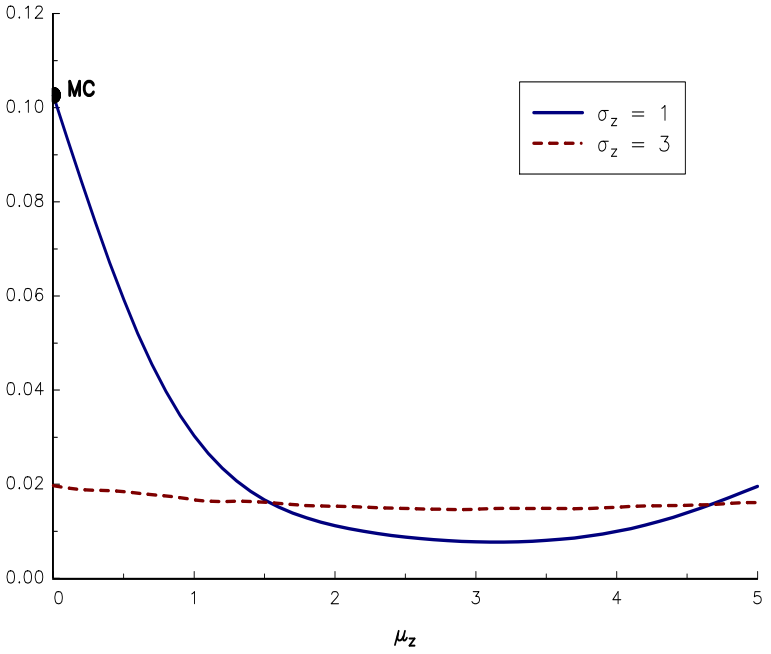


FIGURE 13.41: Standard deviation (in %) of the estimator $\hat{p}_{\text{IS}}$ ($n_S = 1\,000$)

Remark 168 *The previous example is an illustration of rare event simulation. This is why importance sampling is related to the theory of large deviations. Many results of this statistical field (Cramer's theorem, Berry-Esseen bounds) are then obtained using the same approach than the importance sampling method.*

We consider the pricing of the put option:

$$\boldsymbol{\mathcal{P}} = e^{-rT}\mathbb{E}\left[(K - S(T))^+\right]$$

We can estimate the option price by using the Monte Carlo method with:

$$\varphi(x) = e^{-rT}(K - x)^+$$

In the case where $K \ll S(0)$, the probability of exercise $\Pr\{S(T) \leq K\}$ is very small. Therefore, we have to increase the probability of exercise in order to obtain a more efficient estimator. In the case of the Black-Scholes model, the density function of $S(T)$ is equal to:

$$f(x) = \frac{1}{x\sigma_x}\phi\left(\frac{\ln x - \mu_x}{\sigma_x}\right)$$

where $\mu_x = \ln S_0 + \left(r - \sigma^2/2\right)T$ and $\sigma_x = \sigma\sqrt{T}$. Using the same approach than previously, we consider the IS density $g(x)$ defined by:

$$g(x) = \frac{1}{x\sigma_z}\phi\left(\frac{\ln x - \mu_z}{\sigma_z}\right)$$

where $\mu_z = \theta + \mu_x$ and $\sigma_z = \sigma_x$. For instance, we can choose θ such that the probability of exercise is equal to 50%. It follows that:

$$\begin{aligned}\Pr\{Z \leq K\} = \frac{1}{2} &\Leftrightarrow \Phi\left(\frac{\ln K - \theta - \mu_x}{\sigma_x}\right) = \frac{1}{2} \\ &\Leftrightarrow \theta = \ln K - \mu_x \\ &\Leftrightarrow \theta = \ln \frac{K}{S_0} - \left(r - \frac{1}{2}\sigma^2\right) T\end{aligned}$$

We deduce that:

$$\begin{aligned}\boldsymbol{\mathcal{P}} &= \mathbb{E}\left[\varphi\left(S\left(T\right)\right)\right] \\ &= \mathbb{E}\left[\varphi\left(S'\left(T\right)\right) \cdot \mathcal{L}\left(S'\left(T\right)\right)\right]\end{aligned}$$

where:

$$\begin{aligned}\mathcal{L}\left(x\right) &= \frac{\frac{1}{x\sigma_x}\phi\left(\frac{\ln x - \mu_x}{\sigma_x}\right)}{\frac{1}{x\sigma_z}\phi\left(\frac{\ln x - \mu_z}{\sigma_z}\right)} \\ &= \exp\left(\frac{\theta^2}{2\sigma_x^2} - \left(\frac{\ln x - \mu_x}{\sigma_x}\right) \cdot \frac{\theta}{\sigma_x}\right)\end{aligned}$$

and $S'\left(T\right)$ is the same geometric Brownian motion than $S\left(T\right)$, but with another initial value:

$$S'\left(0\right) = S\left(0\right) e^{\theta} = Ke^{-\left(r-\sigma^2/2\right)T}$$

Example 155 *We assume that* $S_0 = 100$, $K = 60$, $r = 5\%$, $\sigma = 20\%$ *and* $T = 2$. *If we consider the previous method, the IS process is simulated using the initial value* $S'\left(0\right) = Ke^{-\left(r-\sigma^2/2\right)T} = 56.506$, *whereas the value of* θ *is equal to* -0.5708. *In Figure 13.42, we report the density function of the estimators* $\hat{\boldsymbol{\mathcal{P}}}_{\text{MC}}$ *and* $\hat{\boldsymbol{\mathcal{P}}}_{\text{IS}}$ *when the number of simulations is equal to* 1 000. *For this example, the variance ratio is equal to* 1.77%, *meaning that the IS method has reduced the variance of the MC estimator by a factor greater than* 50. *If we use another IS scheme with* $S'\left(0\right) = 80$, *the reduction is less important, but remains significant*[43].

13.3.2.4 Other methods

We mention two other methods, which are less used in risk management than the previous methods, but may be very efficient for some financial problems. The first method is known as the conditional Monte Carlo method. Recall that $I = \mathbb{E}\left[Y\right]$ where $Y = \varphi\left(X_1, \ldots, X_n\right)$. Let Z be a random vector and $V = \mathbb{E}\left[Y \mid Z\right]$ be the conditional expectation of Y with respect to Z. It follows that:

$$\begin{aligned}\mathbb{E}\left[V\right] &= \int \mathbb{E}\left[Y \mid Z\right] f_z\left(z\right) \mathrm{d}z \\ &= \mathbb{E}\left[Y\right] \\ &= I\end{aligned}$$

[43] The variance ratio is equal to 13.59%.

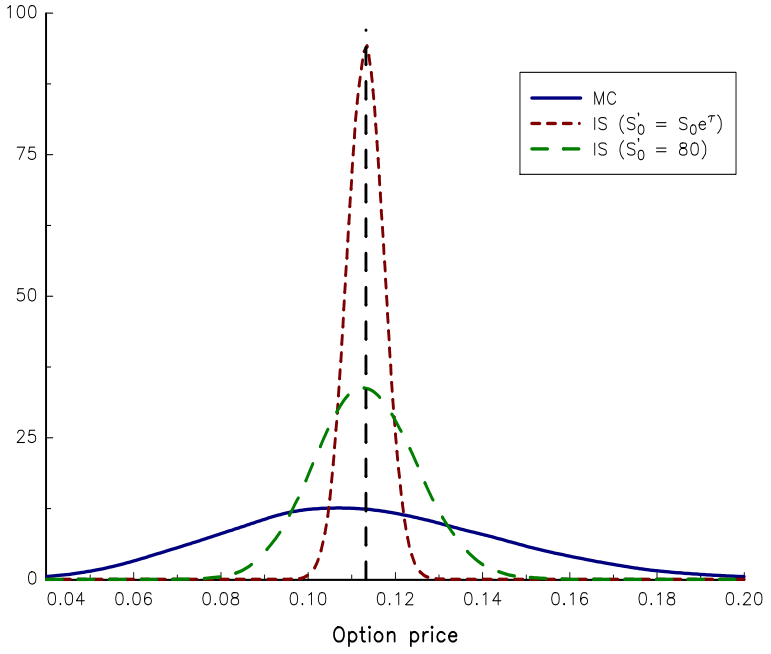


FIGURE 13.42: Density function of the estimators $\hat{\mathcal{P}}_{\text{MC}}$ and $\hat{\mathcal{P}}_{\text{IS}}$ $(n_S = 1\,000)$

where f_z is the probability density function of Z. Recall that:

$$\text{var}(Y) = \mathbb{E}[\text{var}(Y \mid Z)] + \text{var}(\mathbb{E}[Y \mid Z])$$

We deduce that:

$$\begin{aligned} \text{var}(V) &= \text{var}(\mathbb{E}[Y \mid Z]) \\ &= \text{var}(Y) - \mathbb{E}[\text{var}(Y \mid Z)] \\ &\leq \text{var}(Y) \end{aligned}$$

because $\text{var}(Y \mid Z) \geq 0$ implies that $\mathbb{E}[\text{var}(Y \mid Z)] \geq 0$. The idea of the conditional Monte Carlo method is then simulating V instead of Y in order to reduce the variance. For that, we have to find Z such that $\mathbb{E}[Y \mid Z]$ can easily be sampled. It can be the case with some stochastic volatility models.

Example 156 *Let $X = (X_1, X_2)$ be a standardized Gaussian random vector with correlation ρ. We want to calculate $p = \Pr\{X_1 \leq aX_2 + b\}$. We have:*

$$\begin{aligned} Y &= \varphi(X_1, X_2) \\ &= \mathbb{1}\{X_1 \leq aX_2 + b\} \end{aligned}$$

and:

$$\hat{p}_{\text{MC}} = \frac{1}{n_S} \sum_{s=1}^{n_S} \mathbb{1}\{X_{1,s} \leq aX_{2,s} + b\}$$

If we consider $Z = X_2$, we obtain:

$$\begin{aligned} V &= \mathbb{E}[Y \mid Z] \\ &= \mathbb{E}[\mathbb{1}\{X_1 \leq aX_2 + b\} \mid X_2 = x_2] \end{aligned}$$

Because we have $X_2 = \rho X_1 + \sqrt{1-\rho^2} X_3$ *where* $X_3 \sim \mathcal{N}(0,1)$ *is independent from* X_1*, we deduce that:*

$$\begin{aligned} V &= \mathbb{E}\left[\mathbb{1}\left\{X_1 \leq a\left(\rho X_1 + \sqrt{1-\rho^2} X_3\right) + b\right\} \mid X_3 = x_3\right] \\ &= \Phi\left(\frac{a\sqrt{1-\rho^2} x_3 + b}{1 - a\rho}\right) \end{aligned}$$

The conditional Monte Carlo (CMC) estimator is then equal to[44]*:*

$$\hat{p}_{\text{CMC}} = \frac{1}{n_S} \sum_{s=1}^{n_S} \Phi\left(\frac{a\sqrt{1-\rho^2} X_{3,s} + b}{1 - a\rho}\right)$$

where $X_{3,s} \sim \mathcal{N}(0,1)$*. In Table 13.6, we report the variance ratio between the CMC and MC estimators when* a *is equal to* 1*. We verify that the CMC estimator is particularly efficient when* ρ *is negative. For instance, the variance is divided by a factor of* 70 *when* ρ *is equal to* -90% *and* b *is equal to* 3.0*.*

TABLE 13.6: Variance ratio (in %) when $a = 1$

b	Correlation ρ (in %)								
	-90.0	-75.0	-50.0	-25.0	0.0	25.0	50.0	75.0	90.0
0.0	3.2	8.0	16.1	24.5	33.3	43.0	54.0	67.8	79.8
1.0	2.9	7.3	14.8	22.5	30.6	39.3	48.9	60.0	67.6
2.0	2.2	5.6	11.3	17.2	23.3	29.6	35.9	41.9	48.8
3.0	1.4	3.4	7.0	10.7	14.4	18.1	21.4	24.6	

The second method is the stratified sampling. Recall that $X \in \Omega$. Let $\{\Omega_j, j = 1, \ldots, m\}$ be a partition[45] of Ω. We have:

$$\begin{aligned} I &= \mathbb{E}[\varphi(X)] \\ &= \int_\Omega \varphi(x) f(x)\, \mathrm{d}x \\ &= \sum_{j=1}^m \int_{\Omega_j} \varphi(x) f(x)\, \mathrm{d}x \\ &= \sum_{j=1}^m \mathbb{E}[\mathbb{1}\{X \in \Omega_j\} \cdot \varphi(X)] \\ &= \sum_{j=1}^m \Pr\{X \in \Omega_j\} \cdot \mathbb{E}[\varphi(X) \mid X \in \Omega_j] \end{aligned}$$

We introduce the index random variable B:

$$B = j \Leftrightarrow X \in \Omega_j$$

[44] If $a\rho = 1$, we have:

$$\hat{p}_{\text{CMC}} = \frac{1}{n_S} \sum_{s=1}^{n_S} \mathbb{1}\left\{a\sqrt{1-\rho^2} X_{3,s} + b \geq 0\right\} = \hat{p}_{\text{MC}}$$

[45] This means that $\Omega_j \bigcap \Omega_k = \emptyset$ and $\bigcup_{j=1}^m \Omega_j = \Omega$.

We note $p(j) = \Pr\{B = j\} = \Pr\{X \in \Omega_j\}$ and $X(j)$ the random vector, whose probability distribution is the conditional law of $X \mid X \in \Omega_j$. It follows that:

$$\begin{aligned} I &= \sum_{j=1}^{m} p(j) \cdot \mathbb{E}[\varphi(X(j))] \\ &= \sum_{j=1}^{m} p(j) \cdot I(j) \end{aligned}$$

where $I(j) = \mathbb{E}[\varphi(X(j))] = \mathbb{E}[\varphi(X) \mid B = j]$. We define the stratified sampling estimator as follows:

$$\hat{I}_{\text{STR}} = \sum_{j=1}^{m} p(j) \cdot \hat{Y}(j)$$

where:

$$\hat{Y}(j) = \frac{1}{n_S(j)} \sum_{s=1}^{n_S(j)} Y_s(j) = \frac{1}{n_S(j)} \sum_{s=1}^{n_S(j)} \varphi(X_s(j))$$

Recall that the MC estimator is equal to:

$$\hat{I}_{\text{MC}} = \hat{Y}$$

where:

$$\hat{Y} = \frac{1}{n_S} \sum_{s=1}^{n_S} Y_s = \frac{1}{n_S} \sum_{s=1}^{n_S} \varphi(X_s)$$

The MC estimator can be viewed as a stratified sampling estimator with only one stratum: $\Omega_1 = \Omega$. On the contrary, the STR estimator depends on the number m of strata and the distribution of strata.

Like the MC estimator, it is easy to show that the stratified sampling estimator is unbiased[46]:

$$\mathbb{E}\left[\hat{I}_{\text{STR}}\right] = I$$

We introduce the following notations.

1. the conditional expectation $\mu(j)$ is defined as:

$$\mu(j) = \mathbb{E}[\varphi(X(j))] = \mathbb{E}[\varphi(X) \mid B = j]$$

2. the conditional variance $\sigma^2(j)$ is equal to:

$$\sigma^2(j) = \text{var}(\varphi(X(j))) = \text{var}(\varphi(X) \mid B = j)$$

Using the conditional independence of the random variables $X(j)$, it follows that:

$$\text{var}\left(\hat{I}_{\text{STR}}\right) = \sum_{j=1}^{m} \frac{p^2(j) \cdot \sigma^2(j)}{n_S(j)} \tag{13.11}$$

and:

$$\text{var}\left(\hat{I}_{\text{MC}}\right) = \frac{1}{n_S} \left(\sum_{j=1}^{m} p(j) \cdot \sigma^2(j) + \sum_{j=1}^{m} p(j) \cdot (\mu(j) - \bar{\mu})^2 \right) \tag{13.12}$$

[46] We assume that $n_S(j) \neq 0$.

where:

$$\bar{\mu} = \sum_{j=1}^{m} p(j) \cdot \mu(j)$$

Using Equations (13.11) and (13.12), it is not possible to compare directly the variance of the two estimators because the stratified sampling estimator depends on the allocation $(n_S(1), \ldots, n_S(m))$. Therefore, we can have $\text{var}\left(\hat{I}_{\text{STR}}\right) < \text{var}\left(\hat{I}_{\text{MC}}\right)$ or $\text{var}\left(\hat{I}_{\text{STR}}\right) > \text{var}\left(\hat{I}_{\text{MC}}\right)$. However, for many allocation schemes, the stratified sampling approach is an efficient method to reduce the variance of the MC estimator.

To illustrate the interest of the stratified sampling approach, we consider the proportional allocation:

$$n_S(j) = n_S \cdot p(j)$$

It follows that:

$$\text{var}\left(\hat{I}_{\text{STR}}\right) = \frac{1}{n_S} \sum_{j=1}^{m} p(j) \cdot \sigma^2(j)$$

and:

$$\begin{aligned}
\text{var}\left(\hat{I}_{\text{MC}}\right) &= \frac{1}{n_S} \sum_{j=1}^{m} p(j) \cdot \sigma^2(j) + \frac{1}{n_S} \sum_{j=1}^{m} p(j) \cdot (\mu(j) - \bar{\mu})^2 \\
&= \text{var}\left(\hat{I}_{\text{STR}}\right) + \frac{1}{n_S} \sum_{j=1}^{m} p(j) \cdot (\mu(j) - \bar{\mu})^2 \\
&\geq \text{var}\left(\hat{I}_{\text{STR}}\right)
\end{aligned}$$

Therefore, the stratified sampling estimator has a lower variance than the Monte Carlo estimator. In this case, we notice that:

$$\text{var}\left(\hat{I}_{\text{STR}}\right) = \frac{1}{n_S} \cdot \mathbb{E}\left[\text{var}\,\varphi(X) \mid B\right]$$

and:

$$\text{var}\left(\hat{I}_{\text{MC}}\right) = \frac{1}{n_S} \cdot \underbrace{\mathbb{E}\left[\text{var}\,\varphi(X) \mid B\right]}_{\text{intra-strata variance}} + \frac{1}{n_S} \cdot \underbrace{\text{var}\left(\mathbb{E}\left[\varphi(X) \mid B\right]\right)}_{\text{inter-strata variance}}$$

The stratified sampling approach with a proportional allocation consists of removing the inter-strata variance in order to keep only the intra-strata variance. This result gives some ideas about the optimal strata. Indeed, the variance reduction is high if the intra-strata variance is low.

We now consider that the strata are given. We write the allocation as follows:

$$n_S(j) = n_S \cdot q(j)$$

where the $q(j)$'s are arbitrary frequencies such that $\sum_{j=1}^{m} q(j) = 1$. To find the optimal allocation $q^\star$, we have to solve the following variance minimization problem:

$$q^\star = \arg\min \text{var}\left(\hat{I}_{\text{STR}}\right)$$

subject to the constraint $\sum_{j=1}^{m} q(j) = 1$. It follows that the Lagrange function is equal to:

$$\mathcal{L}(q; \lambda) = \frac{1}{n_S} \sum_{j=1}^{m} \frac{p^2(j) \cdot \sigma^2(j)}{q(j)} + \lambda \left(\sum_{j=1}^{m} q(j) - 1 \right)$$

We deduce that the optimal allocation is[47]:

$$q^{\star}(j) = \frac{p(j) \cdot \sigma(j)}{\sum_{j=1}^{m} p(j) \cdot \sigma(j)}$$

In this case, we obtain:

$$\begin{aligned} \text{var}\left(\hat{I}_{\text{STR}}\right) &= \frac{1}{n_S} \sum_{j=1}^{m} \frac{p^2(j) \cdot \sigma^2(j)}{q^{\star}(j)} \\ &= \frac{1}{n_S} \left(\sum_{j=1}^{m} p(j) \cdot \sigma(j) \right)^2 \end{aligned}$$

Example 157 *We have* $I = \int_0^1 \varphi(x)\,\mathrm{d}x = \mathbb{E}[\varphi(X)]$ *where* $X \sim \mathcal{U}_{[0,1]}$. *We consider the following cases for the function* $\varphi(x)$*:*

1. $\varphi(x) = x$
2. $\varphi(x) = x^2$
3. $\varphi(x) = (1 + \cos(\pi x))/2$

These three functions are reported in Figure 13.43. In Table 13.7, we give the exact value of I *and the variance of the estimators. For the MC estimator and the function* $\varphi(x) = x$*, we verify that:*

$$n_S \cdot \text{var}\left(\hat{I}_{\text{MC}}\right) = \text{var}\left(\mathcal{U}_{[0,1]}\right) = \frac{1}{12}$$

For the STR estimator, we consider fixed-space strata $X(j) \in \left[\frac{j-1}{m}, \frac{j}{m}\right]$ *implying that* $p(j) = 1/m$*. We can then simulate the conditional random variable* $X(j)$ *by using the following transformation:*

$$X(j) = \frac{j-1}{m} + \frac{U}{m}$$

where $U \sim \mathcal{U}_{[0,1]}$. *In Table 13.7, we report* $n_S \cdot \text{var}\left(\hat{I}_{\text{STR}}\right)$ *when* m *is equal to* 10. *We notice that the variance is approximately divided by* 100 *when we consider the proportional allocation* $q(j) = p(j)$. *To understand this result, we consider the function* $\varphi(x) = x$. *In this case, the variance of the stratum* j *is equal to*[48]*:*

$$\begin{aligned} \sigma^2(j) &= \mathbb{E}\left[X^2(j)\right] - \mathbb{E}^2\left[X(j)\right] \\ &= \int_{\frac{j-1}{m}}^{\frac{j}{m}} mx^2\,\mathrm{d}x - \left(\int_{\frac{j-1}{m}}^{\frac{j}{m}} mx\,\mathrm{d}x \right)^2 \\ &= \left[m \cdot \frac{x^3}{3} \right]_{\frac{j-1}{m}}^{\frac{j}{m}} - \left(\left[m \cdot \frac{x^2}{2} \right]_{\frac{j-1}{m}}^{\frac{j}{m}} \right)^2 \end{aligned}$$

[47] The first-order condition is:

$$\frac{\partial \mathcal{L}(q;\lambda)}{\partial q(j)} = -\frac{1}{n_S} \cdot \frac{p^2(j) \cdot \sigma^2(j)}{q^2(j)} + \lambda = 0$$

implying that the ratio $\dfrac{p(j) \cdot \sigma(j)}{q(j)}$ is constant.

[48] The density function of $X(j)$ is $f(x) = m$.

We deduce that:

$$\begin{aligned}\sigma^2(j) &= \frac{1}{3m^2}\left(j^3-(j-1)^3\right)-\frac{1}{4m^2}(2j-1)^2 \\ &= \frac{1}{12m^2}\end{aligned}$$

This implies that the variance of strata is equal to the variance of the uniform random variable divided by a factor of m^2*. These intra-strata variances are given in Figure 13.44. For the function* $\varphi(x) = x^2$*, the variance of strata increases with the index* j*. This is normal if we consider the graphic representation of the function* $\varphi(x)$ *in Figure 13.43. Indeed, the curvature of* $\varphi(x)$ *implies that there is more variance when* x *increases. We have also calculated the variance of the STR estimator when we use the optimal allocation* $q^\star$*, which is reported in Figure 13.45. In this case, we allocate a more important number of simulations for strata that present more variance. This is perfectly normal, because these strata are more difficult to simulate than strata with low variance. However, the gain of the optimal allocation is not very significant with respect to the proportional allocation.*

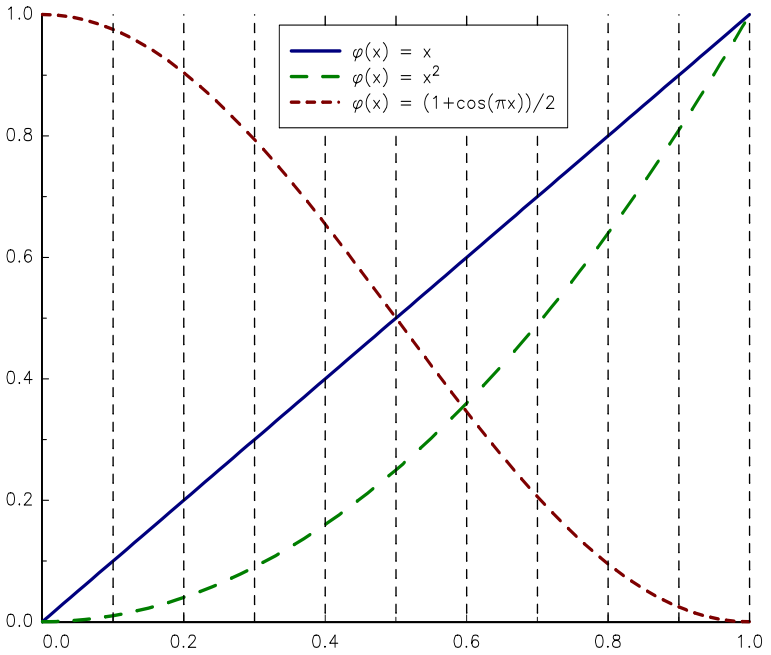


FIGURE 13.43: Function $\psi(x)$

In the case of the uniform distribution $\mathcal{U}_{[0,1]}$, we have used fixed-space strata $X(j) \in \left[\frac{j-1}{m}, \frac{j}{m}\right]$, implying that the probability $p(j)$ is equal for all the strata. This is the most popular method for defining strata. In the case of a general probability distribution $\mathbf{F}$, we define the conditional random variable $X(j)$ as follows:

$$X(j) = \mathbf{F}^{-1}\left(\frac{j-1}{m}+\frac{U}{m}\right) \tag{13.13}$$

TABLE 13.7: Comparison between MC and STR estimators

$\varphi(x)$		x	x^2	$(1+\cos(\pi x))/2$
I		0.50000	0.33333	0.50000
$n_S \text{ var}\left(\hat{I}_{\text{MC}}\right)$		0.08333	0.08890	0.12501
$n_S \text{ var}\left(\hat{I}_{\text{STR}}\right)$	$p(j)$	0.00083	0.00113	0.00105
$n_S \text{ var}\left(\hat{I}_{\text{STR}}\right)$	$q^\star(j)$	0.00083	0.00083	0.00084

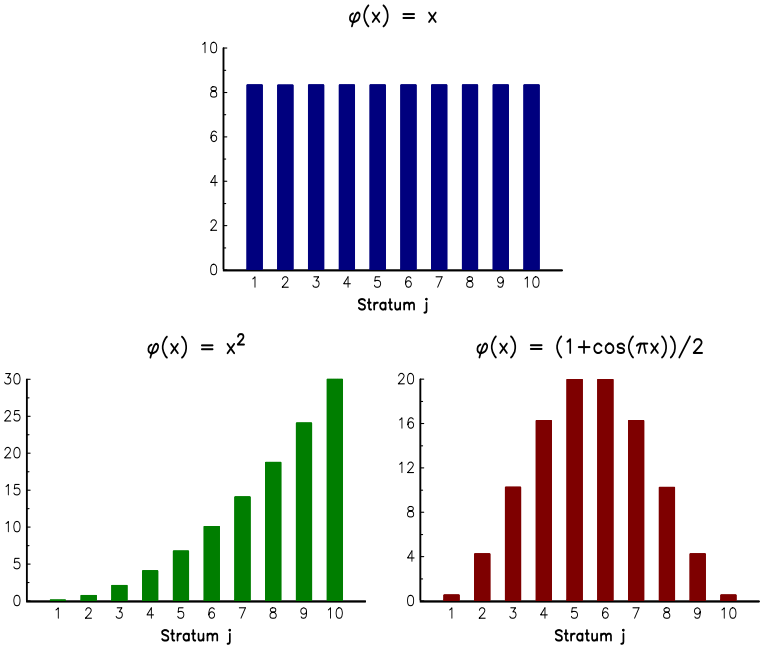


FIGURE 13.44: Intra-strata variance $\sigma^2(j)$ (in bps)

where U is a standard uniform random variate. We deduce that $X(j) \in \left[\mathbf{F}^{-1}\left(\frac{j-1}{m}\right), \mathbf{F}^{-1}\left(\frac{j}{m}\right)\right]$ and:

$$\begin{aligned} p(j) &= \Pr\left\{X \in \left[\mathbf{F}^{-1}\left(\frac{j-1}{m}\right), \mathbf{F}^{-1}\left(\frac{j}{m}\right)\right]\right\} \\ &= \mathbf{F}\left(\mathbf{F}^{-1}\left(\frac{j}{m}\right)\right) - \mathbf{F}\left(\mathbf{F}^{-1}\left(\frac{j-1}{m}\right)\right) \\ &= \frac{1}{m} \end{aligned}$$

In Figure 13.46, we have reported the strata defined by Equation (13.13) for different probability distribution when m is equal to 10.

The previous method consists in defining strata in order to obtain equal probabilities $p(j)$. It can be very different than the optimal method, whose objective is to define strata such that the intra-strata variances $\sigma(j)$ are close to zero. In order to illustrate the two approaches, we consider the pricing of an European call option in the Black-Scholes model. Recall that the price of the call option is equal to:

$$\mathcal{C} = e^{-rT}\mathbb{E}\left[\max\left(0, S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}X} - K\right)\right]$$

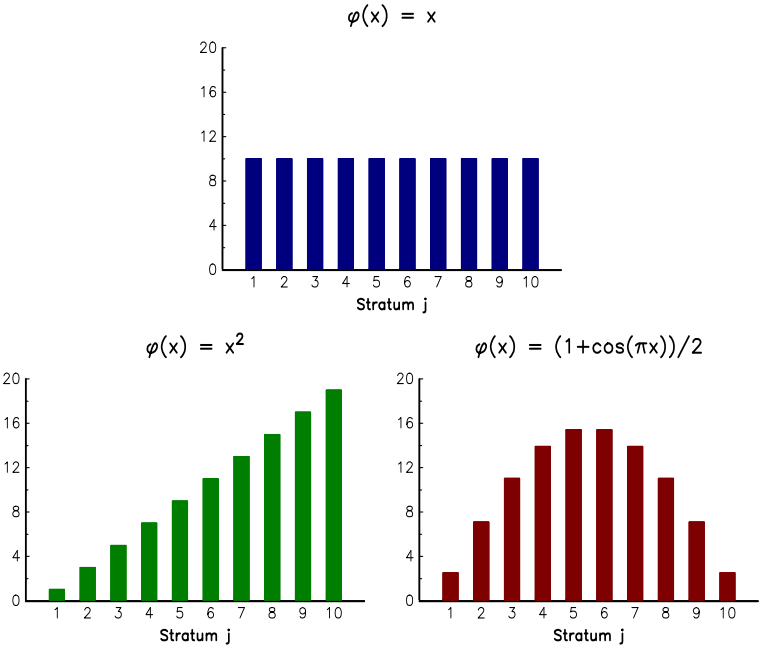


FIGURE 13.45: Optimal allocation $q^\star(j)$ (in %)

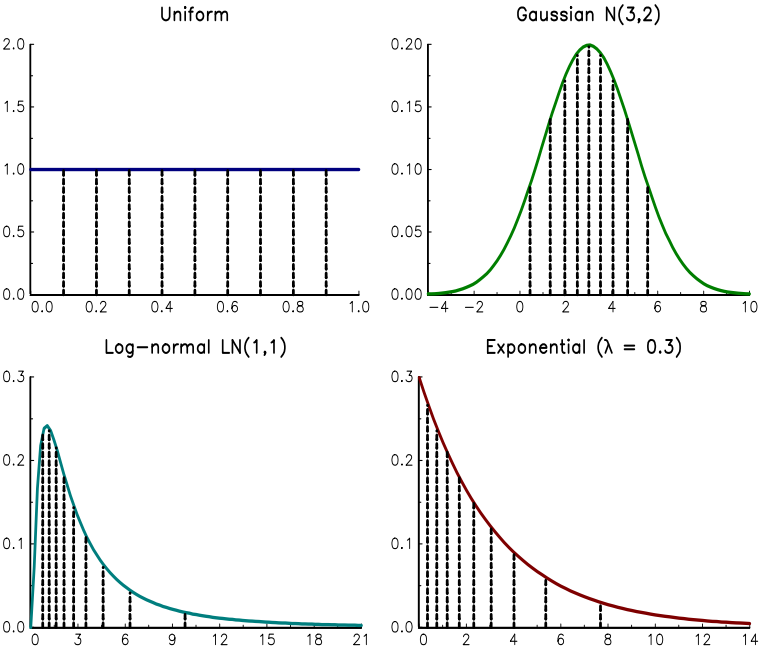


FIGURE 13.46: Strata for different random variables

where $X \sim \mathcal{N}(0,1)$. Let us apply the stratification method with strata defined by Equation (13.13). We have:

$$X(j) = \Phi^{-1}\left(\frac{j-1}{m} + \frac{U}{m}\right)$$

where $U \sim \mathcal{U}_{[0,1]}$. We deduce that:

$$\hat{I}^{(1)}_{\text{STR}}(m) = \frac{e^{-rT}}{m}\sum_{j=1}^{m}\frac{1}{n_S(j)}\sum_{s=1}^{n_s(j)}\max\left(0, S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}X_s(j)} - K\right)$$

However, we notice that $\max(0, S(T) - K)$ is equal to zero if:

$$\begin{aligned} & S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}X} - K \leq 0 \\ \Leftrightarrow \quad & X \leq x_1 = \frac{1}{\sigma\sqrt{T}}\left(\ln\left(\frac{K}{S_0}\right) - rT\right) + \frac{1}{2}\sigma\sqrt{T} \end{aligned}$$

It is then natural to define the first stratum by $]-\infty, x_1]$. Indeed, we have $\varphi(X(1)) = 0$, implying that the intra-strata variance $\sigma(1)$ is equal to zero. For the other strata, we can use the previous approach. For $j \geq 2$, we have:

$$p(j) = \Pr\{X \leq x_j\} - \Pr\{X \leq x_{j-1}\} = \frac{1-p(1)}{m-1}$$

We deduce that:

$$\Pr\{X \leq x_j\} = p(1) + \frac{j-1}{m-1}\cdot(1-p(1))$$

and:

$$x_j = \Phi^{-1}\left(p(1) + \frac{j-1}{m-1}\cdot(1-p(1))\right)$$

The j^{th} stratum is then defined by $X \in [x_{j-1}, x_j]$ with $x_0 = -\infty$, $x_m = +\infty$ and:

$$p(j) = \begin{cases} p(1) & \text{if} \quad j = 1 \\ \dfrac{1-p(1)}{m-1} & \text{if} \quad j \geq 2 \end{cases}$$

To simulate the conditional random variable $X(j)$ for $j \geq 2$, we use the following scheme:

$$X(j) = \Phi^{-1}\left(\Phi(x_{j-1}) + (\Phi(x_j) - \Phi(x_{j-1}))\cdot U\right)$$

where $U \sim \mathcal{U}_{[0,1]}$. We deduce that:

$$\hat{I}^{(2)}_{\text{STR}}(m) = \left(\frac{1-p(1)}{m-1}\right)e^{-rT}\sum_{j=2}^{m}\frac{1}{n_S(j)}\sum_{s=1}^{n_s(j)}\left(S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}X_s(j)} - K\right)$$

In the case of proportional allocation $n_S(j) = n_S \cdot p(j)$, we notice that the total number of simulations is reduced and equal to $(1-p(1))\cdot n_S$ because we don't have to simulate the first stratum.

We consider an European call option with the following parameters: $S_0 = 100$, $K = 105$, $\sigma = 20\%$, $\tau = 1$ and $r = 5\%$. In Figure 13.47, we have reported the variance of the two stratified estimators for different values of m when the number of simulations n_S is equal to 10 000. In the case $m = 1$, we obtain the traditional MC estimator and we have:

$$\text{var}\left(\hat{I}^{(2)}_{\text{STR}}(1)\right) = \text{var}\left(\hat{I}^{(1)}_{\text{STR}}(1)\right) = \text{var}\left(\hat{I}_{\text{MC}}\right)$$

When $m > 1$, we obtain:

$$\text{var}\left(\hat{I}_{\text{STR}}^{(2)}(m)\right) < \text{var}\left(\hat{I}_{\text{STR}}^{(1)}(m)\right) < \text{var}\left(\hat{I}_{\text{MC}}\right)$$

The second stratified estimator is then more efficient than the first stratified estimator, because the design of the first stratum is optimal[49].

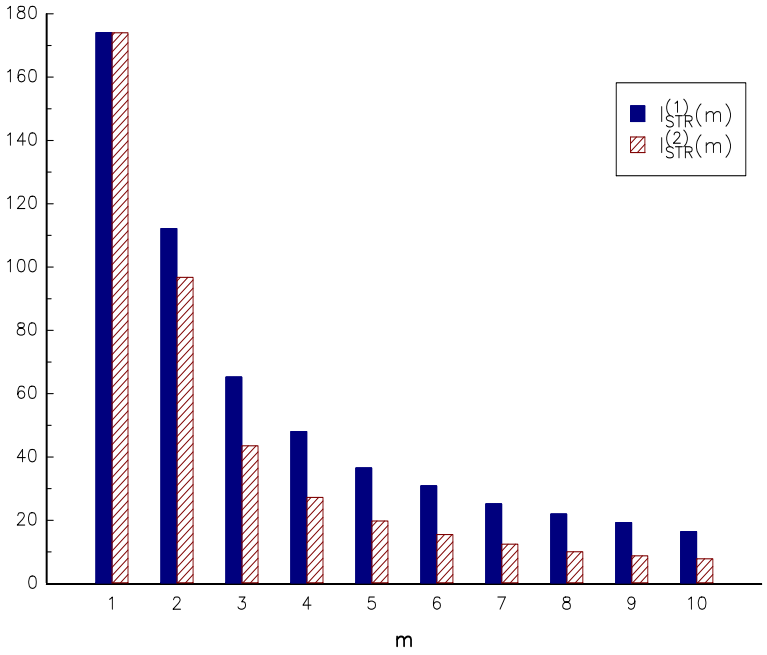


FIGURE 13.47: Variance of the two estimators $\hat{I}_{\text{STR}}^{(1)}(m)$ and $\hat{I}_{\text{STR}}^{(2)}(m)$ for different values of m

13.3.3 MCMC methods

Let us consider a Markov chain with transition density $p\left(x^{(t+1)} \mid x^{(t)}\right) = \Pr\left\{X^{(t+1)} = x^{(t+1)} \mid X^{(t)} = x^{(t)}\right\}$. We also assume that it has a stationary distribution $\pi(x)$. In this case, we can show that the Markov chain satisfies the detailed balance equation[50]:

$$p(y \mid x) \cdot \pi(x) = p(x \mid y) \cdot \pi(y) \tag{13.14}$$

It follows that:

$$\int p(x \mid y)\, \pi(y)\, \mathrm{d}y = \int p(y \mid x)\, \pi(x)\, \mathrm{d}y = \pi(x) \tag{13.15}$$

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for simulating a sample from a probability density function $f(x)$. The underlying idea is to simulate a Markov chain, whose limiting pdf is the desired pdf $f(x)$. It is then equivalent to find the

[49] We have $x_1 = 0.093951$ and $p_1 = 53.74\%$.

[50] In the discrete case, we have:

$$p_{i,j} \cdot \pi_i = p_{j,i} \cdot \pi_j$$

for all the states (i, j) of the Markov chain.

transition kernel $\kappa\left(x^{(t+1)} \mid x^{(t)}\right)$ such that the detailed balance property is satisfied:

$$\kappa(y \mid x) \cdot f(x) = \kappa(x \mid y) \cdot f(y) \tag{13.16}$$

In this case, MCMC methods then generate a sample such that:

$$\lim_{t \to \infty} \Pr\left\{X^{(t)} = x\right\} = f(x)$$

Because the solution of Equation (13.16) is not unique, there are different MCMC methods that will differ by the specification of the transition kernel $\kappa(y \mid x)$.

13.3.3.1 Gibbs sampling

According to Casella and George (1992), the Gibbs sampler has been formulated by Geman and Geman (1984), who studied image-processing models, and popularized in statistics by Gelfand and Smith (1990). Let $f(x_1, \ldots, x_n)$ be the target probability density function and $f(x_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ the conditional density of the i^{th} component of the random vector $X = (X_1, \ldots, X_n)$. At iteration t, the Gibbs sampler (GS) is given by the following steps:

- we draw $x_1^{(t)} \sim f\left(x_1 \mid x_2^{(t-1)}, \ldots, x_n^{(t-1)}\right)$;
- we draw $x_2^{(t)} \sim f\left(x_2 \mid x_1^{(t)}, x_3^{(t-1)}, \ldots, x_n^{(t-1)}\right)$;
- we draw $x_3^{(t)} \sim f\left(x_3 \mid x_1^{(t)}, x_2^{(t)}, x_4^{(t-1)}, \ldots, x_n^{(t-1)}\right)$;
- we draw

$$x_i^{(t)} \sim f\left(x_i \mid \underbrace{x_1^{(t)}, \ldots, x_{i-1}^{(t)}}_{\text{iteration } t}, \underbrace{x_{i+1}^{(t-1)}, \ldots, x_n^{(t-1)}}_{\text{iteration } t-1}\right)$$

  for $i = 4, \ldots, n-1$;
- we draw $x_n^{(t)} \sim f\left(x_n \mid x_1^{(t)}, \ldots, x_{n-1}^{(t)}\right)$;

The algorithm is initialized with $\left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$. After t iterations, we obtain the following Gibbs sequence:

$$\left\{\left(x_1^{(1)}, \ldots, x_n^{(1)}\right), \ldots, \left(x_1^{(t)}, \ldots, x_n^{(t)}\right)\right\} \tag{13.17}$$

Under some conditions and if t is sufficiently large, $\left(x_1^{(t)}, \ldots, x_n^{(t)}\right)$ is a random sample of the joint distribution $f(x_1, \ldots, x_n)$. To obtain n_S simulations of the density $f(x_1, \ldots, x_n)$, Gelfand and Smith (1990) suggested then to generate n_S Gibbs sequences and to take the final value $\left(x_{1,s}^{(t)}, \ldots, x_{n,s}^{(t)}\right)$ from each sequence s. The Monte Carlo estimator of $\mathbb{E}\left[\varphi(X_1, \ldots, X_n)\right]$ is then equal to:

$$\hat{I}_{n_S} = \frac{1}{n_S} \sum_{s=1}^{n_S} \varphi\left(X_{1,s}^{(t)}, \ldots, X_{n,s}^{(t)}\right) \tag{13.18}$$

However, if the Markov chain has reached his stationary state at time n_b, this implies that the Gibbs sequence:

$$\left\{\left(x_1^{(n_b+1)}, \ldots, x_n^{(n_b+1)}\right), \ldots, \left(x_1^{(n_b+n_S)}, \ldots, x_n^{(n_b+n_S)}\right)\right\} \tag{13.19}$$

is also a Gibbs sample of the density $f(x_1, \ldots, x_n)$. We can then formulate another MC estimator[51]:

$$\hat{I}_{n_S} = \frac{1}{n_S} \sum_{t=1}^{n_S} \varphi\left(X_1^{(n_b+t)}, \ldots, X_n^{(n_b+t)}\right) \tag{13.20}$$

However, contrary to the Gelfand-Smith approach, the random vectors of this estimator are correlated. Therefore, the variance of this estimator is larger than in the independent case.

Let us consider the two-dimensional case. We note (X, Y) the random vector and $f(x, y)$ the targeted distribution. At time t, we have $X^{(t)} = x$ and $Y^{(t)} = y$, and we assume that $X^{(t+1)} = x'$ and $Y^{(t+1)} = y'$. If $(x, y) \sim f$, the density $g(x', y')$ of the Gibbs sample is equal to[52]:

$$\begin{aligned} g(x', y') &= \int f(x, y)\, f(x' \mid y)\, f(y' \mid x')\, \mathrm{d}x\, \mathrm{d}y \\ &= \int f(x, y)\, \frac{f(x', y)}{f_y(y)}\, \frac{f(y', x')}{f_x(x')}\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

where f_x and f_y are the marginal densities of the joint distribution $f(x, y)$. It follows that:

$$\begin{aligned} g(x', y') &= \int \frac{f(x, y)}{f_y(y)}\, \frac{f(x', y)}{f_x(x')}\, f(y', x')\, \mathrm{d}x\, \mathrm{d}y \\ &= \int f(x \mid y)\, f(y \mid x')\, f(y', x')\, \mathrm{d}x\, \mathrm{d}y \\ &= f(y', x') \int f(x \mid y)\, f(y \mid x')\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

Because the events $\{X \mid Y = y\}$ and $\{Y \mid X = x'\}$ are independent, we obtain:

$$\begin{aligned} g(x', y') &= f(y', x') \int f(x \mid y)\, \mathrm{d}x \int f(y \mid x')\, \mathrm{d}y \\ &= f(y', x') \end{aligned}$$

We deduce that $(x', y') \sim f$. If the Gibbs sampler reaches a stationary regime, it then converges to the targeted distribution $f(x, y)$.

Remark 169 *In the two dimensional case, we notice that the proposal kernel is equal to:*

$$\kappa(x', y' \mid x, y) \propto f(x' \mid y) \cdot f(y' \mid x')$$

In the general case, we have:

$$\kappa(y \mid x) \propto \prod_{i=1}^{n} f(y_i \mid y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)$$

Example 158 *Casella and George (1992) consider the following joint distribution of X and Y:*

$$f(x, y) \propto \binom{n}{x} y^{x+\alpha-1} (1-y)^{n-x+\beta-1}$$

[51] This approach requires a burn-in period, meaning that an initial number of samples is discarded. However, it is not always obvious to know when the Markov chain has converged.

[52] We have the following sequences $(x, y) \to (x', y)$ where $x' \sim f(X \mid y)$ and $(x', y) \to (x', y')$ where $y' \sim f(Y \mid x')$.

where $x \in \{0, 1, \ldots, n\}$ *and* $y \in [0, 1]$. *It follows that:*

$$\begin{aligned} f(x \mid y) &\propto y^{\alpha-1}(1-y)^{\beta-1} \cdot \binom{n}{x} y^x (1-y)^{n-x} \\ &\propto \binom{n}{x} y^x (1-y)^{n-x} \\ &\sim \mathcal{B}(n, y) \end{aligned}$$

and:

$$\begin{aligned} f(y \mid x) &\propto \binom{n}{x} \cdot y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \\ &\propto y^{x+\alpha-1}(1-y)^{n-x+\beta-1} \\ &\sim \mathcal{B}(x+\alpha, n-x+\beta) \end{aligned}$$

Therefore, $\{X \mid Y = y\}$ *is a Bernoulli random variable* $\mathcal{B}(n, p)$ *with* $p = y$ *and* $\{Y \mid X = x\}$ *is a beta random variable* $\mathcal{B}(\alpha', \beta')$ *with* $\alpha' = x + \alpha$ *and* $\beta' = n - x + \beta$. *In Figure 13.48, we have reported the Gibbs sequence of* 1 000 *iterations* $\left(x^{(t)}, y^{(t)}\right)$ *for the parameters* $n = 5$, $\alpha = 2$ *and* $\beta = 4$. *The initial values are* $x^{(0)} = 5$ *and* $y^{(0)} = 1/2$. *We assume that the burn-in-period corresponds to the initial 200 iterations. We can then calculate* $I = \mathbb{E}[X \cdot Y]$ *by Monte Carlo with the next* 800 *iterations. We obtain* $\hat{I} = 0.71$. *We can also show that the variance of this MCMC estimator is three times larger than the variance of the traditional MC estimator. This is due to the high autocorrelation between the samples*[53].

13.3.3.2 Metropolis-Hastings algorithm

Like the Gibbs sampler, the Metropolis-Hastings algorithm considers a multidimensional probability density function $f(x) = f(x_1, \ldots, x_n)$. Let $q(y \mid x) = q(y_1, \ldots, y_n \mid x_1, \ldots, x_n)$ be the Markov transition density or the proposal density. The Metropolis-Hastings (MH) algorithm consists in the following steps:

1. given the state $x^{(t)}$, we generate $y \sim q\left(Y \mid x^{(t)}\right)$ from the Markov transition density;
2. we generate a uniform random number $u \sim \mathcal{U}_{[0,1]}$;
3. we calculate the density ratio $r\left(x^{(t)}, y\right)$ defined by:

$$r(x, y) = \frac{q(x \mid y) \cdot f(y)}{q(y \mid x) \cdot f(x)}$$

   and $\alpha\left(x^{(t)}, y\right) = \min\left(r\left(x^{(t)}, y\right), 1\right)$;

4. we set:

$$x^{(t+1)} = \begin{cases} y & \text{if } u \leq \alpha\left(x^{(t)}, y\right) \\ x^{(t)} & \text{otherwise} \end{cases}$$

The Metropolis-Hastings algorithm can be viewed as an acceptance-rejection algorithm, when the samples are correlated due to the Markov chain (Hastings, 1970).

[53] We have:

$$\rho\left\langle X^{(t)} \cdot Y^{(t)}, X^{(t-1)} \cdot Y^{(t-1)} \right\rangle = 52\%$$

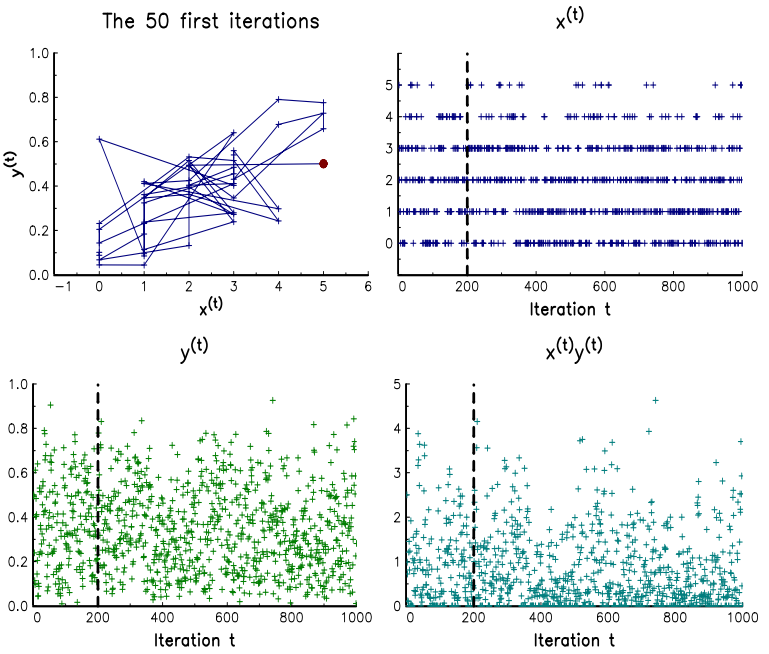


FIGURE 13.48: Illustration of the Gibbs sampler

The underlying idea of the MH algorithm is to build a kernel density $\kappa(y \mid x)$ such that the Markov chain converges to the targeted distribution $f(x)$. In this case, we must verify that:

$$\kappa(y \mid x) \cdot f(x) = \kappa(x \mid y) \cdot f(y)$$

It would be a pure coincidence that the kernel density $\kappa(y \mid x)$ is equal to the proposal density $q(y \mid x)$. Suppose that:

$$q(y \mid x) \cdot f(x) > q(x \mid y) \cdot f(y)$$

Chib and Greenberg (1995) explain that "*the process moves from x to y too often and from y to x too rarely*". To reduce the number of moves from x to y, we can introduce the probability $\alpha(x, y) < 1$ such that $\alpha(y, x) = 1$ and:

$$\begin{aligned} q(y \mid x) \cdot \alpha(x, y) \cdot f(x) &= q(x \mid y) \cdot \alpha(y, x) \cdot f(y) \\ &= q(x \mid y) \cdot f(y) \end{aligned}$$

where $\alpha(y, x) = 1$. We deduce that:

$$\alpha(x, y) = r(x, y) = \frac{q(x \mid y) \cdot f(y)}{q(y \mid x) \cdot f(x)}$$

If $q(y \mid x) \cdot f(x) < q(x \mid y) \cdot f(y)$, we have $q(y \mid x) \cdot f(x) = q(x \mid y) \cdot \alpha(y, x) \cdot f(y)$. Because the Markov chain must be reversible, we finally obtain that:

$$\alpha(x, y) = \min\left(\frac{q(x \mid y) \cdot f(y)}{q(y \mid x) \cdot f(x)}, 1\right)$$

From the previous analysis, we deduce that the kernel density is $\kappa(y \mid x) = q(y \mid x)\,\alpha(x, y)$. However, this result does not take into account that the Markov chain can remain at x.

Therefore, Chib and Greenberg (1995) show that the kernel density of the MH algorithm has the following form:

$$\begin{aligned}\kappa(y \mid x) &= q(y \mid x)\,\alpha(x,y) + \\ &\quad \left(1 - \int q(y \mid x)\,\alpha(x,y)\,\mathrm{d}y\right)\delta_x(y)\end{aligned}$$

where $\delta_x(y)$ is the Dirac delta function.

Remark 170 *The previous analysis shows that $\alpha(x,y)$ is the probability to move from x to y. $\alpha(x,y)$ is then the acceptance ratio of the MH algorithm. Contrary to the rejection sampling algorithm, the efficiency of the MH algorithm is not necessarily obtained when $\alpha(x,y)$ is equal to 1. Indeed, there are two sources of correlation between $x^{(t)}$ and $x^{(t+1)}$: (1) the correlation $\rho\left(x^{(t)},y\right)$ between $x^{(t)}$ and y, and (2) the correlation $\rho\left(x^{(t)},x^{(t)}\right)$ because y is rejected. Therefore, we face a trade-off between reducing the correlation $\rho\left(x^{(t)},y\right)$ and increasing the acceptance ratio $\alpha\left(x^{(t)},y\right)$. Suppose for instance that we use a proposal distribution with small variance, the correlation $\rho\left(x^{(t)},y\right)$ is high, but the acceptance ratio $\alpha\left(x^{(t)},y\right)$ is high. On the contrary, if we use a proposal distribution with small variance, the correlation $\rho\left(x^{(t)},y\right)$ is low, but the acceptance ratio $\alpha\left(x^{(t)},y\right)$ is also low. Therefore, it is extremely difficult to find a proposal distribution such that the correlation $\rho\left(x^{(t)},y\right)$ is low and the acceptance ratio $\alpha\left(x^{(t)},y\right)$ is high.*

In the original Metropolis algorithm (Metropolis *et al.*, 1953), the authors assumed that the proposal distribution is symmetric: $q(y \mid x) = q(x \mid y)$. In this case, the acceptance ratio is equal to:

$$\alpha(x,y) = \min\left(\frac{f(y)}{f(x)}, 1\right)$$

An example of such algorithm is the random walk sampler:

$$Y = x^{(t)} + Z$$

where the random vector Z follows a symmetric distribution. Another special case of the Metropolis-Hastings algorithm is the independence sampler: $q(y \mid x) = q(y)$. The proposal distribution does not depend on x and the acceptance ratio becomes:

$$\alpha(x,y) = \min\left(\frac{q(x)\cdot f(y)}{q(y)\cdot f(x)}, 1\right)$$

The MH algorithm is then very similar to the rejection sampling method, except that it produces correlated samples. We also notice that the Gibbs sampler is a special case of the MH algorithm where[54]:

$$q(y \mid x) \sim f(x_i \mid x_{-i})$$

Example 159 *We consider the simulation of the bivariate Gaussian random vector $X = (X_1, X_2) \sim \mathcal{N}(\mu, \Sigma)$ with the Metropolis-Hastings algorithm and a symmetric proposal distribution. It follows that:*

$$\alpha(x,y) = \min\left(\frac{\exp\left(-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)\right)}{\exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)}, 1\right)$$

[54] At each iteration, we have $y_i \sim f(x_i \mid x_{-i})$, $y_j = x_j$ if $j \neq i$, and $\alpha(x,y) = 1$.

The parameters are $\mu_1 = 1$, $\mu_2 = -1$, $\sigma_1 = 2$, $\sigma_2 = 1$ *and* $\rho = 99\%$. *We use the random walk sampler* $Y_i = x_i^{(t)} + Z_i$ *for* $i = 1, 2$. *The random vector* (Z_1, Z_2) *is generated using the following four proposal distributions:*

(a) $Z_1 \sim \mathcal{N}(0, 1)$ *and* $Z_2 \sim \mathcal{N}(0, 1)$;

(b) $Z_1 \sim \mathcal{N}(0, 0.1)$ *and* $Z_2 \sim \mathcal{N}(0, 0.1)$;

(c) $Z_1 \sim \mathcal{U}_{[-2,2]}$ *and* $Z_2 \sim \mathcal{U}_{[-2,2]}$;

(d) $Z_1 \sim \mathcal{U}_{[-0.2,0.2]}$ *and* $Z_2 \sim \mathcal{U}_{[-0.2,0.2]}$.

In Figure 13.49, we have reported the simulated samples of the first 2 000 *iterations for the four cases when we use a burn-in-period of* 500 *iterations. The sampler is initialized with* $x_1^{(0)} = x_2^{(0)} = 0$. *The acceptance ratio is respectively equal to* 15% *(a),* 43% *(b),* 10% *(c) and* 72% *(d). The acceptance ratio is the highest for the case (d). However, we notice that this proposal distribution is slow to explore the entire space. To obtain a sample such that* $x_1^{(t)} > 3$ *and* $x_2^{(t)} > 0$, *we need more iterations* ($n_S > 5000$). *On the contrary, proposal distributions (a) and (c) have a high variance. The exploration of the probability space is then faster, but the acceptance ratio is also lower. This example illustrates the trade-off between the autocorrelation and the acceptance ratio.*

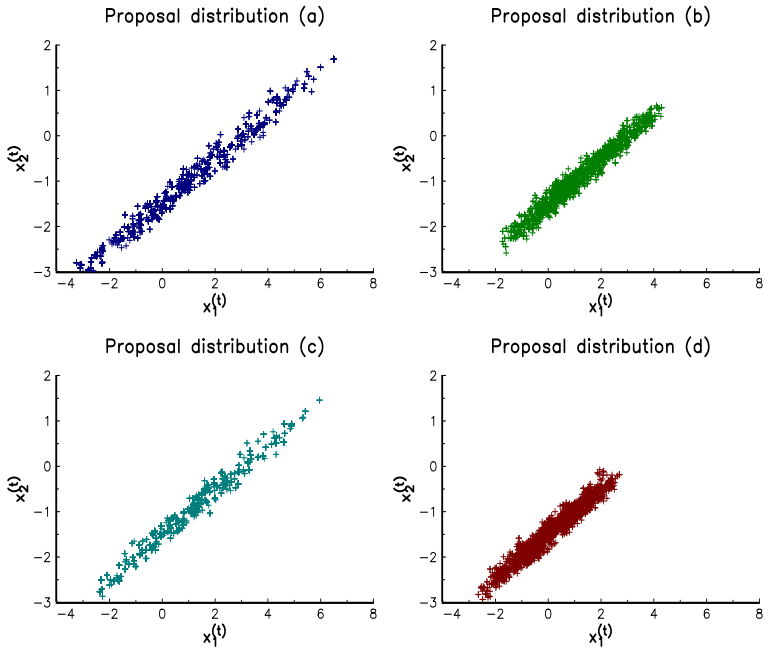


FIGURE 13.49: Illustration of the random walk sampler

The previous example is purely illustrative, because we don't need to use MCMC for simulating Gaussian random vectors. Let us now consider the following bivariate probability density functions:

(a) the pdf is a perturbation of the Gaussian density function:

$$f(x_1, x_2) \propto \exp\left(-x_1^2 - x_2^2 - x_1\right)$$

(b) the pdf is a mixture of two Gaussian density functions:

$$f(x_1, x_2) \propto \exp\left(-x_1^2 - x_2^2\right) + \exp\left(-x_1^2 - x_2^2 - 1.8 \cdot x_1 \cdot x_2\right)$$

(c) the pdf is a complex function of $\Phi(x_1)$ and $\Phi(x_2)$:

$$f(x_1, x_2) \propto \exp\left(-0.1 \cdot \Phi(x_1) \cdot \Phi(x_2) \cdot x_2^2\right)$$

(d) we consider the pdf of the Clayton copula with two exponential marginal distributions[55]:

$$f(x_1, x_2) = (1+\theta)\lambda_1\lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2}\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta - 2}(u_1 u_2)^{-1-\theta}$$

where $u_1 = 1 - e^{-\lambda_1 x_1}$ and $u_2 = 1 - e^{-\lambda_2 x_2}$.

We notice that we don't know the normalization constant for the first three pdfs. The third pdf is very complex and needs a very accurate algorithm for computing the Gaussian cdf. The fourth case is a copula model. We use the bivariate Gaussian probability distribution with a correlation equal to 50% for the proposal distribution. A sample of 1500 iterations simulated with the random walk sampler is given in Figure 13.50. The fourth panels have been obtained with the same random numbers of U, Z_1 and Z_2.

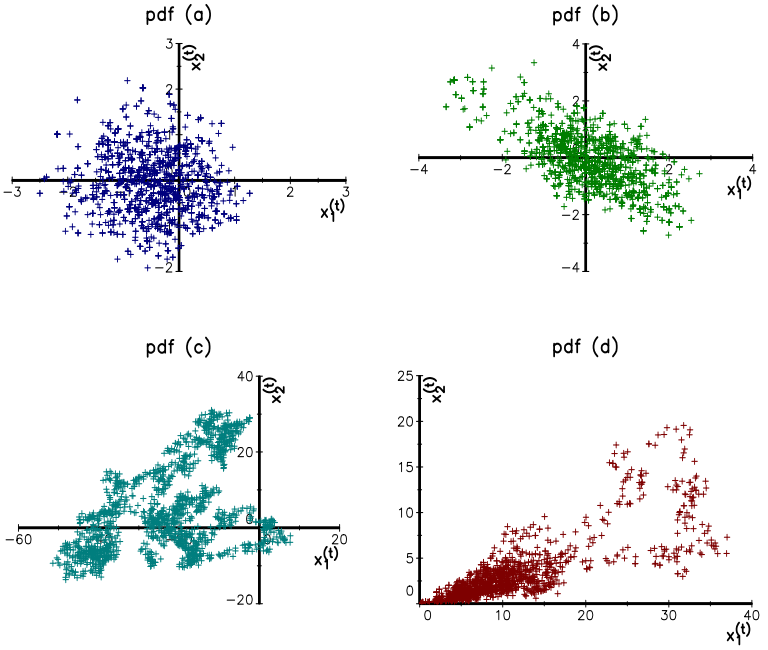


FIGURE 13.50: Simulating bivariate probability distributions with the MH algorithm

Remark 171 *For high-dimensional MCMC problems, the proposal distribution is generally the Gaussian distribution* $\mathcal{N}(\mu, \Sigma)$ *or the multivariate Student's t distribution* $t_n(\Sigma, \nu)$.

[55] We use the parameters $\theta = 2.5$, $\lambda_1 = 1\%$ and $\lambda_2 = 5\%$.

13.3.3.3 Sequential Monte Carlo methods and particle filters

Sequential Monte Carlo methods (or SMC) and particle filters (PF) are used when we consider non-linear/non-Gaussian state space models (or hidden Markov models). In these models, the state vector $X^{(t)}$ is characterized by the transition density:

$$X^{(t+1)} = x' \mid X^{(t)} = x \sim f\left(x' \mid x\right)$$

We assume that the state vector $X^{(t)}$ is not directly observed. The process that is observed is denoted by $Y^{(t)}$ and is characterized by the measurement density:

$$Y^{(t)} = y \mid X^{(t)} = x \sim f\left(y \mid x\right)$$

Let $x^{(0:T)} = \left(x^{(0)}, \ldots, x^{(T)}\right)$ be the exhaustive statistic of the system. By the Markov property, we have:

$$f\left(x^{(0:T)}\right) = f\left(x^{(0)}\right) \prod_{t=1}^{T} f\left(x^{(t)} \mid \mathcal{I}_t\right)$$

where $\mathcal{I}_t$ represents all the information available at time t (including $y^{(1)}, \ldots, y^{(t)}$). To calculate $f\left(x^{(t)} \mid \mathcal{I}_t\right)$, we have:

$$f\left(x^{(t)} \mid \mathcal{I}_t\right) \propto f\left(y^{(t)} \mid x^{(t)}\right) f\left(x^{(t)} \mid \mathcal{I}_{t-1}\right) \tag{13.21}$$

where $f\left(x^{(t)} \mid \mathcal{I}_{t-1}\right)$ is the prior density of $X^{(t)}$ and $f\left(y^{(t)} \mid x^{(t)}\right)$ is the log-likelihood function of the observed variable $Y^{(t)}$. This equation is known as the Bayes update step. We recall that the prior density of the state variable $X^{(t)}$ is given by the Chapman-Kolmogorov equation:

$$f\left(x^{(t)} \mid \mathcal{I}_{t-1}\right) = \int f\left(x^{(t)} \mid x^{(t-1)}\right) f\left(x^{(t-1)} \mid \mathcal{I}_{t-1}\right) \mathrm{d}x^{(t-1)} \tag{13.22}$$

This equation is also known as the Bayes prediction step. It gives an estimate of the probability density function of $X^{(t)}$ given all the information until $t-1$. A Bayesian filter corresponds to the system of the two recursive equations (13.21) and (13.22). In order to initialize the recurrence algorithm, we assume that the density function of the initial state vector $f\left(x^{(0)}\right)$ is known.

The underlying idea of SMC methods is then to estimate the density functions $f\left(x^{(t)} \mid \mathcal{I}_{t-1}\right)$ and $f\left(x^{(t)} \mid \mathcal{I}_t\right)$. Given these estimates, we may also compute the best estimates $\hat{x}^{(t)} \mid \mathcal{I}_{t-1}$ and $\hat{x}^{(t)}$, which are given by:

$$\hat{x}^{(t)} \mid \mathcal{I}_{t-1} = \mathbb{E}\left[X^{(t)} \mid \mathcal{I}_{t-1}\right] = \int x^{(t)} f\left(x^{(t)} \mid \mathcal{I}_{t-1}\right) \mathrm{d}x^{(t)}$$

and:

$$\hat{x}^{(t)} = \mathbb{E}\left[X^{(t)} \mid \mathcal{I}_t\right] = \int x^{(t)} f\left(x^{(t)} \mid \mathcal{I}_t\right) \mathrm{d}x^{(t)}$$

For that, we estimate the density $f\left(x^{(t)} \mid \mathcal{I}_t\right)$ by the Monte Carlo method. At time $t-1$, we assume that $f\left(x^{(t-1)} \mid \mathcal{I}_{t-1}\right)$ is approximated by a sample $\left\{x_1^{(t-1)}, \ldots, x_{n_S}^{(t-1)}\right\}$ and a vector of associated weights $\left\{w_1^{(t-1)}, \ldots, w_{n_S}^{(t-1)}\right\}$, where n_S is the number of simulated particles. We deduce that:

$$f\left(x^{(t)} \mid \mathcal{I}_t\right) \propto f\left(y^{(t)} \mid x^{(t)}\right) \sum_{s=1}^{n_S} w_s^{(t-1)} f\left(x^{(t)} \mid x_s^{(t-1)}\right)$$

The problem consists then in estimating the states $\left\{x_1^{(t)}, \ldots, x_{n_S}^{(t)}\right\}$ and the corresponding weights $\left\{w_1^{(t)}, \ldots, w_{n_S}^{(t)}\right\}$.

Computation of weights Following Arulampalam *et al.* (2002), we apply the method of importance sampling to the joint distribution of $x^{(0:t)} = \left(x^{(0)}, \ldots, x^{(t)}\right)$:

$$\begin{aligned} f\left(x^{(0:t)} \mid y^{(t)}, \mathcal{I}_{t-1}\right) &= f\left(x^{(t)} \mid y^{(t)}, x^{(0:t-1)}, \mathcal{I}_{t-1}\right) f\left(x^{(0:t-1)} \mid \mathcal{I}_{t-1}\right) \\ &= \frac{f\left(y^{(t)} \mid x^{(t)}\right) f\left(x^{(t)} \mid x^{(0:t-1)}\right)}{f\left(y^{(t)} \mid \mathcal{I}_{t-1}\right)} f\left(x^{(0:t-1)} \mid \mathcal{I}_{t-1}\right) \\ &\propto f\left(y^{(t)} \mid x^{(t)}\right) f\left(x^{(t)} \mid x^{(0:t-1)}\right) f\left(x^{(0:t-1)} \mid \mathcal{I}_{t-1}\right) \end{aligned}$$

Let q be the instrumental density such that it factorizes in the following way:

$$q\left(x_s^{(0:t)} \mid y^{(t)}, \mathcal{I}_{t-1}\right) = q\left(x_s^{(t)}, x_s^{(0:t-1)} \mid y^{(t)}, \mathcal{I}_{t-1}\right) q\left(x_s^{(0:t-1)}, \mathcal{I}_{t-1}\right)$$

We deduce that[56]:

$$\begin{aligned} w_s^{(t)} &\propto \frac{f\left(x_s^{(0:t)} \mid y^{(t)}, \mathcal{I}_{t-1}\right)}{q\left(x_s^{(0:t)} \mid y^{(t)}, \mathcal{I}_{t-1}\right)} \\ &\propto \frac{f\left(y^{(t)} \mid x_s^{(t)}\right) f\left(x_s^{(t)} \mid x_s^{(0:t-1)}\right) f\left(x_s^{(0:t-1)} \mid \mathcal{I}_{t-1}\right)}{q\left(x_s^{(t)}, x_s^{(0:t-1)} \mid y^{(t)}, \mathcal{I}_{t-1}\right)} \\ &= \frac{f\left(y^{(t)} \mid x_s^{(t)}\right) f\left(x_s^{(t)} \mid x_s^{(0:t-1)}\right)}{q\left(x_s^{(t)} \mid y^{(t)}, x_s^{(0:t-1)}, \mathcal{I}_{t-1}\right)} \cdot \frac{f\left(x_s^{(0:t-1)} \mid \mathcal{I}_{t-1}\right)}{q\left(x_s^{(0:t-1)} \mid y^{(t)}, \mathcal{I}_{t-1}\right)} \\ &= \frac{f\left(y^{(t)} \mid x_s^{(t)}\right) f\left(x_s^{(t)} \mid x_s^{(0:t-1)}\right)}{q\left(x_s^{(t)} \mid y^{(t)}, x_s^{(0:t-1)}, \mathcal{I}_{t-1}\right)} \cdot w_s^{(t-1)} \end{aligned} \tag{13.23}$$

The posterior density at time t can then be approximated as:

$$f\left(x^{(t)} \mid y^{(t)}\right) = \sum_{s=1}^{n_S} w_s^{(t)} \cdot \delta_x\left(x^{(t)} - x_s^{(t)}\right) \tag{13.24}$$

The previous computations lead to the sequential importance sampling (or SIS) algorithm:

1. at time t, we simulate $x_s^{(t)} \sim q\left(x_s^{(t)} \mid y^{(t)}, x_s^{(0:t-1)}, \mathcal{I}_{t-1}\right)$;
2. we update the weight $w_s^{(t)}$ using Equation (13.23);
3. we repeat steps 1 and 2 in order to obtain a sample of n_S particles;
4. we normalize the weights:

$$w_s^{(t)} \leftarrow \frac{w_s^{(t)}}{\sum_{s=1}^{n_S} w_s^{(t)}}$$

Remark 172 *With the SIS algorithm, the variance of the weights increases exponentially with the number of particles. This implies that some particles have negligible weights whereas others have large weights. This is why resampling techniques are generally added in order to reduce this phenomenon. Another method consists in simulating new (or auxiliary) particles at each time. Therefore, there exist several SMC algorithms (Arulampalam* et al.*, 2002; Doucet and Johansen, 2009): auxiliary particle filter (APF), generic particle filter (GPF), regularized particle filter (RPF), sampling importance resampling (SIR), etc.*

[56] Note that $y^{(t)} \in \mathcal{I}_t$.

An example We consider the following example:

$$\begin{cases} f\left(x^{(t)} \mid x^{(t-1)}\right) = \mathcal{N}\left(x^{(t)}; T\left(t, x^{(t-1)}\right), Q\right) \\ f\left(y^{(t)} \mid x^{(t)}\right) = \mathcal{N}\left(y^{(t)}; \kappa\left(x^{(t)}\right)^2, H\right) \end{cases}$$

The corresponding state space model is equal to:

$$\begin{cases} x^{(t)} = T\left(t, x^{(t-1)}\right) + \eta^{(t)} \\ y^{(t)} = \kappa\left(x^{(t)}\right)^2 + \varepsilon^{(t)} \end{cases} \tag{13.25}$$

where $\eta^{(t)} \sim \mathcal{N}(0, Q)$ and $\varepsilon^{(t)} \sim \mathcal{N}(0, H)$. We notice that the state space model is non-linear. The previous example has been extensively studied with the following specification (Carlin *et al.*, 1992; Kitagawa, 1996):

$$T(t, x) = \frac{x}{2} + \frac{25 \cdot x}{1 + x^2} + 8 \cdot \cos(1.2 \cdot t)$$

With the following values of parameters $\kappa = 1/20$, $Q = 1$ and $R = 10$, we simulate the model (13.25) and estimate $x^{(t)}$ by considering different particle filters. The likelihood and prior transition densities are given by $f\left(y^{(t)} \mid x^{(t)}\right)$ and $f\left(x^{(t)} \mid x^{(t-1)}\right)$. We assume that the instrumental density $q\left(x^{(t)} \mid y^{(t)}, x^{(t-1)}, \mathcal{I}_{t-1}\right)$ is equal to the transition density $f\left(x^{(t)} \mid x^{(t-1)}\right)$, meaning that the knowledge of $y^{(t)}$ does not improve the estimate of the state $x^{(t)}$. The particles are simulated according to the following scheme:

$$x_s^{(t)} = T\left(t, x_s^{(t-1)}\right) + \eta_s^{(t)}$$

where $\eta_s^{(t)} \sim \mathcal{N}(0, Q)$. In Figure 13.51, we report one simulation of the state space model. Recall that we observe $y^{(t)}$ and not $x^{(t)}$. The estimate $\hat{x}^{(t)}$ is equal to:

$$\hat{x}^{(t)} = \sum_{s=1}^{n_S} w_s^{(t)} \cdot x_s^{(t)}$$

where $x_s^{(t)}$ is the simulated value of $x^{(t)}$ for the s^{th} particle and $w_s^{(t)}$ is the importance weight given by Equation (13.23). For each simulation, we also calculate the root mean squared error:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(x^{(t)} - \hat{x}^{(t)}\right)^2}$$

where $x^{(t)}$ is the true value of the state and $\hat{x}^{(t)}$ is the estimate. We report the probability density function of the RMSE statistic in Figure 13.52 when the number of particles is equal to 1 000. We notice that the SIR algorithm is better than the other SMC algorithms for this example. Figure 13.53 illustrates the convergence of the SIS algorithm with respect to the number of particles.

13.3.4 Quasi-Monte Carlo simulation methods

We consider the following Monte Carlo problem:

$$I = \int \cdots \int_{[0,1]^n} \varphi(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

Let X be the random vector of independent uniform random variables. It follows that $I = \mathbb{E}[\varphi(X)]$. The Monte Carlo method consists in generating uniform coordinates in

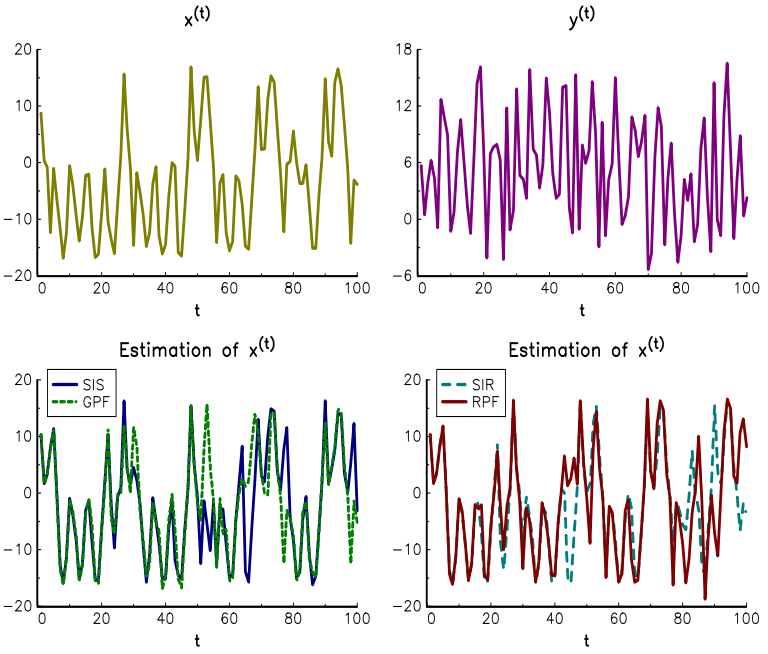


FIGURE 13.51: An example of a SMC run with 1 000 particles

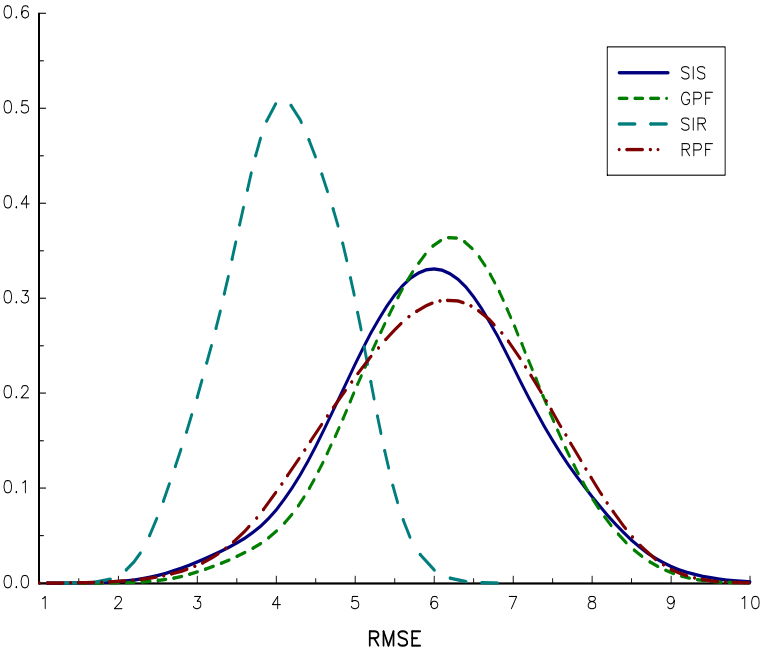


FIGURE 13.52: Density of the RMSE statistic for 1 000 particles

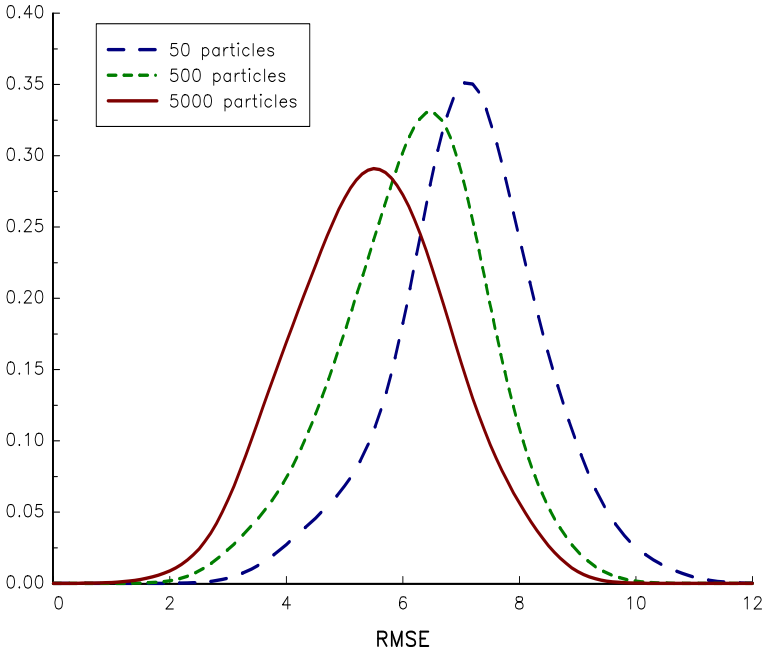


FIGURE 13.53: Density of the RMSE statistic for the SIS algorithm

the hypercube $[0,1]^n$. Quasi-Monte Carlo methods use non-random coordinates in order to obtain a more nicely uniform distribution. A low discrepancy sequence $\mathcal{U} = \{u_1, \ldots, u_{n_S}\}$ is then a set of deterministic points distributed in the hypercube $[0,1]^n$. Let us define the star discrepancy of $\mathcal{U}$ by $D^\star$:

$$D^\star_{n_s}(\mathcal{U}) = \sup_{x \in [0,1]^n} \left| \frac{1}{n_S} \sum_{s=1}^{n_S} \prod_{i=1}^{n} \mathbb{1}\{u_{i,s} \leq x_i\} - \prod_{i=1}^{n} x_i \right|$$

We could interpret $D^\star$ as the $\mathcal{L}_\infty$ norm between the theoretical continuous uniform distribution and the discrete uniform distribution generated by the low discrepancy sequence $\mathcal{U}$. We note that if $\mathcal{U}$ is really uniform, then $\lim_{n_S \to \infty} D^\star_{n_S}(\mathcal{U}) = 0$ for every dimension n. Moreover, Morokoff and Caflisch (1994) noticed that:

$$|I_{n_S} - I| \leq D^\star_{n_S}(\mathcal{U}) \cdot \mathcal{V}(f)$$

where $\mathcal{V}(f)$ is the Hardy-Krause variation of f. We could find low discrepancy sequences such that the error is of order $n_S^{-1} (\ln n_S)^n$ in probability (Morokoff and Caflisch, 1994). If we compare this bound with the order convergence $n_S^{-1/2}$ of MC, we notice that QMC is theoretically better than MC for small dimensions, but MC is better than QMC for high dimensions. However, in practice, it appears that QMC could be more accurate than MC even for large dimension n.

Glasserman (2003) reviewed different quasi-random sequences. The most known are the Halton, Sobol and Faure sequences and corresponding numerical codes are available in different programming languages (Press *et al.*, 2007). The techniques to generate these sequences are based on number theory. For example, the Halton sequence is based on the p-adic expansion of integers $n = d_k p^k + \cdots + d_1 p + d_0$ and the radical-inverse function $\varrho_b(n) = \sum_{i=0}^{k} d_k p^{-(i+1)}$ where $d_i \in \{0, \ldots, p-1\}$ for $i = 0, \ldots, k$. The d-dimensional

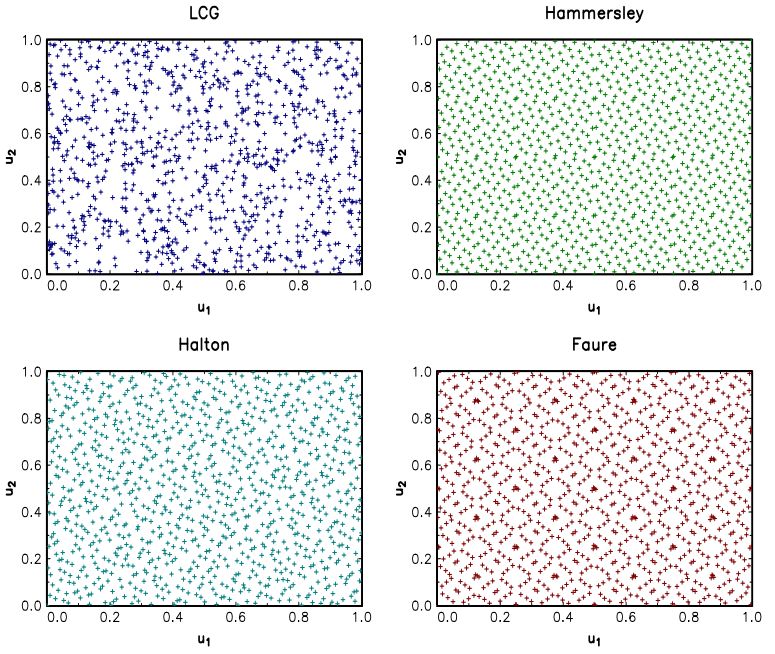


FIGURE 13.54: Comparison of different low discrepancy sequences

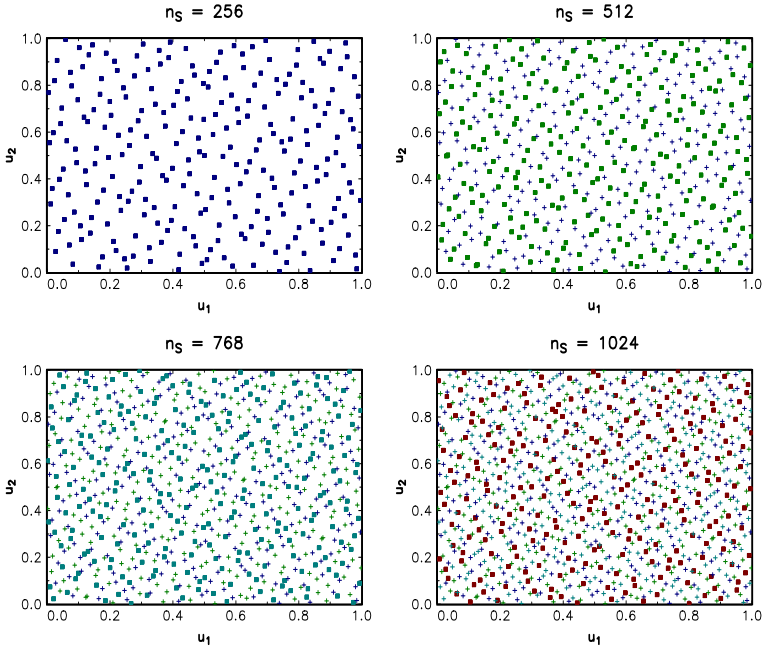


FIGURE 13.55: The Sobol generator

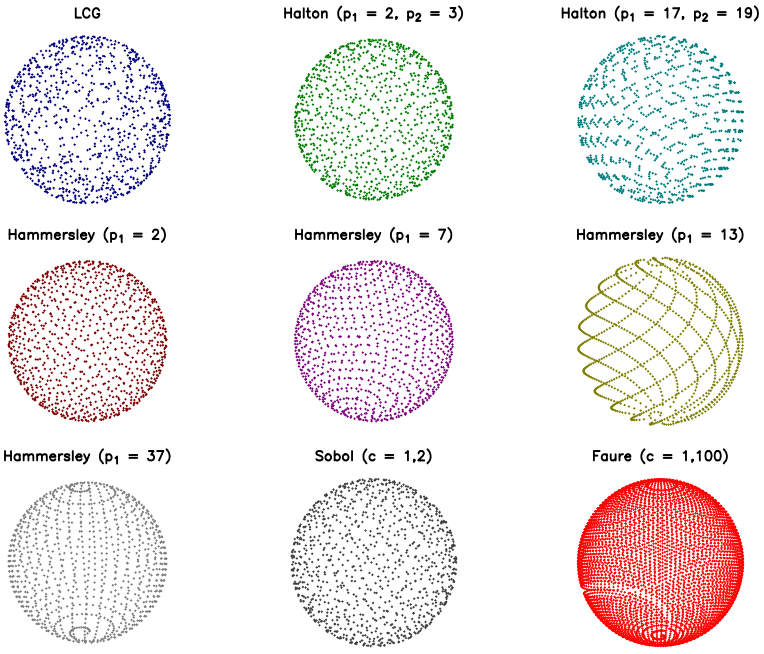


FIGURE 13.56: Quasi-random points on the unit sphere

TABLE 13.8: Pricing of the spread option using quasi-Monte Carlo methods

n_S	10^2	10^3	10^4	10^5	10^6	5×10^6
LCG (1)	4.3988	5.9173	5.8050	5.8326	5.8215	5.8139
LCG (2)	6.1504	6.1640	5.8370	5.8219	5.8265	5.8198
LCG (3)	6.1469	5.7811	5.8125	5.8015	5.8142	5.8197
Hammersley (1)	32.7510	26.5326	21.5500	16.1155	9.0914	5.8199
Hammersley (2)	32.9082	26.4629	21.5465	16.1149	9.0914	5.8199
Halton (1)	8.6256	6.1205	5.8493	5.8228	5.8209	5.8208
Halton (2)	10.6415	6.0526	5.8544	5.8246	5.8208	5.8207
Halton (3)	8.5292	6.0575	5.8474	5.8235	5.8212	5.8208
Sobol	5.7181	5.7598	5.8163	5.8190	5.8198	5.8198
Faure	5.7256	5.7718	5.8157	5.8192	5.8197	5.8198

Halton sequence $\varrho = \{\varrho_n\}$ is then defined by $\varrho_n = (\varrho_{p_1}(n), \ldots, \varrho_{p_d}(n))$ where $p_1, \ldots, p_d$ are integers that are greater than one and pairwise relatively prime. We represent this sequence when $d = 2$ and $n_S = n = 1024$, and compare it to LCG random variates and Hammersley and Faure sequences in Figure 13.54. The underlying idea of QMC is to add the new points not randomly, but between the existing points. For example, we have added 256 points in the Sobol sequence[57] in each panel of Figure 13.55. Finally, we report the projection of different low discrepancy sequences on the unit sphere in Figure 13.56. We notice that we can generate some '*hole area*'.

Example 160 *We consider a spread option whose payoff is equal to* $(S_1(T) - S_2(T) - K)^+$. *The price is calculated using the Black-Scholes model, and the following parameters:*

[57] The new points correspond to a square symbol.

$S_1(0) = S_2(0) = 100$, $\sigma_1 = \sigma_2 = 20\%$, $\rho = 50\%$ and $r = 5\%$. The maturity T of the option is set to one year, whereas the strike K is equal to 5. We estimate the option price with several QMC methods and different number of simulations n_S. Results are given in Table 13.8. We consider three seed values for the LCG pseudorandom sequences. In the case of Hammersley sequences, we use $p_1 = 2$ (sequence 1) and $p_1 = 7$ (sequence 2). We also use three Halton sequences based on the following values of (p_1, p_2): $(2,3)$, $(17,19)$ and $(2,19)$. Finally, we consider Sobol and Faure sequences when the dimension is equal to 2. For simulating Gaussian random variates, we use the Box-Muller method. The true price of the spread option is equal to 5.8198. We notice that only the Sobol and Faure generators have converged to this price when the number of simulations is equal to one million.

13.4 Exercises

13.4.1 Simulating random numbers using the inversion method

1. Propose an algorithm to simulate random variates for the following distribution functions:
   (a) the generalized extreme value distribution $\mathcal{GEV}(\mu, \sigma, \xi)$;
   (b) the log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$;
   (c) the log-logistic distribution $\mathcal{LL}(\alpha, \beta)$.

2. When we model operational risk losses, we are interested in the conditional random variable $L = X \mid X \geq H$ where H is a given threshold.
   (a) How can we simulate L if we consider a random number generator of X?
   (b) Let $\mathbf{F}_X$ be the distribution function of X. Give the conditional distribution $\mathbf{F}_L$.
   (c) Find the inverse function $\mathbf{F}_L^{-1}$ and propose an algorithm to simulate L.
   (d) Compare the two algorithms in terms of efficiency.
   (e) Apply algorithms (a) and (c) to the log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$. We assume that $\mu = 7$, $\sigma = 2.3$ and $H = \$50\,000$. Simulate 100 random numbers[58] and draw the scatterplot between the uniform random numbers u_i and the random variates L_i.

3. We consider the extreme order statistics $X_{1:n} = \min(X_1, \ldots, X_n)$ and $X_{n:n} = \max(X_1, \ldots, X_n)$.
   (a) How can we simulate $X_{1:n}$ and $X_{n:n}$ if we have a random generator for X_i?
   (b) Calculate the distribution functions $\mathbf{F}_{1:n}$ and $\mathbf{F}_{n:n}$. Deduce an efficient algorithm to simulate $X_{1:n}$ and $X_{n:n}$.
   (c) Using the previous algorithm, simulate 1000 random numbers of $X_{1:50}$ and $X_{50:50}$ when $X_i \sim \mathcal{N}(0,1)$.

[58] We can use the Lewis-Goodman-Miller generator with a seed equal to 123 456.

13.4.2 Simulating random numbers using the transformation method

1. We consider the random variable X, whose probability density function is given by:

$$f(x) = \frac{\beta^\alpha x^{-\alpha-1} e^{-\beta/x}}{\Gamma(\alpha)}$$

Calculate the density function of $Y = 1/X$. Deduce that X follows the inverse-gamma distribution $\mathcal{IG}(\alpha, \beta)$. Find an algorithm to simulate X.

2. Let $X \sim \mathcal{G}(\alpha, \beta)$ be a gamma distributed random variable.

   (a) Show that the case $\alpha = 1$ corresponds to the exponential distribution with parameter $\lambda = \beta$:

$$\mathcal{G}(1, \beta) = \mathcal{E}(\beta)$$

   Deduce an algorithm to simulate $\mathcal{G}(1, \beta)$.

   (b) When α is an integer and is equal to n, show that $X \sim \mathcal{G}(n, \beta)$ is the sum of n independent exponential random variables with parameter β. Deduce an algorithm to simulate $\mathcal{G}(n, \beta)$.

3. Let $X \sim \mathcal{B}(\alpha, \beta)$ be a beta distributed random variable.

   (a) We note $Y \sim \mathcal{G}(\alpha, \delta)$ and $Z \sim \mathcal{G}(\beta, \delta)$ two independent gamma distributed random variable. Show that[59]:

$$X = \frac{Y}{Y + Z}$$

   (b) Deduce an algorithm to simulate X.

4. The polar method considers the random vector (X, Y) defined by:

$$\begin{cases} X = R \cdot \cos\Theta \\ Y = R \cdot \cos\Theta \end{cases}$$

where R and Θ are two independent random variables. We assume that $R \sim \mathbf{F}_R$ and $\Theta \sim \mathcal{U}_{[0,2\pi]}$.

   (a) Show that the joint density of (X, Y) is equal to:

$$f_{X,Y}(x, y) = \frac{f_R\left(\sqrt{x^2 + y^2}\right)}{2\pi\sqrt{x^2 + y^2}}$$

   Deduce the expression of the density function $f_X(x)$ of X.

   (b) We assume that $R = \sqrt{2 \cdot E}$ where $E \sim \mathcal{E}(1)$.

      i. Show that $f_R(r) = re^{-r^2/2}$.

      ii. Calculate $f_X(x)$.

      iii. Deduce the Box-Muller algorithm to simulate normal distributed random variables.

[59] Hint: consider the change of variables $X = Y/(Y + Z)$ and $S = Y + Z$, and calculate the joint density function $f_{X,S}$.

(c) Bailey (1994) proposed to simulate the Student's t_ν distribution with the polar method. For that, he considered the distribution:

$$\mathbf{F}_R(r) = 1 - \left(1 + \frac{r^2}{\nu}\right)^{-\nu/2}$$

where $r \geq 0$ and $\nu > 0$.

i. Calculate $f_R(r)$.

ii. Show that:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\left(1 + \frac{x^2+y^2}{\nu}\right)^{-\nu/2-1}$$

iii. Find the expression[60] of $f_X(x)$. Deduce that X is a Student's t random variable with ν degrees of freedom.

iv. Find an algorithm to simulate R.

v. Deduce an algorithm to simulate Student's t_ν random variables.

vi. What is the main difference with the Box-Muller algorithm?

13.4.3 Simulating random numbers using rejection sampling

1. We consider the beta distributed random variable $X \sim \mathcal{B}(\alpha, \beta)$. We assume that $\alpha \geq 1$ and $\beta \geq 1$.

   (a) We use the proposal density function $g(x) = 1$. Calculate the function $h(x)$ defined as follows:

   $$h(x) = \frac{f(x)}{g(x)}$$

   Show that $h(x)$ achieves its maximum at the point:

   $$x^\star = \frac{\alpha - 1}{\alpha + \beta - 2}$$

   Deduce the value of c that maximizes the acceptance ratio.

   (b) Plot the functions $f(x)$ and $cg(x)$ for the following parameters (α, β): $(1.5, 1.5)$, $(3, 2)$, $(1, 8)$ and $(5, 7)$. Comment on these results.

   (c) Propose an algorithm to simulate $\mathcal{B}(\alpha, \beta)$ when $\alpha \geq 1$ and $\beta \geq 1$.

2. We consider the beta distributed random variable $X \sim \mathcal{B}(\alpha, \beta)$. We assume that $\alpha < 1$ and $\beta \geq 1$.

   (a) We use the proposal density function $g(x) = \alpha x^{\alpha-1}$. Find the value of c that maximizes the acceptance ratio.

   (b) Give an algorithm to simulate the random variable $X \sim \mathbf{G}$.

   (c) Give an algorithm to simulate $\mathcal{B}(\alpha, \beta)$ when $\alpha < 1$ and $\beta \geq 1$.

3. We consider the standard Gaussian random variable $X \sim \mathcal{N}(0, 1)$.

[60] Hint: consider the change of variable $u = \left(1 + \frac{y^2}{\nu + x^2}\right)^{-1}$.

(a) We use the Laplace distribution as the proposal distribution:

$$g(x) = \frac{1}{2}e^{-|x|}$$

Calculate $\mathbf{G}(x)$ and $\mathbf{G}^{-1}(x)$. Give an algorithm to simulate the Laplace distribution.

(b) Find the value of c that maximizes the acceptance ratio. Draw the functions $f(x)$ and $cg(x)$.

(c) Deduce the acceptance-rejection algorithm.

4. We consider the standard gamma random variable $X \sim \mathcal{G}(\alpha)$ where $\alpha \geq 1$.

(a) We use the Cauchy distribution as the proposal distribution:

$$g(x) = \frac{1}{\pi(1+x^2)}$$

Show that:

$$\frac{f(x)}{g(x)} \leq \frac{2\pi}{\Gamma(\alpha)}x^{\alpha+1}e^{-x}$$

Find the value of c that maximizes the acceptance ratio.

(b) We use the Student's t distribution with 2 degrees of freedom as the proposal distribution. Calculate the analytical expression of $\mathbf{G}(x)$. Deduce an algorithm to simulate X.

(c) In the case (b), Devroye (1986) showed that:

$$f(x) \leq cg(x)$$

where:

$$c = \frac{\sqrt{3(4\alpha-1)}}{\Gamma(\alpha)}(\alpha-1)^{\alpha-1}e^{-(\alpha-1)}$$

What is the most efficient method between algorithms (a) and (b)?

5. We consider a discrete random variable X with a finite number of states x_k where $k = 1, \ldots, K$. We note $p(k) = \Pr\{X = x_k\}$ its probability mass function.

(a) We consider the following proposal distribution:

$$q(k) = \frac{1}{K}$$

Find the value of c that maximizes the acceptance ratio.

(b) We consider the following distribution:

x_k	0	1.5	3.3	5.6	8.9
$p(k)$	10%	20%	40%	20%	10%

Simulate 1 000 random numbers using the acceptance-rejection algorithm and draw the histogram of accepted and rejected values. Comment on these results.

13.4.4 Simulation of Archimedean copulas

We recall that an Archimedean copula has the following expression:

$$\mathbf{C}(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

where φ is the generator function.

1. Retrieve the Genest-MacKay algorithm to simulate Archimedean copulas.
2. We assume that $\varphi(u) = (-\ln u)^{\theta}$ with $\theta \geq 1$. Find the corresponding copula.
3. Calculate the conditional distribution $\mathbf{C}_{2|1}$ associated to the previous Archimedean copula. Deduce an algorithm to simulate it.
4. We consider the Frank copula defined as follows:

$$\mathbf{C}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)}{e^{-\theta} - 1} \right)$$

   where $\theta \in \mathbb{R}$. Calculate the conditional distribution $\mathbf{C}_{2|1}$ and deduce an algorithm to simulate this copula.
5. The Ali-Mikhail-Haq family of copulas is given by:

$$\mathbf{C}(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}$$

   where θ lies in $[-1, 1]$. Verify that the generator of this family is:

$$\varphi(u) = \ln \left(\frac{1 - \theta(1 - u)}{u} \right)$$

6. Simulate 5 random vectors of the Gumbel-Hougaard ($\theta = 1.8$), Frank ($\theta = 2.1$) and Ali-Mikhail-Haq copulas ($\theta = 0.6$) by using the following uniform random variates:

v_1	0.117	0.607	0.168	0.986	0.765
v_2	0.498	0.400	0.269	0.892	0.109

13.4.5 Simulation of conditional random variables

Let $Z \sim \mathcal{N}(\mu_z, \Sigma_{z,z})$ be a Gaussian random vector of dimension n_z. We consider the partition $Z = (X, Y)$ where $n_x + n_y = n_z$, $\mu_z = (\mu_x, \mu_y)$ and:

$$\Sigma_{zz} = \begin{pmatrix} \Sigma_{x,x} & \Sigma_{x,y} \\ \Sigma_{y,x} & \Sigma_{y,y} \end{pmatrix}$$

1. Let T be the random vector Y given that $X = x^\star$. Give the distribution of T. Deduce an algorithm to simulate T.
2. We consider the random vector $\tilde{T}$ defined by:

$$\tilde{T} = Y - \Sigma_{y,x} \Sigma_{x,x}^{-1} (X - x^\star)$$

   Show that $\tilde{T} = T$. Deduce an algorithm to simulate T.

3. How can we simulate the Gaussian random vector Z without using the Cholesky decomposition?

4. We assume that the vector of means is $(1, 2, 3)$, the vector of standard deviations is $(1, 0.5, 5)$ and the correlation matrix is:

$$\mathbb{C} = \begin{pmatrix} 1.00 & & \\ 0.50 & 1.00 & \\ 0.20 & 0.30 & 1.00 \end{pmatrix}$$

Apply the algorithm described in Question 3 by using the following independent Gaussian random variates $\mathcal{N}(0, 1)$:

u_1	-1.562	-0.563	-0.573	-0.596	0.984
u_2	0.817	0.845	0.872	-1.303	-0.433
u_3	-0.670	0.126	0.884	-0.918	-0.052

13.4.6 Simulation of the bivariate Normal copula

Let $X = (X_1, X_2)$ be a standard Gaussian vector with correlation ρ. We note $U_1 = \Phi(X_1)$ and $U_2 = \Phi(X_2)$.

1. We note Σ the matrix defined as follows:

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Calculate the Cholesky decomposition of Σ. Deduce an algorithm to simulate X.

2. Show that the copula of (X_1, X_2) is the same that the copula of the random vector (U_1, U_2).

3. Deduce an algorithm to simulate the Normal copula with parameter ρ.

4. Calculate the conditional distribution of X_2 knowing that $X_1 = x$. Then show that:

$$\Phi_2(x_1, x_2; \rho) = \int_{-\infty}^{x_1} \Phi\left(\frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right) \phi(x)\, \mathrm{d}x$$

5. Deduce an expression of the Normal copula.

6. Calculate the conditional copula function $\mathbf{C}_{2|1}$. Deduce an algorithm to simulate the Normal copula with parameter ρ.

7. Show that this algorithm is equivalent to the Cholesky algorithm found in Question 3.

13.4.7 Computing the capital charge for operational risk

We assume that the mapping matrix contains two cells. For each cell, the aggregate loss S_k is defined as:

$$S_k = \sum_{i=1}^{N_k} X_{k,i}$$

where $N_k \sim \mathcal{P}(\lambda_k)$ is the number of losses and $X_{k,i} \sim \mathcal{LN}(\mu_k, \sigma_k^2)$ are the individual losses. The total loss for the bank is then equal:

$$L = S_1 + S_2$$

We calculate the capital-at-risk $\mathrm{CaR}(\alpha)$ for different confidence levels: 90%, 95%, 99% and 99.9%. For that, we use one million simulations.

1. We consider the first cell $k = 1$ and we assume that $\lambda_1 = 50$, $\mu_1 = 7$ and $\sigma_1 = 1.5$. Using 100 replications, calculate the mean and standard deviation of the estimator $\widehat{\mathrm{CaR}}_1(\alpha)$. Do you think that one million simulations is sufficient?
2. Same question for the second cell $k = 2$ if we assume that $\lambda_2 = 100$, $\mu_2 = 5.5$ and $\sigma_2 = 1.8$.
3. Represent the probability density function of $\ln L$ when the aggregate losses S_1 and S_2 are independent and perfectly dependent. Calculate the diversification ratio when we assume that S_1 and S_2 are independent.
4. We assume that the dependence function $\mathbf{C}\langle S_1, S_2\rangle$ is a Normal copula with parameter ρ. Calculate the capital-at-risk of the bank for the following values of ρ: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. Compare these estimates with those obtained with a Gaussian approximation.
5. Same question if the dependence function between S_1 and S_2 is a t_4 copula.
6. Same question if the dependence function between S_1 and S_2 is a t_1 copula.
7. Comment on these results.

13.4.8 Simulating a Brownian bridge

We consider a Brownian bridge $B(t)$ such that $s \leq t \leq u$, $W(s) = w_s$ and $W(u) = w_u$.

1. Find the distribution of the random vector $(W(s), W(t), W(u))$.
2. Calculate the conditional distribution of $W(t)$ given that $W(s) = w_s$ and $W(u) = w_u$.
3. Deduce an algorithm to simulate $B(t)$.

13.4.9 Optimal importance sampling

We consider the estimation of the probability $p = \Pr\{X \geq c\}$ when $X \sim \mathcal{N}(0,1)$.

1. We note $\hat{p}_{\mathrm{MC}}$ the MC estimator of p for one simulation. Calculate $\mathbb{E}[\hat{p}_{\mathrm{MC}}]$ and $\mathrm{var}(\hat{p}_{\mathrm{MC}})$. What is the probability distribution of $\hat{p}_{\mathrm{MC}}$?
2. Let $\mathcal{N}(\mu, \sigma^2)$ be the importance sampling distribution. Give the expression of the IS estimator $\hat{p}_{\mathrm{IS}}$ for one simulation. Calculate $\mathbb{E}[\hat{p}_{\mathrm{IS}}]$ and $\mathrm{var}(\hat{p}_{\mathrm{IS}})$. What do you notice about the probability distribution of $\hat{p}_{\mathrm{IS}}$?
3. We assume that $c = 3$. Calculate $\mathrm{var}(\hat{p}_{\mathrm{MC}})$. Draw the relationship between μ and $\mathrm{var}(\hat{p}_{\mathrm{IS}})$ when σ is respectively equal to 0.8, 1, 2 and 3. Find the optimal value of μ. What hypothesis can we make?
4. We assume that σ is equal to 1. Find the first-order condition if we would like to select the optimal important sampling scheme. Draw the relationship between c and the optimal value of μ. Deduce an heuristic approach for defining a good IS scheme.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 14

Stress Testing and Scenario Analysis

In 1996, the Basel Committee proposed that banks regularly conduct stress testing programs in the case of market risk. The underlying idea was to identify events that could generate exceptional losses and understand the vulnerability of a bank. The use of stress tests has been increasing with the implementation of the Basel II Accord. Indeed, stress testing is the core of the Pillar 2 supervision, in particular for credit risk. At the same time, stress testing programs have been extended to the financial sector taken as a whole. In this case, they do not concern a given financial institution, but a set of banks or institutions. For example, the financial sector assessment program (FSAP) conducted by the International Monetary Fund and the World Bank measures the resilience of the financial sector of a given country or region. In Europe, EBA and ECB are in charge of the EU-wide stress testing. Since the 2008 Global Financial Crisis, they have conducted six stress testing surveys. In the US, the Fed performs every year a stress testing program that concerns the largest 30 banks. This annual assessment includes two related programs: The '*Comprehensive Capital Analysis and Review*' (CCAR) and the '*Dodd-Frank Act stress testing*' (DFAST). The objective of this last program is to evaluate the impact of stressful economic and financial market conditions on the bank capital. Recently, the Basel Committee on Banking Supervision has published a consultative document on stress testing principles. It highlights the growing importance of stress testing in the banking supervision model:

> "*Stress testing is now a critical element of risk management for banks and a core tool for banking supervisors and macroprudential authorities*" (BCBS, 2017a, page 5).

During long times, stress testing mainly concerned market risk, and later credit risk. These last years, it has been extended to other risks: funding risk, liquidity risk, and spillover risk (Tarullo, 2016). Moreover, stress testing is now extended to other financial sectors such as insurance and asset management. For instance, the Financial Stability Board (2017) encourages national financial regulators to conduct system-wide stress testing of asset managers, in particular for measuring the liquidity risk[1]. These views are also supported by IMF (Bouveret, 2017) and some national regulators (AMF, 2017; BaFin, 2017). The counterparty credit risk is another topic where stress testing could help. This explains that ESMA and CFTC have conducted specific stress testing of central counterparty clearing houses. We could then expect that the use of stress testing programs will increase across financial industries in the coming years, not only at the level of financial institutions, but also for regulatory purposes.

[1] "*Although such system-wide stress testing exercises are still in an exploratory stage, over time they may provide useful insights that could help inform both regulatory actions and funds' liquidity risk management practices*" (FSB, 2017, page 23).

14.1 Stress test framework

14.1.1 Definition

14.1.1.1 General objective

There are several definitions of stress testing, because stress tests can be used for different objectives. Lopez (2005) describes stress testing as "*a risk-management tool used to evaluate the potential impact on portfolio values of unlikely, although plausible, events or movements in a set of financial variables*". In this case, stress testing is a complementary tool for VaR analysis. Jorion (2007) considers that stress testing encompasses scenario analysis and the impact of stressed model parameters. Scenario analysis consists in measuring the potential loss due to a given economic or financial stress scenario. For example, the bank could evaluate the impact on its balance sheet if the world GDP decreases by 5% in the next two years. Stress testing of model parameters consists in evaluating the impact of stressed parameters on the P&L or the balance sheet of the bank. For example, the bank could evaluate the impact of more severe LGD parameters on its risk-weighted assets or the impact of higher correlations between banks on its CVA P&L. In the case of market risk, we can use stressed covariance matrices. In this context, stress testing can be viewed as an extension of the historical value-at-risk (Kupiec, 1998). More generally, stress testing aims to provide a forward-looking assessment of losses that would be suffered under adverse economic and financial conditions (BCBS, 2017a). In the case of a trading book, we recall that the loss of Portfolio w is equal to:

$$L_s(w) = P_t(w) - g(\mathcal{F}_{1,s}, \ldots, \mathcal{F}_{m,s}; w)$$

where g is the pricing function and $(\mathcal{F}_{1,s}, \ldots, \mathcal{F}_{m,s})$ is the value of risk factors for the scenario s. When considering the historical value-at-risk, we calculate the quantile of the P&L obtained for n_S historical scenarios of risk factors $(s = 1, \ldots, n_S)$. When considering the stress testing, we evaluate the portfolio loss for only one scenario:

$$L_{\text{stress}}(w) = P_t(w) - g(\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}}; w)$$

However, this scenario represented by the risk factors $(\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}})$ is supposed to be severe. Contrary to the value-at-risk, stress testing is then not built from a probability distribution.

14.1.1.2 Scenario design and risk factors

In the previous section, we feel that the stress scenario $\mathbb{S}$ is given by the set of risk factors $\mathcal{F}_{\text{stress}} = (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}})$. It is only the case when we consider a historical scenario, e.g. the stock market crash in 1987 or the bond market crash in 1994. This type of approach is related to the concept of market price-based stress test, when the stress scenario is entirely defined by a set of market prices, for example the level of the VIX index, the return of the S&P 500 index, etc. However, most of the time, the scenario $\mathbb{S}$ is defined by a set $(\mathbb{S}_1, \ldots, \mathbb{S}_q)$ of q stress factors, which are not necessarily the market risk factors of the pricing function. This is particularly true when we consider hypothetical and macroeconomic stress tests. The difficulty is then to deduce the value of risk factors from the stress scenario:

$$\mathbb{S} = (\mathbb{S}_1, \ldots, \mathbb{S}_q) \Rightarrow \mathcal{F}_{\text{stress}} = (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}})$$

Let us consider the FSAP stress scenarios used for the assessment of the stability of the French banking system (De Bandt and Oung, 2004). They tested 13 stress scenarios: 9 single- and multi-factor shocks (F_1 – F_9) and 4 macroeconomic shocks (M_1 – M_4). We report here the F_1, F_5 and F_9 shocks:

- F_1 flattening of the yield curve due to an increase in interest rates: increase of 150 basis points (bp) in overnight rates, increase of 50 bp in 10-year rates, with interpolation for intermediate maturities;
- F_5 share price decline of 30% in all stock markets;
- F_9 flattening of the yield curve (increase of 150 basis points in overnight rates, increase of 50 bp in 10-year rates) together with a 30% drop in stock markets.

We denote by $\mathbb{S}_1$ and $\mathbb{S}_2$ the stress factors defined by the single-factor shocks F_1 and F_5. We have:

$$\begin{aligned} F_1 &: \mathbb{S}_1 \Rightarrow (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}}) \\ F_5 &: \mathbb{S}_2 \Rightarrow (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}}) \\ F_9 &: (\mathbb{S}_1, \mathbb{S}_2) \Rightarrow (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}}) \end{aligned}$$

We notice that F_9 corresponds to the simultaneous shocks F_1 and F_5. It is obvious that the three shocks will impact differently the market risk factors $(\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}})$. However, the transformation of the stress $\mathbb{S}$ into $\mathcal{F}_{\text{stress}}$ is complex and depends on the modeling process of the financial institution. For instance, we can imagine that most of models will associate to the scenario $\mathbb{S}_1$ a negative impact on stock markets. For Bank A, it could be a 10% drop in stock markets while the model of Bank B may imply a share price decline of 20% in stock markets. It follows that stress testing is highly model-dependent. Let us now consider the M_2 macroeconomic shock:

- M_2 increase to USD 40 in the price per barrel of Brent crude for two years (an increase of 48% compared with USD 27 per barrel in the baseline case), without any reaction from the central bank; the increase in the price of oil leads to an increase in the general rate of inflation and a decline in economic activity in France together with a drop in global demand.

Again, the stress factor $\mathbb{S}_3$ can produce different outcomes in terms of market risk factors depending on the model:

$$M_2 : \mathbb{S}_3 \Rightarrow (\mathcal{F}_{1,\text{stress}}, \ldots, \mathcal{F}_{m,\text{stress}})$$

Therefore, stress testing models are more sensitive to value-at-risk models. This is the main drawback of this approach. For instance, if we want to compare two banks, it is important to describe more precisely the stress scenarios than the shocks above. Moreover, having the stressed market risk factors of the two banks $\mathcal{F}^A_{\text{stress}} = \left(\mathcal{F}^A_{1,\text{stress}}, \ldots, \mathcal{F}^A_{m,\text{stress}}\right)$ and $\mathcal{F}^B_{\text{stress}} = \left(\mathcal{F}^B_{1,\text{stress}}, \ldots, \mathcal{F}^B_{m,\text{stress}}\right)$ is also relevant to understand how the initial shock spreads through the financial system, and the underlying assumptions of the models of banks A and B. The sensitivity to models and assumptions is even more pronounced in the case of liquidity stress tests. Indeed, the model must take into account spillover effects between financial institutions. In the case of funding liquidity, it requires modeling the network between banks, but also the monetary policy reaction function. In the case of market liquidity, the losses will depend on the behavior of all market participants, including asset managers and investors.

The previous introduction shows that we can classify stress scenarios into 4 main categories:

1. historical scenario: "*a stress test scenario that aims at replicating the changes in risk factor shocks that took place in an actual past episode*[2]" (BCBS, 2017a, page 60);

2. hypothetical scenario: "*a stress test scenario consisting of a hypothetical set of risk factor changes, which does not aim to replicate a historical episode of distress*" (BCBS, 2017a, page 60);

3. macroeconomic scenario: "*a stress test that implements a link between stressed macroeconomic factors [...] and the financial sustainability of either a single financial institution or the entire financial system*" (BCBS, 2017a, page 61);

4. liquidity scenario: "*a liquidity stress test is the process of assessing the impact of an adverse scenario on institution's cash flows as well as on the availability of funding sources, and on market prices of liquid assets*" (BCBS, 2017a, page 60).

Concerning hypothetical stress tests, we can also make the distinction between three types of scenarios: baseline, adverse and severely adverse. Since a baseline scenario corresponds to the best forecast of future economic conditions, it is not necessarily a stress scenario but serves as a benchmark. An adverse scenario is a scenario, where the economic conditions are assumed to be worse than for the baseline scenario. The distinction between an adverse and a severely adverse scenario is the probability of occurrence, which is very low for this latter. Therefore, we notice that defining a stress scenario is a two-step process. We first have to select the types of shocks, and then we have to calibrate the severity of the scenario. In Figures 14.1 and 14.2, we have reported the three scenarios of the 2017 Dodd-Frank Act stress test exercises[3] that were developed by the Board of Governors of the Federal Reserve System (2017). The baseline scenario for the United States is a moderate economic expansion, while the US economy experiences a moderate recession in the adverse scenario. The severely adverse scenario is characterized by a severe global recession that is accompanied by a period of heightened stress in corporate loan markets and commercial real estate markets. The baseline, adverse and severely adverse scenarios use the same set of stress factors, but the magnitude of the shocks are different.

14.1.1.3 Firm-specific versus supervisory stress testing

In the 1990s, stress tests were mainly conducted by banks in order to understand their hidden vulnerabilities:

> "*The art of stress testing should give the institution a deeper understanding of the specific portfolios that could be put in jeopardy given a certain situation. The question then would be: Would this be enough to bring down the firm? That way, each institution can know exactly what scenario they do not want to engage in*" (Dunbar and Irving, 1998).

More precisely, stress testing first emerged in trading activities. This explains that stress testing was presented by the 1996 amendment to the capital accord as an additional tool to the value-at-risk. It was an extreme risk measure, a tool for risk management, a requirement in order to validate internal models, but it was not used for calculating the regulatory

[2] According to BCBS (2017a), it may also result from "*a combination of changes in risk factor shocks observed during different past episodes*".

[3] The data are available at the following website: www.federalreserve.gov/supervisionreg/dfast-archive.htm.

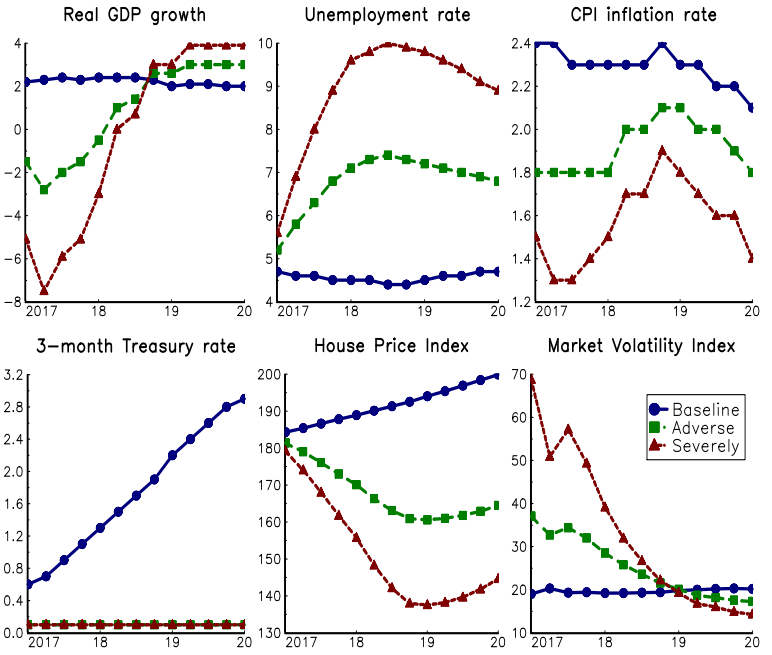


FIGURE 14.1: 2017 DFAST supervisory scenarios: Domestic variables

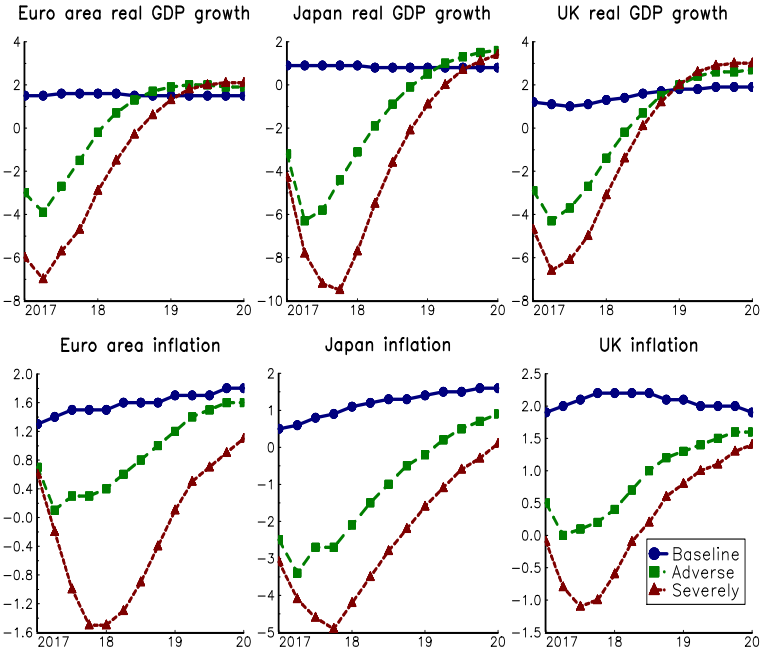


FIGURE 14.2: 2017 DFAST supervisory scenarios: International variables

capital. The Basel 2.5 framework has changed this situation since the capital depends on the stressed value-at-risk. In trading activities, stress scenarios are mainly historical. Besides the vulnerability analysis, stress testing have also been extensively used in the setting of trading limits. In the case of derivatives portfolios, trading limits are defined using sensitivities or VaR metrics. However, some situations can lead the bank to determine hard trading limits based on stress testing:

- some trading portfolios are sensitive to parameters that are unobservable or unstable; for example, a basket option depends on correlations, that can change faster in a crisis period;
- some underlying assets may become less liquid in a period of stress, for example volatility indices, dividends futures, small cap stocks, high yield bonds, etc.

In these cases, stress exposure limits are better than delta or vega exposure limits, because it is difficult to manage portfolios in non-normal situations. In the 2000s, the Basel II Accord has encouraged banks to apply stress testing techniques to credit risk, and some operational risk events such as rogue trading. However, firm-wide stress testing has made little progress before the development of supervisory stress tests (CGFS, 2001, 2005).

Supervisory stress tests starts in 1996 with the amendment to the Basel I Accord. However, they mainly concerned micro-prudential analysis. It was also the case with the Basel II Accord. The development of stress testing for macro-prudential purposes really begins to take off after the Global Financial Crisis. Before 2008, only the financial sector assessment program (FSAP), which was launched by the International Monetary Fund and the World Bank, can be considered as a system-wide stress testing exercise. Since the GFC, supervisory stress tests has become a standard for the different policymakers:

> "*Regulatory stress tests moved from being small-scale, isolated exercises within the broader risk assessment programme, to large-scale, comprehensive risk-assessment programmes in their own right leading directly to policy responses*" (Dent *et al.*, 2016, page 133).

Most of the time, policymakers and supervisors develop concurrent stress tests, meaning that the stress tests are applied to all banks of the system. Generally, these concurrent stress tests are used for setting capital buffers of banks. In this case, it is important to distinguish stress tests under the constant or dynamic balance sheet assumption (Busch *et al.*, 2017). Below, we review three supervisory stress testing frameworks:

- Financial sector assessment program (FSAP)[4]
  The FSAP exercise is conducted by the IMF and the World Bank. It is an in-depth assessment of a country's financial sector. According to the IMF, "*FSAPs analyze the resilience of the financial sector, the quality of the regulatory and supervisory framework, and the capacity to manage and resolve financial crises. Based on its findings, FSAPs produce recommendations of a micro- and macro-prudential nature, tailored to country-specific circumstances*". For instance, FSAPs have been tested for the following countries in 2017: Bulgaria, China, Finland, India, Indonesia, Japan, Lebanon, Luxembourg, Netherlands, New Zealand, Saudi Arabia, Spain, Sweden, Turkey and Zambia. Generally, the FSAP exercise includes one or two stress scenarios.

[4]The FSAP website is www.imf.org/external/np/fsap/fsap.aspx.

- Dodd-Frank Act stress test (DFAST)[5]
  According to the Fed, DFAST is a "*forward-looking quantitative evaluation of the impact of stressful economic and financial market conditions on bank holding companies' capital*". The results of DFAST are incorporated into the comprehensive capital analysis and review (CCAR), which evaluates the vulnerability of each bank on an annual basis. The DFAST exercise includes three types of scenario: baseline, adverse and severely adverse.

- EU-wide stress testing[6]
  EU-wide stress tests are conducted by the European Banking Authority (EBA), the European Systemic Risk Board (ESRB), the European Central Bank (ECB) and the European Commission (EC) in a regular basis, generally every two years[7]. According to the EBA, the aim of such tests is to "*assess the resilience of financial institutions to adverse market developments, as well as to contribute to the overall assessment of systemic risk in the EU financial system*".

Supervisory stress tests are not limited to these three examples, since most of developed central banks also use stress testing approaches, for instance the Bank of England (`www.bankofengland.co.uk/stress-testing`) or the Bank of Japan (`www.boj.or.jp/en/research/brp/fsr/index.htm`).

14.1.2 Methodologies

There are three main approaches for building stress scenarios. The historical approach can be viewed as an extension of the historical value-at-risk. The macroeconomic approach consists in developing hypothetical scenarios based on a macro-econometric model. Hypothetical scenarios can also be generated by the probabilistic approach. In this case, the probability distribution of risk factors is estimated and extreme scenarios are computed analytically or by Monte Carlo simulations.

14.1.2.1 Historical approach

This approach is the first method that have been used by banks in the early 1990s. It consists in identifying the worst period for a given risk factor. For instance, a stress scenario for equity markets may be the one that occurred during the Black Monday (1987) or the collapse of Lehman Brothers (2008). A typical adverse scenario for sovereign bonds is the US interest rate shock in 1994, also known as the '*great bond massacre*'. For currencies and commodities, historical stress scenarios can be calibrated using the Mexican peso crisis in 1994, the Asian crisis in 1997, or the commodity price crash in 2015. This approach is very simple and objective since it is based on past values of risk factors. However, it has two main drawbacks. First, the past worst scenario is not necessarily a good estimate of a future stress scenario. A typical example is the subprime crisis. Second, it is difficult to compare the severity of different historical stress scenarios.

The loss (or drawdown) function is defined by $\mathcal{L}(h) = \min_t R(t;h)$ where $R(t;h)$ is the asset return for the period $[t, t+h]$. In Table 14.1, we have reported the 5 maximum values of $\mathcal{L}(h)$ for the S&P 500 index and different values of h. For instance, the maximum of the daily drawdown is reached on 19 October 1987, where we observe a daily return of -20.47%. On 15 October 2008, a loss of -9% is observed. If we consider a monthly period, the maximum loss is about 30%. In Figure 14.3, we have reported the drawdown function

[5] The DFAST website is `www.federalreserve.gov/supervisionreg/dfa-stress-tests.htm`.
[6] The corresponding website is `www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing`.
[7] They took place in 2009, 2010, 2011, 2014, 2016 and 2018.

TABLE 14.1: Worst historical scenarios of the S&P 500 index

Sc.	1D		1W		1M	
1	1987-10-19	−20.47	1987-10-19	−27.33	2008-10-27	−30.02
2	2008-10-15	−9.03	2008-10-09	−18.34	1987-10-26	−28.89
3	2008-12-01	−8.93	2008-11-20	−17.43	2009-03-09	−22.11
4	2008-09-29	−8.79	2008-10-27	−13.85	2002-07-23	−19.65
5	1987-10-26	−8.28	2011-08-08	−13.01	2001-09-21	−16.89
Sc.	2M		3M		6M	
1	2008-11-20	−37.66	2008-11-20	−41.11	2009-03-09	−46.64
2	1987-10-26	−31.95	1987-11-30	−30.17	1974-09-13	−34.33
3	2002-07-23	−27.29	1974-09-13	−28.59	2002-10-09	−31.29
4	2009-03-06	−26.89	2002-07-23	−27.55	1962-06-27	−26.59
5	1962-06-22	−23.05	2009-03-09	−25.63	1970-05-26	−25.45

$\mathcal{L}(h)$. We notice that the drawdown increases with the time period at the beginning, but decreases when the time period is sufficiently long. The maximum loss is called the maximum drawdown:

$$\mathcal{MDD} = \min_{\Delta t} \mathcal{L}(\Delta t)$$

In the case of the S&P 500 index, the maximum drawdown is equal to −56.8% and has been observed between 9 October 2007 and 9 March 2009.

Remark 173 *In practice, the maximum drawdown is calculated using this formula:*

$$\mathcal{MDD} = -\max_t \left(\frac{\max_{[0,t]} P_t - P_t}{\max_{[0,t]} P_t} \right)$$

where P_t is the asset price or the risk factor.

The choice of the lag window h is important. Indeed, defining a stress scenario of −30% for US stocks is not the same if the time period is one day, one week or one month. Another important factor is the time period. For instance, a 50% drawdown for US stocks is observed many times in the last 50 years. However, it is not the same thing to consider the subprime crisis, the dot.com crisis or the 1973-1974 crisis of the stock market. Even if these three historical periods experience similar losses for stocks, the fixed income market reacts differently. It is then obvious that defining a stress scenario cannot be reduced to a single number for one risk factor. It is also important to define how the other risk factors will react and be impacted.

14.1.2.2 Macroeconomic approach

The macroeconomic approach consists in developing a macroeconometric model and considering an exogenous shock in order to generate adverse stress scenarios. The advantages of this approach are manifold. First, the macroeconomic model takes into account the current economic environment. Stress scenarios are then seen as more plausible than using the historical approach. For example, the drawdowns observed in the stock market in 1974, 2000 and 2008 are comparable in terms of magnitude, but not in terms of economic conditions. The origin of a financial market crisis is different each time. This is true for the stock market, but also for the other asset classes. Macroeconomic modeling may then help to develop the relationships between risk factors and the interconnectedness between asset classes for the next crisis. This is why the macroeconomic approach is certainly not better

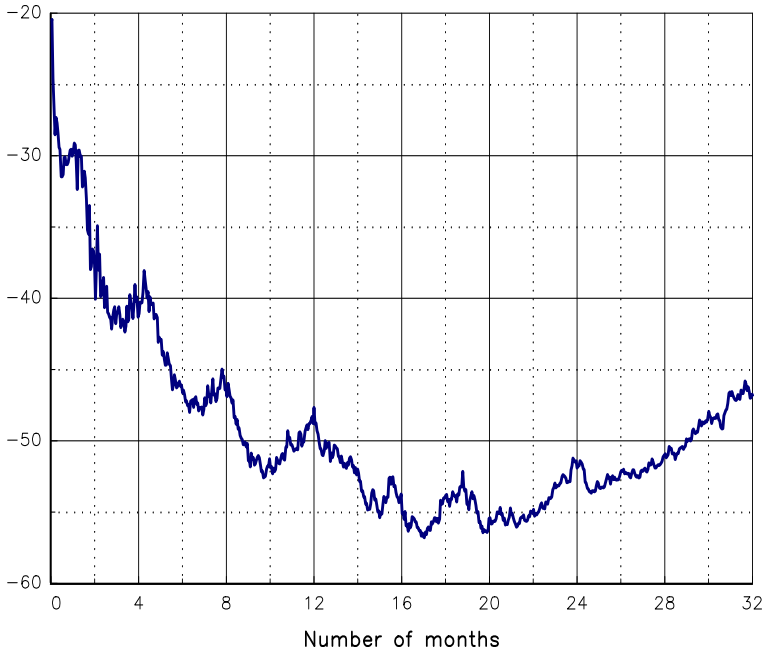


FIGURE 14.3: Loss function of the S&P 500 index

than the historical approach for defining single-factor stress testing, but it is more adapted for building multi-factor stress testing. Therefore, a second advantage is to describe the sequence of the crisis, and the dynamics between risk factors. Another advantage is that many scenarios can be generated by the model. For instance, we have previously seen that the DFAST program defines three scenarios: baseline, adverse and severely adverse. However, it is obvious that more scenarios are generated by the model. At the end, only two or three scenarios are selected, because some of them produce unrealistic outcomes, others may generate similar results, etc.

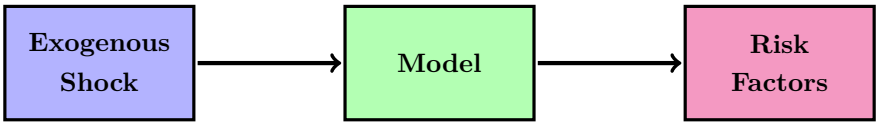


FIGURE 14.4: Macroeconomic approach of stress testing

However, we must be careful with the macroeconomic approach since it has also weaknesses. Stress testing always contains a side of uncertainty. In the case of the historical approach, this is obvious since there is no chance that the next crisis will look like a previous crisis. In the case of the macroeconomic approach, we generally expect to predict the future crisis, but we certainly expect too much (Borio *et al.*, 2014). In Figure 14.4, we have represented the traditional way to describe and think the macroeconomic approach of stress testing. The model uses input parameters (exogenous shocks) in order to produce output parameters (risk factors). In the real life, the impact of the risk factors on financial entities are not direct and deterministic. Indeed, we generally observe feedback effects from the stressed entities $(E_1, \ldots, E_n)$ on the economic situation (Figure 14.5). For instance, the default of one financial institution may lead monetary authorities to change their interest

rate policy. These feedback effects are the most challenging point of the macroeconomic stress testing framework.

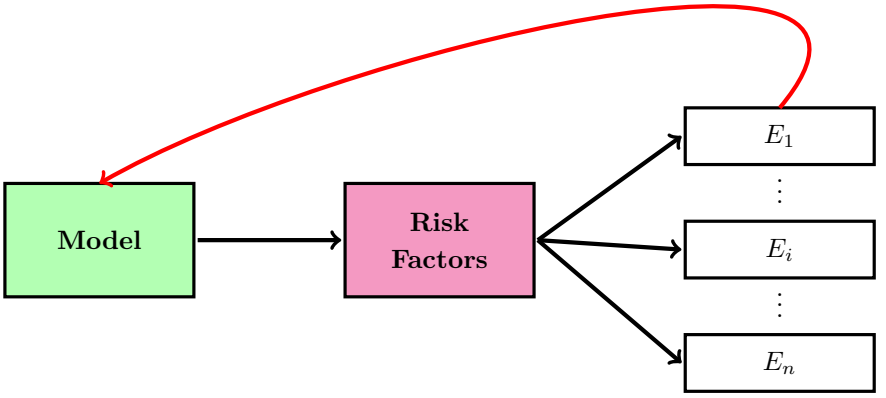


FIGURE 14.5: Feedback effects in stress testing models

A macroeconomic stress testing model is not only a macro-econometric model, that is based on a reduced form or a vector autoregressive process (Sims, 1980). Modeling activity (GDP, unemployment rate, etc.), interest rates and inflation (3M and 10Y interest rates, CPI, etc.) is the first step of the global process. It must also indicate the impact of the economic regime on credit risk parameters (default rates, CDS spreads, recovery rates, etc.) and fundamental variables (earnings, dividends, etc.). Finally, it must define the shocks on financial asset prices (stocks, bonds, commodities, real estate, etc.). For instance, the DFAST program defines 16 domestic and 12 international economic variables:

- Domestic variables: (1) Real GDP growth; (2) Nominal GDP growth; (3) Real disposable income growth; (4) Nominal disposable income growth; (5) Unemployment rate; (6) CPI inflation rate; (7) 3-month Treasury rate; (8) 5-year Treasury yield; (9) 10-year Treasury yield; (10) BBB corporate yield; (11) Mortgage rate; (12) Prime rate; (13) Dow Jones Total stock market index (Level); (14) House price index (Level); (15) Commercial real estate price index (Level); (16) Market volatility index (Level).
- International variables: (1) Euro area real GDP growth; (2) Euro area inflation; (3) Euro area bilateral dollar exchange rate (USD/euro); (4) Developing Asia real GDP growth; (5) Developing Asia inflation; (6) Developing Asia bilateral dollar exchange rate (F/USD, index); (7) Japan real GDP growth; (8) Japan inflation; (9) Japan bilateral dollar exchange rate (yen/USD); (10) UK real GDP growth; (11) UK inflation; (12) UK bilateral dollar exchange rate (USD/pound).

These variables concern activity, interest rates, inflation but also the prices of financial assets: the scenario for equities is given by the level of the Dow Jones index; the slope of the yield curve defines the scenario for fixed income instruments; the BBB corporate yield, the mortgage rate and the prime rate can be used to shape the scenario for credit products like corporate bonds or CDS; the scenario for real estate is given by house and commercial RE price indices; the level of the VIX indicates the scenario of the implied volatility for options and derivatives; the four exchange rates determine the stress scenario, which is valid for currency markets.

We notice that the DFAST program defines the major trends of the financial asset prices, but not a detailed scenario for each asset class. For equity markets, only the stress scenario for US large cap stocks is specified. Using this figure, one has to deduce the stress scenario for European equities, Japanese equities, EM equities, small cap equities, etc. So, there is room for interpretation. And there is a gap between the stress scenario given by the macroeconomic model and the outcome of the stress scenario. Contrary to the historical approach, the macroeconomic approach requires translating the big trends into the detailed path of risk factors. This step can only be done using parametric models: CAPM or APT models for stocks, Nelson-Siegel model for interest rates, Merton model for credit, etc.

14.1.2.3 Probabilistic approach

Until now, we have presented the outcome of a stress scenario as an extreme loss. However, the term '*extreme*' has little meaning and is not precise. It is obvious that a GDP growth of -10% is more extreme than a GDP growth of -5%. The extreme nature of a stress scenario can then be measured by its severity. However, we may wonder if a GDP growth of -50% is conceivable for instance. There is then a trade-off between the severity of a stress scenario and its probability or likelihood.

At first approximation, a stress scenario can be seen as an extreme quantile or value-at-risk. In this case, the aim of stress testing is not to estimate the maximum loss but an extreme loss. For instance, if we consider the univariate stress scenarios $F_1 - F_9$ of De Bandt and Oung (2004) presented on page 895, the authors indicate that the corresponding frequency is 1% over the last thirty years. In the case of multivariate stress scenarios, the probability of M_2 is equal to 1% while the probability of M_5 is equal to 5%. However, most of the time, the occurrence probability of a stress scenario is not discussed.

Let Y, X_1 and X_2 be three random variables that we would like to stress. These random variables may represent macroeconomic variables, market risk factors or parameters of risk models. We note $\mathbb{S}(Y)$, $\mathbb{S}(X_1)$ and $\mathbb{S}(X_2)$ the corresponding stressed values. Evaluating the likelihood of a stress scenario consists in calculating its probability of occurrence. The calculation depends on the relationship between the portfolio loss $L(w)$ and the random variable to stress. For instance, if the relationship between $L(w)$ and X_1 is decreasing, the probability of the stress $\mathbb{S}(X_1)$ is equal to:

$$\alpha_1 = \Pr\{X_1 \leq \mathbb{S}(X_1)\} = \mathbf{F}_1(\mathbb{S}(X_1))$$

If the relationship between $L(w)$ and X_2 is increasing, we have:

$$\alpha_2 = \Pr\{X_2 \geq \mathbb{S}(X_2)\} = 1 - \mathbf{F}_2(\mathbb{S}(X_2))$$

α_1 and α_2 measures the probability of univariate stress scenarios $\mathbb{S}(X_1)$ and $\mathbb{S}(X_2)$. Similarly, we may compute the joint probability of the stress scenario $(\mathbb{S}(X_1), \mathbb{S}(X_2))$:

$$\begin{aligned}
\alpha_{1,2} &= \Pr\{X_1 \leq \mathbb{S}(X_1), X_2 \geq \mathbb{S}(X_2)\} \\
&= \Pr\{X_1 \leq \mathbb{S}(X_1)\} - \Pr\{X_1 \leq \mathbb{S}(X_1), X_2 \leq \mathbb{S}(X_2)\} \\
&= \mathbf{F}_1(\mathbb{S}(X_1)) - \mathbf{C}_{1,2}(\mathbf{F}_1(\mathbb{S}(X_1)), \mathbf{F}_2(\mathbb{S}(X_2)))
\end{aligned}$$

While the univariate stress scenarios depend on the cumulative distribution functions $\mathbf{F}_1$ and $\mathbf{F}_2$, the bivariate stress scenario also depends on the copula function $\mathbf{C}_{1,2}$ between X_1 and X_2. If we assume that X_1 and X_2 are independent, we obtain:

$$\begin{aligned}
\alpha_{1,2} &= \alpha_1 - \alpha_1 \cdot (1 - \alpha_2) \\
&= \alpha_1 \cdot \alpha_2
\end{aligned}$$

If we assume that X_1 and X_2 are perfectly dependent — $\mathbf{C}_{1,2} = \mathbf{C}^+$, we have[8]:

$$\begin{aligned}\alpha_{1,2} &= \alpha_1 - \min(\alpha_1, 1 - \alpha_2) \\ &= 0\end{aligned}$$

This result is perfectly normal because X_1 and X_2 impact $L(w)$ in an opposite way. If $\mathbf{C}_{1,2} = \mathbf{C}^-$, we have:

$$\begin{aligned}\alpha_{1,2} &= \alpha_1 - \max(0, \alpha_1 - \alpha_2) \\ &= \min(\alpha_1, \alpha_2)\end{aligned}$$

We deduce that the probability of the bivariate stress scenario is lower than the probability of the univariate stress scenarios:

$$0 \leq \alpha_{1,2} \leq \min(\alpha_1, \alpha_2)$$

We now consider that the stress scenario $\mathbb{S}(Y)$ is deduced from $\mathbb{S}(X_1)$ and $\mathbb{S}(X_2)$. The conditional probability of the stress scenario $\mathbb{S}(Y)$ is then given by:

$$\alpha = \Pr\{Y \leq \mathbb{S}(Y) \mid (X_1, X_2) = (\mathbb{S}(X_1), \mathbb{S}(X_2))\}$$

It follows that α depends on the conditional distribution of Y given X_1 and X_2. These three concepts of probability — univariate, joint and conditional — drive the quantitative approaches of stress testing that are presented below. They highlight the importance of quantifying the likelihood of the stress scenario, that is the probability of outcomes.

14.2 Quantitative approaches

The previous breakdown is used to classify the models into three main categories. The univariate case generally consists in modeling the probability distribution of a risk factor in an extreme situation. It is generally based on the extreme value theory. The multivariate case is a generalization of the first approach, and requires specifying the dependence between the risk factors. Copula functions are then the right tool for this task. The third approach uses more or less complex econometric models, in particular time series models.

14.2.1 Univariate stress scenarios

Let X be the random variable that produces the stress scenario $\mathbb{S}(X)$. If X follows the probability distribution $\mathbf{F}$, we have[9] $\Pr\{X \leq \mathbb{S}(X)\} = \mathbf{F}(\mathbb{S}(X))$. Given a stress scenario $\mathbb{S}(X)$, we may deduce its severity:

$$\alpha = \mathbf{F}(\mathbb{S}(X))$$

We may also compute the stressed value given the probability of occurrence α:

$$\mathbb{S}(X) = \mathbf{F}^{-1}(\alpha)$$

Even if this framework is exactly the approach used by the value-at-risk, there is a big difference between value-at-risk and stress testing. Indeed, the probability α used for stress testing is much lower than for value-at-risk.

[8] We recall that $\alpha_1 \approx 0$ and $\alpha_2 \approx 0$.

[9] We assume that the relationship between $L(w)$ and X is decreasing.

TABLE 14.2: Probability (in %) associated to the return period $\mathcal{T}$ in years

Return period	1	5	10	20	30	50
Daily	0.3846	0.0769	0.0385	0.0192	0.0128	0.0077
Weekly	1.9231	0.3846	0.1923	0.0962	0.0641	0.0385
Monthly	8.3333	1.6667	0.8333	0.4167	0.2778	0.1667
$1 - \alpha_{\text{GEV}}$	7.6923	1.5385	0.7692	0.3846	0.2564	0.1538

We recall that the return period $\mathcal{T}$ is related to the probability α by the relationship $\mathcal{T} = \alpha^{-1}$. We deduce that $\alpha = \mathcal{T}^{-1}$. In Table 14.2, we report the probability α for different return periods and different frequencies (daily, weekly and monthly). In the case where $\mathbf{F}$ is the cumulative distribution function of daily returns[10], the probability α is equal to 0.0769% when $\mathcal{T}$ is equal to 5 years, and 0.0128% when $\mathcal{T}$ is equal to 30 years. There are extreme probabilities in comparison to the confidence level $\alpha = 1\%$ for the value-at-risk. Therefore, we can use the extreme value theory to calculate these quantities. We reiterate that:

$$\mathcal{T} = \alpha^{-1} = n \cdot (1 - \alpha_{\text{GEV}})^{-1}$$

where n is the length of the block maxima[11].

TABLE 14.3: GEV parameter estimates (in %) of MSCI USA and MSCI EMU indices

Parameter	Long position		Short position	
	MSCI USA	MSCI EMU	MSCI USA	MSCI EMU
μ	1.242	1.572	1.317	1.599
σ	0.720	0.844	0.577	0.730
ξ	19.363	21.603	26.341	26.494

TABLE 14.4: Stress scenarios (in %) of MSCI USA and MSCI EMU indices

Year	Long position		Short position	
	MSCI USA	MSCI EMU	MSCI USA	MSCI EMU
5	−5.86	−7.27	5.69	7.16
10	−7.06	−8.83	7.01	8.84
25	−8.92	−11.29	9.17	11.60
50	−10.56	−13.49	11.18	14.17
75	−11.62	−14.94	12.54	15.91
100	−12.43	−16.05	13.59	17.26
Extreme statistic	−9.51	−10.94	11.04	10.87
$\mathcal{T}^\star$	32.49	22.24	47.87	20.03

Let us consider the MSCI USA and MSCI EMU indices from 1990 to 2017. We calculate the daily returns R_t. Then we take the block maxima ($X = R_t$) and the block minima ($X = -R_t$) for modeling short and long exposures. Finally, we estimate the parameters (μ, σ, ξ) by the method of maximum likelihood and calculate the corresponding stress scenario $\mathbb{S}(X) =$

[10] We assume that there are 260 trading days in one year.

[11] For instance, when $\mathcal{T}$ is equal to 5 years and n is equal to 20 days, we obtain $\alpha_{\text{GEV}} = 1.5385\%$.

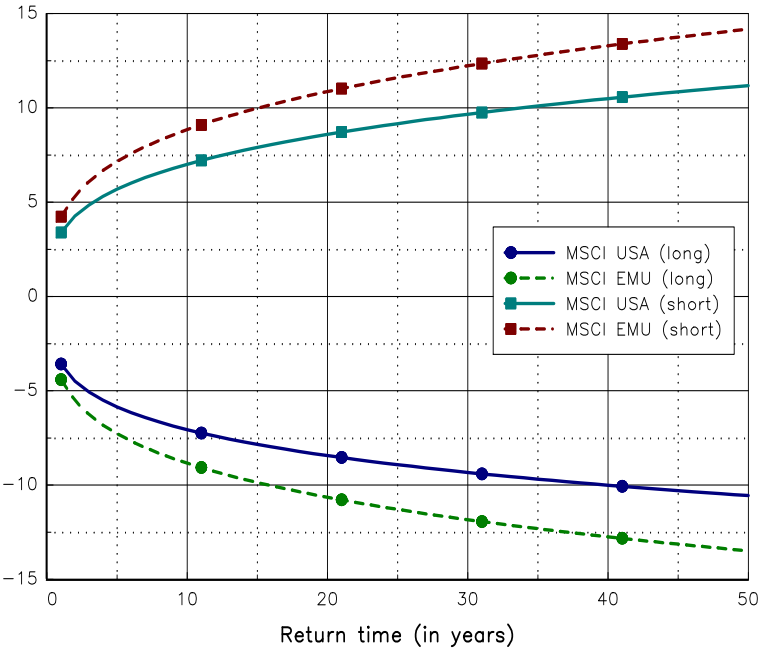


FIGURE 14.6: Stress scenarios (in %) of MSCI USA and MSCI EMU indices

$\hat{\mathbf{G}}^{-1}\left(1-n\mathcal{T}^{-1}\right)$ where $\hat{\mathbf{G}}$ is the estimated GEV distribution. Results are given in Tables 14.3 and 14.4 and Figure 14.6 when the size of blocks is equal to 20 trading days. We notice that the magnitude of stress scenarios is higher for the MSCI EMU index than for the MSCI USA index. For each extreme statistic[12], we have reported the associated return period $\mathcal{T}^\star$. For the MSCI EMU index, $\mathcal{T}^\star$ is close to 20 years. For the MSCI USA index, we obtain a larger return period. This indicates that the stress scenarios for the MSCI USA index may be underestimated. Therefore, it may be appropriate to take the same stress scenario for the two indices, because the differences are not justified.

14.2.2 Joint stress scenarios

14.2.2.1 The bivariate case

Let $X_{n:n,1}$ and $X_{n:n,2}$ be the maximum order statistics of the random variables X_1 and X_2. We note $p = \Pr\{X_{n:n,1} > \mathbb{S}(X_1), X_{n:n,2} > \mathbb{S}(X_2)\}$ the joint probability of stress scenarios $(\mathbb{S}(X_1), \mathbb{S}(X_2))$. We have:

$$\begin{aligned} p &= 1 - \Pr\{X_{n:n,1} \leq \mathbb{S}(X_1)\} - \Pr\{X_{n:n,2} \leq \mathbb{S}(X_1)\} + \\ &\quad \Pr\{X_{n:n,1} \leq \mathbb{S}(X_1), X_{n:n,2} \leq \mathbb{S}(X_2)\} \\ &= 1 - \mathbf{F}_1(\mathbb{S}(X_1)) - \mathbf{F}_2(\mathbb{S}(X_2)) + \mathbf{C}(\mathbf{F}_1(\mathbb{S}(X_1)), \mathbf{F}_2(\mathbb{S}(X_2))) \\ &= \bar{\mathbf{C}}(\mathbf{F}_1(\mathbb{S}(X_1)), \mathbf{F}_2(\mathbb{S}(X_2))) \end{aligned}$$

where $\bar{\mathbf{C}}(u_1, u_2) = 1 - u_1 - u_2 + \mathbf{C}(u_1, u_2)$. We deduce that the failure area is represented by:

$$\left\{(\mathbb{S}(X_1), \mathbb{S}(X_2)) \in \mathbb{R}_+^2 \mid \bar{\mathbf{C}}(\mathbf{F}_1(\mathbb{S}(X_1)), \mathbf{F}_2(\mathbb{S}(X_2))) \leq \frac{n}{\mathcal{T}}\right\}$$

[12] They correspond to the minimum and maximum of daily returns.

Given a return period $\mathcal{T}$, we don't have a unique joint stress scenario $(\mathbb{S}(X_1), \mathbb{S}(X_2))$, but an infinite number of bivariate stress scenarios.

The previous result argues for computing the implied return period of a given scenario, and not the opposite:

$$\mathcal{T} = \frac{n}{\bar{\mathbf{C}}\left(\mathbf{F}_1\left(\mathbb{S}(X_1)\right), \mathbf{F}_2\left(\mathbb{S}(X_2)\right)\right)}$$

In the univariate case, the implied return period of the stress $\mathbb{S}(X_i)$ is equal to:

$$\mathcal{T}_i = \frac{n}{1 - \mathbf{F}_i\left(\mathbb{S}(X_i)\right)}$$

Since an extreme value copula satisfies the property $\mathbf{C}^{\perp} \prec \mathbf{C} \prec \mathbf{C}^{+}$, we deduce that:

$$\max\left(\mathcal{T}_1, \mathcal{T}_2\right) \leq \mathcal{T} \leq n\mathcal{T}_1\mathcal{T}_2$$

In Table 14.5, we report the upper and lower bounds of $\mathcal{T}$ for different values of n, $\mathcal{T}_1$ and $\mathcal{T}_2$ by assuming that a year contains 260 trading days. We observe that the range of $\mathcal{T}$ is wide. For instance, when n is equal to 20 days, and $\mathcal{T}_1$ and $\mathcal{T}_2$ are equal to 5 years, the return period of the joint stress scenario is equal to 5 years if the two scenarios are completely dependent and 325 years if they are independent.

TABLE 14.5: Upper and lower bounds of the return time $\mathcal{T}$ (in years)

n (in days)	$\mathcal{T}_1$	$\mathcal{T}_2$	Lower bound	Upper bound
1	5	5	5	6500
5	5	5	5	1300
20	5	5	5	325
260	5	5	5	25
260	10	5	10	50
260	1	1	1	1

We consider the previous example with MSCI USA and EMU indices. We have reported the failure area in Figure 14.7. For that, we have estimated the copula $\mathbf{C}$ by assuming a Gumbel copula function:

$$\mathbf{C}(u_1, u_2) = \exp\left(-\left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right)^{1/\theta}\right)$$

We estimate θ by the method of maximum likelihood for each quadrant and obtain the following results:

MSCI USA MSCI EMU	Positive Positive	Positive Negative	Negative Negative	Negative Positive
$\hat{\theta}$	1.7087	1.4848	1.7430	1.4697

This means that $\hat{\theta}$ is equal to 1.7087 if the stress for MSCI USA and EMU indices are both positive. We have also reported the solution in the two extremes cases $\mathbf{C}^{\perp}$ and $\mathbf{C}^{+}$. We observe that the dependence plays a major role when considering joint scenarios. For instance, if we consider a scenario of -10% for the MSCI USA index and -10% for the MSCI EMU index, the return period is respectively equal to 39.9, 55.1 and 8 197 years for the product, Gumbel and Fréchet copulas.

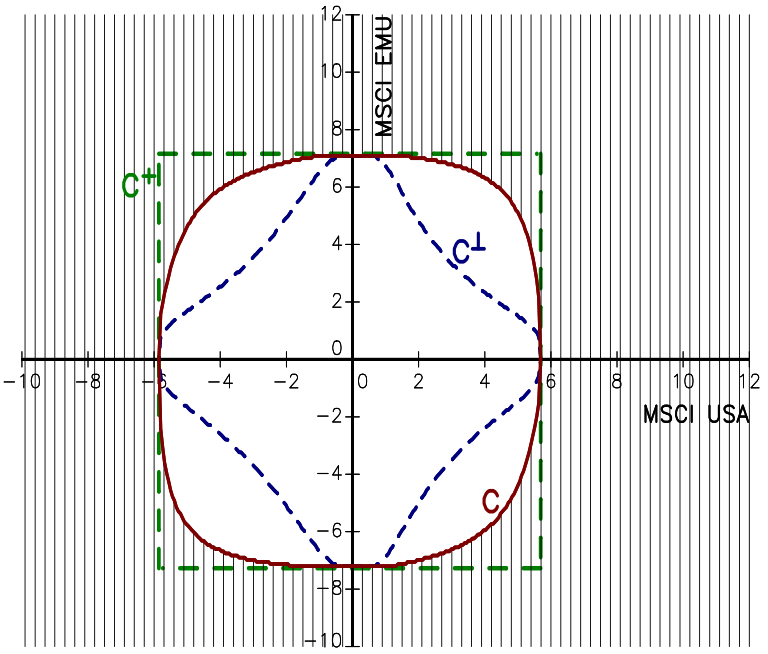


FIGURE 14.7: Failure area of MSCI USA and MSCI EMU indices (blockwise dependence)

Remark 174 *The previous exercise illustrates the limits of blockwise analysis. Let us consider the case of a negative stress for the MSCI USA index and a positive stress for the MSCI EMU index. When n is equal to 20 days, we calculate for each block the worst daily return for the first index and the best daily return for the second index. However, during 4 weeks, these two extreme returns do not certainly occur the same day. It follows that the dependence is overestimated for the two quadrants (Positive, Negative) and (Negative, Positive). This is why it is better to estimate the copula function using daily returns and not blockwise data. In this case, we obtain the results given in Figure 14.8.*

14.2.2.2 The multivariate case

In the multivariate case, the failure area is defined by:

$$\left\{(\mathbb{S}(X_1),\ldots,\mathbb{S}(X_p)) \in \mathbb{R}_+^p \mid \bar{\mathbf{C}}(\mathbf{F}_1(\mathbb{S}(X_1)),\ldots,\mathbf{F}_m(\mathbb{S}(X_p))) \leq \frac{n}{\mathcal{T}}\right\}$$

where:

$$\bar{\mathbf{C}}(u_1,\ldots,u_p) = \sum_{i=0}^{p}\left[(-1)^i \sum_{\mathbf{v}\in\mathcal{Z}(p-i,p)} \mathbf{C}(\mathbf{v})\right]$$

and $\mathcal{Z}(m,p)$ denotes $\{\mathbf{v}\in[0,1]^p \mid \sum_{i=1}^{p} \mathbb{1}\{v_i = 1\} = m\}$. In the case $p = 2$, we retrieve the previous expression:

$$\bar{\mathbf{C}}(u_1,u_2) = 1 - u_1 - u_2 + \mathbf{C}(u_1,u_2)$$

When p is equal to 3, we obtain:

$$\begin{aligned}\bar{\mathbf{C}}(u_1,u_2,u_3) &= 1 - u_1 - u_2 - u_3 + \\ &\quad \mathbf{C}(u_1,u_2) + \mathbf{C}(u_1,u_3) + \mathbf{C}(u_2,u_3) - \\ &\quad \mathbf{C}(u_1,u_2,u_3)\end{aligned}$$

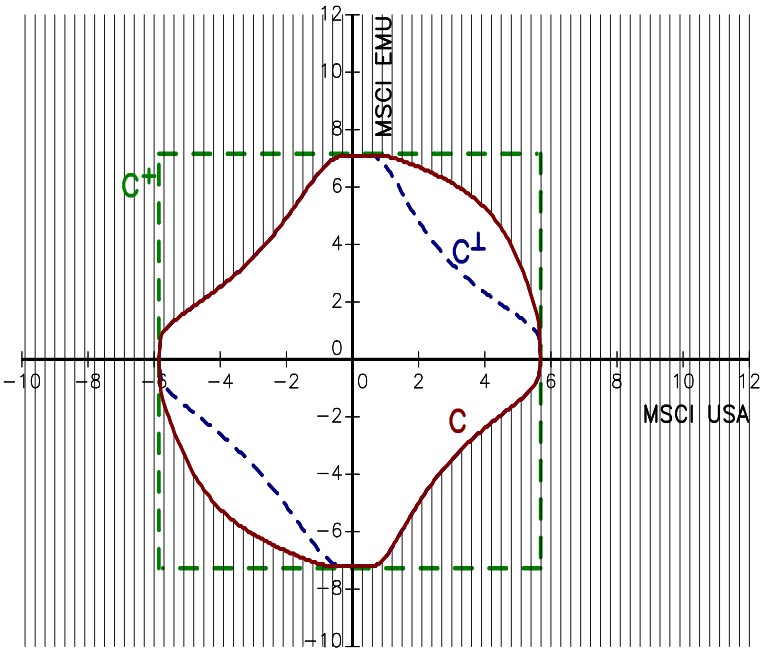


FIGURE 14.8: Failure area of MSCI USA and MSCI EMU indices (daily dependence)

Remark 175 *Bouyé et al. (2000) used this framework for evaluating stress scenarios associated to five commodities of the London Metal Exchange. Since commodity returns are not necessarily positively correlated, they showed that collecting univariate stress scenarios to form a multivariate stress scenario is completely biased. In particular, they presented an example where the return period of univariate stress scenarios is* 5 *years while the return period of the multivariate stress scenario is* 50 000 *years.*

14.2.3 Conditional stress scenarios

In supervisory stress testing, the goal is to impact the parameters of the risk model according to a given scenario. For example, these parameters may be the systematic risk factor in market risk factors, the probability of default and the loss given default in credit risk modeling, or the frequency of the Poisson distribution and the parameters of the severity distribution in operational risk modeling. Therefore, we have to estimate the relationship between these parameters and the variables of the scenario, and deduce their stressed values.

14.2.3.1 The conditional expectation solution

Let us assume a linear model between the independent variable Y and the explanatory variables $X = (X_1, \ldots, X_n)$:

$$Y_t = \beta_0 + \sum_{i=1}^{n} \beta_i X_{i,t} + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$. By assuming that the standard properties of the linear regression model hold, we obtain:

$$\mathbb{E}\left[Y_t\right] = \beta_0 + \sum_{i=1}^{n} \beta_i \mathbb{E}\left[X_{i,t}\right]$$

We can also calculate the conditional expectation of Y_t:

$$\mathbb{E}\left[Y_t \mid X_t = (x_1, \ldots, x_n)\right] = \beta_0 + \sum_{i=1}^{n} \beta_i x_i$$

Given a joint stress scenario $\mathbb{S}(X) = (\mathbb{S}(X_1), \ldots, \mathbb{S}(X_n))$, we deduce the conditional stress scenario of Y and we have:

$$\begin{aligned} \mathbb{S}(Y) &= \mathbb{E}\left[Y_t \mid X_t = (\mathbb{S}(X_1), \ldots, \mathbb{S}(X_n))\right] \\ &= \beta_0 + \sum_{i=1}^{n} \beta_i \mathbb{S}(X_i) \end{aligned}$$

In some cases, assuming a linear relationship is not relevant, in particular for the probability of default or the loss given default. It is then common to use the following transformation (Dees *et al.*, 2017):

$$Z_t = \ln\left(\frac{Y_t}{1 - Y_t}\right)$$

We have:

$$\begin{aligned} Y_t &= \frac{\exp(Z_t)}{1 + \exp(Z_t)} \\ &= \frac{1}{1 + \exp(-Z_t)} \\ &= h(Z_t) \end{aligned}$$

where $h(z)$ is the logit transformation. We verify that $Y_t \in [0, 1]$. Since the statistical model becomes $Z_t = \beta_0 + \sum_{i=1}^{n} \beta_i X_{i,t} + u_t$, we deduce that:

$$\mathbb{E}\left[Y_t \mid X_t = (x_1, \ldots, x_n)\right] = \int_{-\infty}^{\infty} h\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_{i,t} + \omega\right) \frac{1}{\sigma} \phi\left(\frac{\omega}{\sigma}\right) \mathrm{d}\omega \tag{14.1}$$

This conditional expectation can be calculated thanks to numerical integration algorithms.

Remark 176 *The previous model can also be extended in order to take into account fixed effects (panel data) or lag dynamics. For instance, we can use an ARX(p) model:*

$$Y_t = \beta_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{n} \beta_i X_{i,t} + u_t$$

Example 161 *We assume that the probability of default* PD_t *at time* t *is explained by the following linear regression model:*

$$\ln\left(\frac{\mathrm{PD}_t}{1 - \mathrm{PD}_t}\right) = -2.5 - 5g_t - 3\pi_t + 2u_t + \varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}(0, 0.25)$*,* g_t *is the growth rate of the GDP,* π_t *is the inflation rate, and* u_t *is the unemployment rate. The baseline scenario is defined by* $g_t = 2\%$*,* $\pi_t = 2\%$ *and* $u_t = 5\%$*.*

In Figure 14.9, we have reported the probability density function of PD_t for the baseline scenario and the following stress scenario: $g_t = -8\%$, $\pi_t = 5\%$ and $u_t = 10\%$. The conditional expectation[13] is respectively equal to 7.90% and 12.36%. The figure of 7.90% can

[13] We use a Gauss-Legendre quadrature method with an order of 512 for computing the conditional expectation given by Equation (14.1).

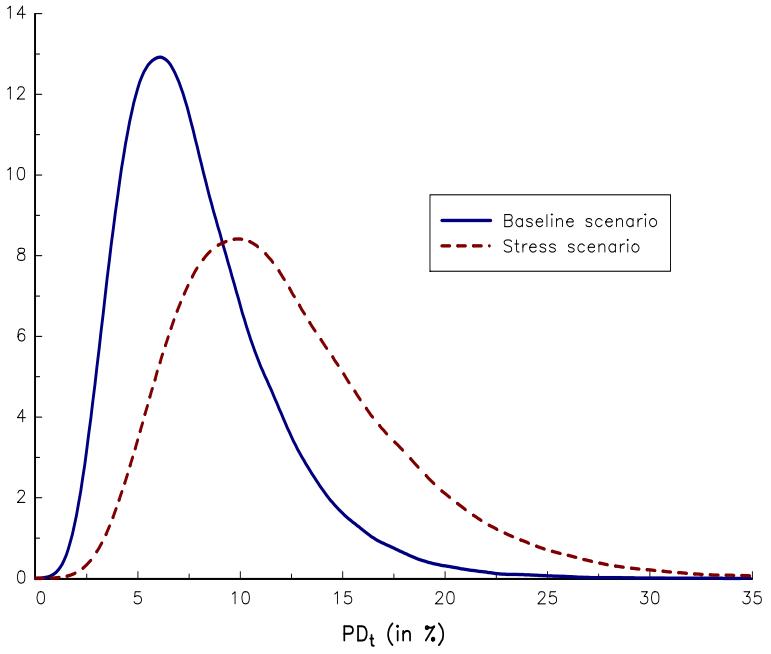


FIGURE 14.9: Probability density function of PD_t

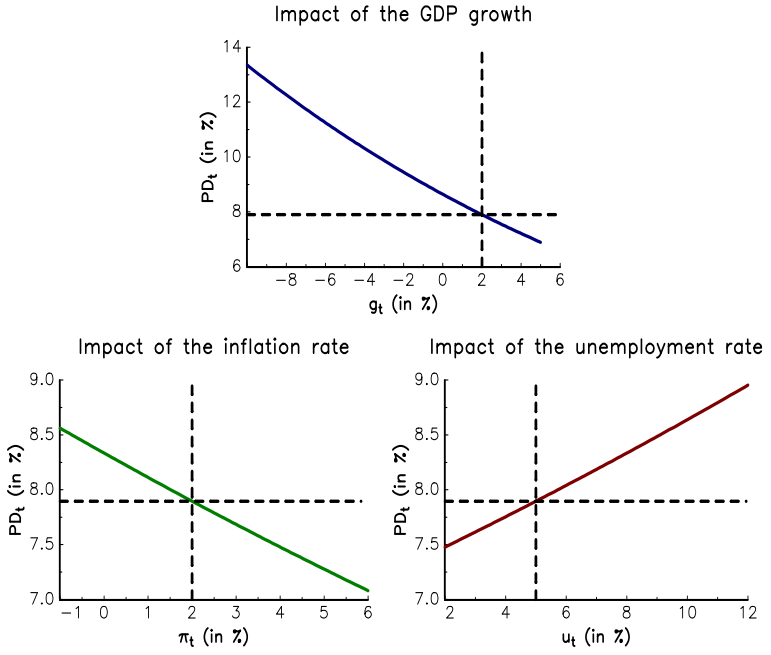


FIGURE 14.10: Relationship between the macroeconomic variables and PD_t

TABLE 14.6: Stress scenario of the probability of default

t	g_t	π_t	u_t	$\mathbb{E}[\text{PD}_t \mid \mathbb{S}(X)]$	$q_{90\%}(\mathbb{S}(X))$
0	2.00	2.00	5.00	7.90	12.78
1	−6.00	2.00	6.00	11.45	18.26
2	−7.00	1.00	7.00	12.47	19.79
3	−9.00	1.00	9.00	14.03	22.14
4	−7.00	1.00	10.00	13.12	20.78
5	−7.00	2.00	11.00	13.01	20.59
6	−6.00	2.00	10.00	12.26	19.49
7	−4.00	4.00	9.00	10.49	16.80
8	−2.00	3.00	8.00	9.70	15.58
9	−1.00	3.00	7.00	9.11	14.68
10	2.00	3.00	6.00	7.82	12.68
11	4.00	3.00	6.00	7.14	11.60
12	4.00	3.00	6.00	7.14	11.60

be interpreted as the long-run (or unconditional) probability of default that is used in the IRB formula. The relationship between the macroeconomic variables and the conditional expectation of PD_t is shown in Figure 14.10. For each panel, we consider the baseline scenario and we vary one parameter each time. In Table 14.6, we consider a stress scenario for the next 3 years, and we indicate the values taken by g_t, π_t and u_t for each quarter t. Then, we calculate the conditional expectation and the conditional quantile at the 90% confidence level of the probability of default PD_t. The stress scenario occurs at time $t = 1$ and propagates until $t = 12$. This is why we initially observe a jump in the probability of default, since it goes from 7.90% to 11.45%. The conditional expectation continues to increase and reaches a top at 14.03%. Then, it decreases and we obtain a new equilibrium after 3 years.

In the previous example, we have also reported the conditional quantile $q_{90\%}(\mathbb{S}(X))$. We observed that its values are larger than those given by the conditional expectation $\mathbb{E}[\text{PD}_t \mid \mathbb{S}(X)]$. These differences raise the question of defining a conditional stress scenario. Indeed, the previous framework defines the conditional stress scenario as the conditional expectation of the linear model $Y_t = \beta_0 + \sum_{i=1}^n \beta_i X_{i,t} + \varepsilon_t$. In this case, the vector of parameters $\beta = (\beta_0, \beta_1, \ldots, \beta_n)$ is estimated by ordinary least squares. We could also define the conditional stress scenario $\mathbb{S}(Y) = q_\alpha(\mathbb{S}(X))$ as the solution of the quantile regression:

$$\Pr\{Y_t \leq q_\alpha(\mathbb{S}) \mid X_t = \mathbb{S}\} = \alpha$$

In this case, we can use the tools presented on pages 613 (parametric approach) and 643 (non-parametric approach). The parametric approach assumes that the probability distribution between Y and X is Gaussian. The non-parametric approach is more adapted when this assumption is not satisfied, for example when the stochastic dependence is not linear.

14.2.3.2 The conditional quantile solution

In order to understand the impact of the dependence on the conditional stress scenario, we consider again the copula framework. If we consider the bivariate random vector (X, Y), using the linear regression is equivalent to assume that:

$$Y_t = \mathbb{E}[Y_t \mid X_t = x] + \varepsilon_t$$

This implies that (X, Y) is a bivariate Gaussian random vector. The average dependence structure between X and Y is then linear and can be represented by the parametric function $y = m(x)$ where $m(x)$ is the conditional expectation function $\mathbb{E}[Y_t \mid X_t = x]$. However, the conditional expectation is not appropriate when (X, Y) is not Gaussian.

The statistical framework We have defined the conditional quantile function $q_\alpha(x)$ as the solution of the equation $\Pr\{Y \leq q_\alpha(x) \mid X = x\} = \alpha$. Let $\mathbf{F}(x, y)$ be the probability distribution of (X, Y). By using the integral transforms $U_1 = \mathbf{F}_x(X)$ and $U_2 = \mathbf{F}_y(Y)$ where $\mathbf{F}_x$ and $\mathbf{F}_y$ are the marginal distributions, we have:

$$\Pr\left\{Y \leq \mathbf{F}_y^{-1}(u_2) \mid X = \mathbf{F}_x^{-1}(u_1)\right\} = \alpha$$

where $u_1 = \mathbf{F}_x(x)$ and $u_2 = \mathbf{F}_y(y) = \mathbf{F}_y(q_\alpha(x))$. It follows that the quantile regression of Y on X is equivalent to solve the following statistical problem:

$$\Pr\{U_2 \leq u_2 \mid U_1 = u_1\} = \alpha$$

or:

$$\frac{\partial}{\partial u_1}\mathbf{C}(u_1, u_2) = \alpha$$

where $\mathbf{C}(u_1, u_2)$ is the copula function associated to probability distribution $\mathbf{F}(x, y)$. We have $u_2 = \mathbf{C}_{2|1}^{-1}(u_1, \alpha)$ where $\mathbf{C}_{2|1}(u_1, u_2) = \partial_1 \mathbf{C}(u_1, u_2)$. It follows that:

$$\mathbf{F}_y(y) = \mathbf{C}_{2|1}^{-1}(\mathbf{F}_x(x), \alpha)$$

Finally, we obtain $y = q_\alpha(x)$ where:

$$q_\alpha(x) = \mathbf{F}_y^{-1}\left(\mathbf{C}_{2|1}^{-1}(\mathbf{F}_x(x), \alpha)\right)$$

Remark 177 *In the case where X and Y are independent, we have $\mathbf{C}(u_1, u_2) = u_1 u_2$, $\partial_1 \mathbf{C}(u_1, u_2) = u_2$, $\mathbf{C}_{2|1}^{-1}(u_1, \alpha) = \alpha$ and:*

$$y = q_\alpha(x) = \mathbf{F}_y^{-1}(\alpha)$$

Therefore, the conditional quantile $q_\alpha(x)$ of Y with respect to $X = x$ is equal to the unconditional quantile $\mathbf{F}_y^{-1}(\alpha)$ of Y.

Some special cases Let us assume that the dependence structure is a Normal copula with parameter ρ. On page 737, we have shown that:

$$\mathbf{C}(u_1, u_2) = \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \mathrm{d}u$$

We deduce that:

$$\partial_1 \mathbf{C}(u_1, u_2) = \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right)$$

Solving the equation $\partial_1 \mathbf{C}(u_1, u_2) = \alpha$ gives:

$$u_2 = \Phi\left(\rho\Phi^{-1}(u_1) + \sqrt{1-\rho^2}\Phi^{-1}(\alpha)\right)$$

The conditional quantile function is then:

$$y = q_\alpha(x) = \mathbf{F}_y^{-1}\left(\Phi\left(\rho\Phi^{-1}(\mathbf{F}_x(x)) + \sqrt{1-\rho^2}\Phi^{-1}(\alpha)\right)\right)$$

In the case of the Student's t copula, we have demonstrated that[14]:

$$\mathbf{C}_{2|1}(u_1, u_2; \rho, \nu) = \mathbf{T}_{\nu+1}\left(\left(\frac{\nu+1}{\nu + \left[\mathbf{T}_\nu^{-1}(u_1)\right]^2}\right)^{1/2} \frac{\mathbf{T}_\nu^{-1}(u_2) - \rho\mathbf{T}_\nu^{-1}(u_1)}{\sqrt{1-\rho^2}}\right)$$

Solving the equation $\mathbf{C}_{2|1}(u_1, u_2; \rho, \nu) = \alpha$ gives:

$$u_2 = \mathbf{T}_\nu\left(\rho\mathbf{T}_\nu^{-1}(u_1) + \sqrt{1-\rho^2}\left(\frac{\nu + \left[\mathbf{T}_\nu^{-1}(u_1)\right]^2}{\nu+1}\right)^{1/2} \mathbf{T}_{\nu+1}^{-1}(\alpha)\right)$$

The conditional quantile function is then:

$$y = q_\alpha(x) = \mathbf{F}_y^{-1}\left(\mathbf{T}_\nu\left(\rho\mathbf{T}_\nu^{-1}(\mathbf{F}_x(x)) + \eta\sqrt{1-\rho^2}\right)\right)$$

where:

$$\eta = \left(\frac{\nu + \left[\mathbf{T}_\nu^{-1}(\mathbf{F}_x(x))\right]^2}{\nu+1}\right)^{1/2} \mathbf{T}_{\nu+1}^{-1}(\alpha)$$

Illustration Let us consider an example with two asset returns $(R_{1,t}, R_{2,t})$. We assume that they follow a bivariate Gaussian distribution with $\mu_1 = 3\%$, $\mu_2 = 5\%$, $\sigma_1 = 10\%$, $\sigma_2 = 20\%$ and $\rho = -20\%$. In Figure 14.11, we have reported the conditional quantile function $R_{2,t} = q_\alpha(R_{1,t})$ for different confidence levels α. We verify that the median regression corresponds to the linear regression. The quantile regression shifts the intercept below when $\alpha < 50\%$ and above when $\alpha > 50\%$. We now assume two variants of this example:

1. the dependence structure is the previous Normal copula, but the marginal distributions follow a Student's t_1 distribution[15];
2. the marginal distributions are the previous Gaussian distributions, but the dependence structure is a Student's t_1 copula.

Results are given in Figures 14.12 and 14.13. We deduce that the linearity of the conditional quantile vanishes if the marginals are not Gaussian or the dependence structure is not Gaussian. In the first case, assuming a linear dependence between $R_{1,t}$ and $R_{2,t}$ implies to overestimate on average the conditional return $R_{2,t} \mid R_{1,t}$ when the first asset has high negative returns. In the second case, we obtain the contrary result.

[14] See page 738.

[15] We have

$$\frac{R_{i,t} - \mu_i}{\sigma_i} \sim t_1$$

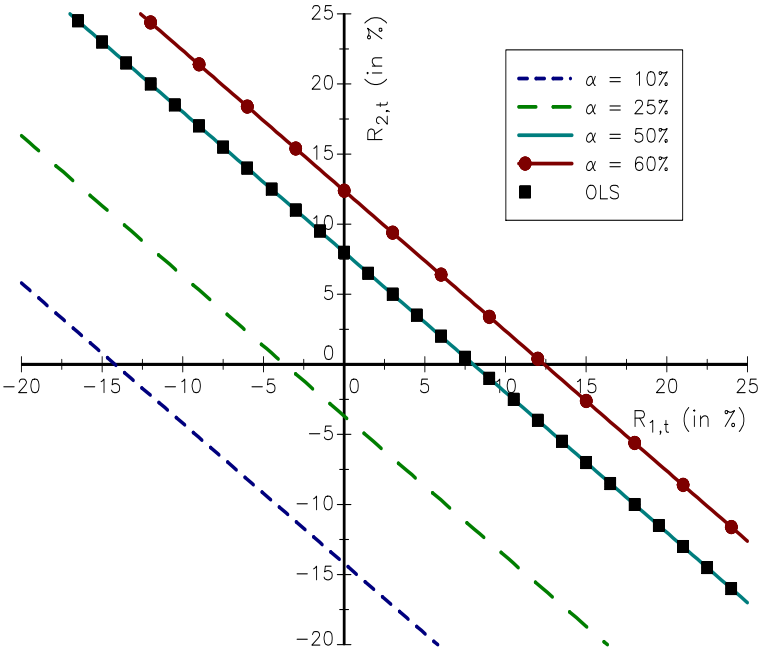


FIGURE 14.11: Conditional quantile (Gaussian distribution)

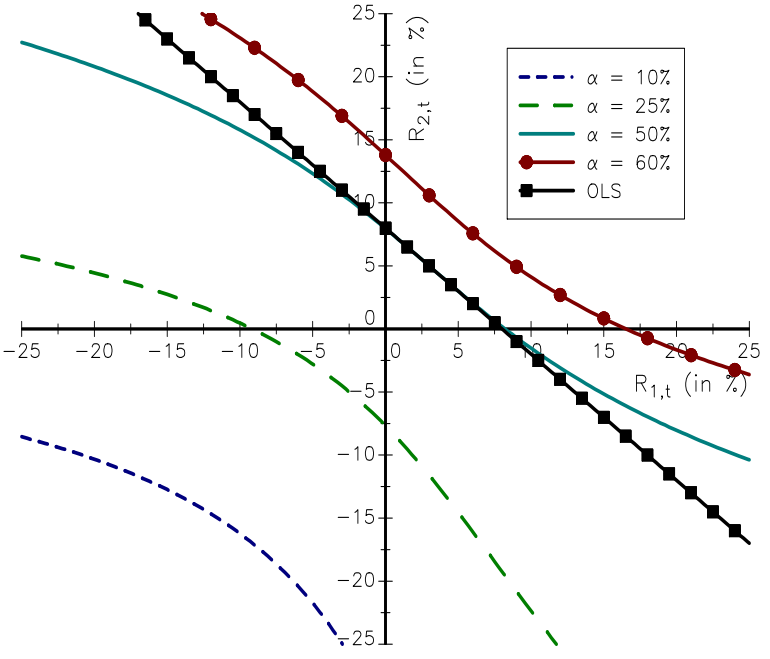


FIGURE 14.12: Conditional quantile (Normal copula and Student's t marginals)

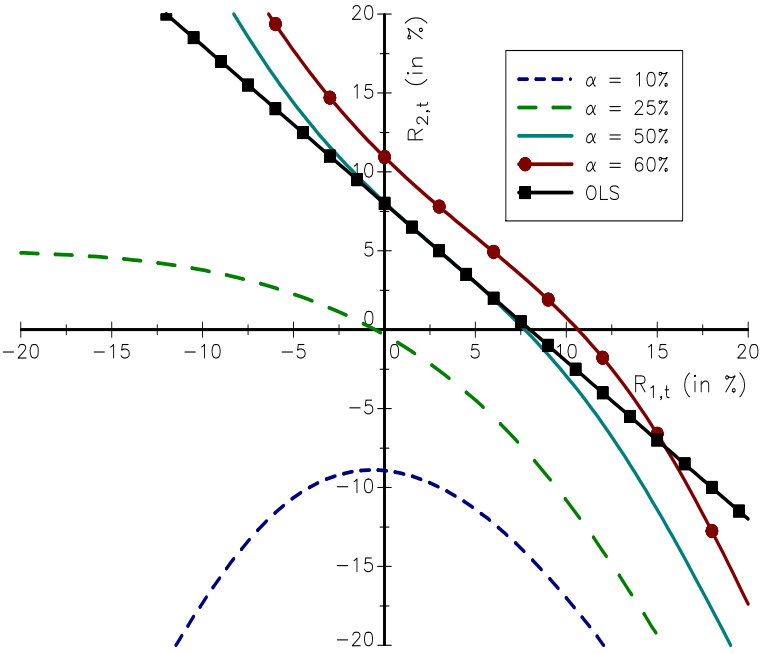


FIGURE 14.13: Conditional quantile (Student's t copula and Gaussian marginals)

14.2.4 Reverse stress testing

According to EBA (2018b), reverse stress test "*means an institution stress test that starts from the identification of the pre-defined outcome (e.g. points at which an institution business model becomes unviable, or at which the institution can be considered as failing or likely to fail) and then explores scenarios and circumstances that might cause this to occur*". The underlying idea is then to identify the set of risk factors that may cause the bankruptcy of the bank (or the financial institution). The difference between stress testing and reverse stress testing can be summarized as follows:

- In stress testing, extreme scenarios of risk factors are used to test the viability of the bank:

$$(\mathbb{S}(\mathcal{F}_1), \ldots, \mathbb{S}(\mathcal{F}_m)) \Rightarrow \mathbb{S}(L(w)) \Rightarrow \begin{cases} D = 0 & \text{if } \mathbb{S}(L(w)) < C \\ D = 1 & \text{otherwise} \end{cases}$$

  Using the set of stressed risk factors, we then compute the corresponding loss $\mathbb{S}(L(w))$ of the portfolio. This stress can cause the default of the bank if the stressed loss is larger than its capital C.

- In reverse stress testing, extreme scenarios of risk factors are deduced from the bankruptcy scenario:

$$D = 1 \Rightarrow \mathbb{RS}(L(w)) \Rightarrow (\mathbb{RS}(\mathcal{F}_1), \ldots, \mathbb{RS}(\mathcal{F}_m))$$

  We first assume that the bank defaults and compute the associated stressed loss. Then, we deduce the implied set of risk factors that has produced the bankruptcy.

Therefore, reverse stress testing can be viewed as an inverse problem, which can face very quickly a curse of dimensionality.

14.2.4.1 Mathematical computation of reverse stress testing

We assume that the portfolio loss is a function of the risk factors:

$$L(w) = \ell(\mathcal{F}_1, \ldots, \mathcal{F}_m; w)$$

Let $(\mathbb{S}(\mathcal{F}_1), \ldots, \mathbb{S}(\mathcal{F}_m))$ be the stress scenario. The associated loss is given by:

$$\mathbb{S}(L(w)) = \ell(\mathbb{S}(\mathcal{F}_1), \ldots, \mathbb{S}(\mathcal{F}_m); w)$$

Reverse stress testing assumes that the financial institution has calculated the reverse stressed loss $\mathbb{RS}(L(w))$ that may produce its bankruptcy. It follows that the reverse stress scenario $\mathbb{RS}$ is the set of risk factors that corresponds to this stressed loss:

$$\mathbb{RS} = \{(\mathbb{RS}(\mathcal{F}_1), \ldots, \mathbb{RS}(\mathcal{F}_m)) : \ell(\mathbb{S}(\mathcal{F}_1), \ldots, \mathbb{S}(\mathcal{F}_m); w) = \mathbb{RS}(L(w))\}$$

Since we have one equation with m unknows, there is not a unique solution except in some degenerate cases. The issue is then to choose the most plausible reverse stress scenario. For instance, we can consider the following optimization program[16]:

$$\begin{aligned}(\mathbb{RS}(\mathcal{F}_1), \ldots, \mathbb{RS}(\mathcal{F}_m)) &= \arg\max \ln f(\mathcal{F}_1, \ldots, \mathcal{F}_m) \\ &\text{s.t.} \quad \ell(\mathbb{S}(\mathcal{F}_1), \ldots, \mathbb{S}(\mathcal{F}_m); w) = \mathbb{RS}(L(w))\end{aligned} \tag{14.2}$$

where $f(x_1, \ldots, x_m)$ is the probability density function of the risk factors $(\mathcal{F}_1, \ldots, \mathcal{F}_m)$.

The linear Gaussian case We assume that $\mathcal{F} \sim \mathcal{N}(\mu_{\mathcal{F}}, \Sigma_{\mathcal{F}})$ and $L(w) = \sum_{j=1}^m w_j \mathcal{F}_j = w^\top \mathcal{F}$. Problem (14.2) becomes:

$$\begin{aligned}\mathbb{RS}(\mathcal{F}) &= \arg\min \frac{1}{2}(\mathcal{F} - \mu_{\mathcal{F}})^\top \Sigma_{\mathcal{F}}^{-1}(\mathcal{F} - \mu_{\mathcal{F}}) \\ &\text{s.t.} \quad w^\top \mathcal{F} = \mathbb{RS}(L(w))\end{aligned}$$

The Lagrange function is:

$$\mathcal{L}(\mathcal{F}; \lambda) = \frac{1}{2}(\mathcal{F} - \mu_{\mathcal{F}})^\top \Sigma_{\mathcal{F}}^{-1}(\mathcal{F} - \mu_{\mathcal{F}}) - \lambda\left(w^\top \mathcal{F} - \mathbb{RS}(L(w))\right)$$

We deduce the first-order condition:

$$\frac{\partial \mathcal{L}(\mathcal{F}; \lambda)}{\partial \mathcal{F}} = \Sigma_{\mathcal{F}}^{-1}(\mathcal{F} - \mu_{\mathcal{F}}) - \lambda w = \mathbf{0}$$

It follows that $\mathcal{F} = \mu_{\mathcal{F}} + \lambda \Sigma_{\mathcal{F}} w$. Since we have $w^\top \mathcal{F} = w^\top \mu_{\mathcal{F}} + \lambda w^\top \Sigma_{\mathcal{F}} w$, we obtain:

$$\lambda = \frac{\mathbb{RS}(L(w)) - w^\top \mu_{\mathcal{F}}}{w^\top \Sigma_{\mathcal{F}} w}$$

and:

$$\mathbb{RS}(\mathcal{F}) = \mu_{\mathcal{F}} + \frac{\Sigma_{\mathcal{F}} w}{w^\top \Sigma_{\mathcal{F}} w}\left(\mathbb{RS}(L(w)) - w^\top \mu_{\mathcal{F}}\right) \tag{14.3}$$

Another approach for solving the inverse problem is to consider the joint distribution of $\mathcal{F}$ and $L(w)$:

$$\begin{pmatrix} \mathcal{F} \\ L(w) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\mathcal{F}} \\ w^\top \mu_{\mathcal{F}} \end{pmatrix}, \begin{pmatrix} \Sigma_{\mathcal{F}} & \Sigma_{\mathcal{F}} w \\ w^\top \Sigma_{\mathcal{F}} & w^\top \Sigma_{\mathcal{F}} w \end{pmatrix}\right)$$

[16] We notice that maximizing the density is equivalent to maximizing its logarithm.

Using Appendix A.2.2.4 on page 1062, we deduce that the conditional distribution of $\mathcal{F}$ given $L(w) = \mathbb{RS}(L(w))$ is Gaussian:

$$\mathcal{F} \mid L(w) = \mathbb{RS}(L(w)) \sim \mathcal{N}\left(\mu_{\mathcal{F}|L(w)}, \Sigma_{\mathcal{F}|L(w)}\right)$$

where:

$$\mu_{\mathcal{F}|L(w)} = \mu_{\mathcal{F}} + \frac{\Sigma_{\mathcal{F}} w}{w^\top \Sigma_{\mathcal{F}} w}\left(\mathbb{RS}(L(w)) - w^\top \mu_{\mathcal{F}}\right)$$

and:

$$\Sigma_{\mathcal{F}|L(w)} = \Sigma_{\mathcal{F}} - \frac{\Sigma_{\mathcal{F}} w w^\top \Sigma_{\mathcal{F}}}{w^\top \Sigma_{\mathcal{F}} w}$$

We know that the maximum of the probability density function of the multivariate normal distribution is reached when the random vector is exactly equal to the mean. We deduce that:

$$\begin{aligned}\mathbb{RS}(X) &= \mu_{\mathcal{F}|L(w)} \\ &= \mu_{\mathcal{F}} + \frac{\Sigma_{\mathcal{F}} w}{w^\top \Sigma_{\mathcal{F}} w}\left(\mathbb{RS}(L(w)) - w^\top \mu_{\mathcal{F}}\right)\end{aligned} \tag{14.4}$$

Example 162 *We assume that* $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2)$, $\mu_{\mathcal{F}} = (5, 8)$, $\sigma_{\mathcal{F}} = (1.5, 3.0)$ *and* $\rho(\mathcal{F}_1, \mathcal{F}_2) = -50\%$. *The sensitivity vector* w *to the risk factors is equal to* $(10, 3)$.

The stress scenario is the collection of univariate stress scenarios at the 99% confidence level:

$$\begin{aligned}\mathbb{S}(\mathcal{F}_1) &= 5 + 1.5 \cdot \Phi^{-1}(99\%) = 8.49 \\ \mathbb{S}(\mathcal{F}_2) &= 8 + 3.0 \cdot \Phi^{-1}(99\%) = 14.98\end{aligned}$$

The stressed loss is then equal to:

$$\mathbb{S}(L(w)) = 10 \cdot 8.49 + 3 \cdot 14.98 = 129.53$$

We assume that the reverse stressed loss is equal to 129.53. Using Formula (14.4), we deduce that $\mathbb{RS}(\mathcal{F}_1) = 10.14$ and $\mathbb{RS}(\mathcal{F}_2) = 9.47$. The reverse stress scenario is very different than the stress scenario even if they give the same loss. In fact, we have $f(\mathbb{S}(\mathcal{F}_1), \mathbb{S}(\mathcal{F}_2)) = 0.8135 \cdot 10^{-6}$ and $f(\mathbb{RS}(\mathcal{F}_1), \mathbb{RS}(\mathcal{F}_2)) = 4.4935 \cdot 10^{-6}$, meaning that the occurrence probability of the reverse stress scenario is more than five times higher than the occurrence probability of the stress scenario.

The general case In the general case, we use a copula function $\mathbf{C}$ in order to describe the joint distribution of the risk factors. We have:

$$\ln f(\mathcal{F}_1, \ldots, \mathcal{F}_m) = \ln c(\mathbf{F}_1(\mathcal{F}_1), \ldots, \mathbf{F}_m(\mathcal{F}_m)) + \sum_{j=1}^{m} \ln f_j(\mathcal{F}_j)$$

where $c(u_1, \ldots, u_m)$ is the copula density, $\mathbf{F}_j$ is the cdf of $\mathcal{F}_j$ and f_j is the pdf of $\mathcal{F}_j$. Finally, we obtain a non-linear optimization problem subject to a non-linear constraint.

14.2.4.2 Practical solutions

There are very few articles on reverse stress testing, and a lack of statistical methods. However, we can cite Grundke (2011), Kopeliovich *et al.* (2015), Glasserman *et al.* (2015) and, Grundke and Pliszka (2018). In these research papers, the optimization problem is generally approximated. For instance, Kopeliovich *et al.* (2015) and Grundke and Pliszka (2018) consider the PCA method to reduce the problem dimension. Glasserman *et al.* (2015) propose to use the method of empirical likelihood in order to evaluate the probability of a reverse stress test.

From a practical point of view, banks generally use a fewer number of risk factors. This helps to reduce the problem dimension. They can also consider a Gaussian approximation. In fact, the main difficulty lies in the equality constraint. This is why they generally consider the following optimization problem:

$$\begin{aligned}(\mathbb{RS}(\mathcal{F}_1),\ldots,\mathbb{RS}(\mathcal{F}_m)) &= \arg\max \ln f(\mathcal{F}_1,\ldots,\mathcal{F}_m)\\ &\text{s.t.}\quad \ell(\mathbb{S}(\mathcal{F}_1),\ldots,\mathbb{S}(\mathcal{F}_m);w) \geq \mathbb{S}(L(w))\end{aligned}$$

In this case, they can use the Monte Carlo simulation method to estimate the reverse stress scenario.

14.3 Exercises

14.3.1 Construction of a stress scenario with the GEV distribution

1. We note a_n and b_n the normalization constants and $\mathbf{G}$ the limit distribution of the Fisher-Tippet theorem.

   (a) Find the limit distribution $\mathbf{G}$ when $X \sim \mathcal{E}(\lambda)$, $a_n = \lambda^{-1}$ and $b_n = \lambda^{-1}\ln n$.

   (b) Same question when $X \sim \mathcal{U}_{[0,1]}$, $a_n = n^{-1}$ and $b_n = 1 - n^{-1}$.

   (c) Same question when X is a Pareto distribution $\mathcal{P}(\alpha,\theta)$:

   $$\mathbf{F}(x) = 1 - \left(\frac{\theta+x}{\theta}\right)^{-\alpha}$$

   and the normalization constants are $a_n = \theta\alpha^{-1}n^{1/\alpha}$ and $b_n = \theta n^{1/\alpha} - \theta$.

2. We denote by $\mathbf{G}$ the GEV probability distribution:

   $$\mathbf{G}(x) = \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

   What is the interest of this probability distribution? Write the log-likelihood function associated to the sample $\{x_1,\ldots,x_T\}$.

3. Show that for $\xi \to 0$, the distribution $\mathbf{G}$ tends toward the Gumbel distribution:

   $$\mathbf{\Lambda}(x) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right)$$

4. We consider the minimum value of daily returns of a portfolio for a period of n trading days. We then estimate the GEV parameters associated to the sample of the opposite of the minimum values. We assume that ξ is equal to 1.

   (a) Show that we can approximate the portfolio loss (in %) associated to the return period $\mathcal{T}$ with the following expression:

   $$R(\mathcal{T}) \simeq -\left(\hat{\mu} + \left(\frac{\mathcal{T}}{n} - 1\right)\hat{\sigma}\right)$$

   where $\hat{\mu}$ and $\hat{\sigma}$ are the ML estimates of the GEV parameters.

   (b) We set n equal to 21 trading days. We obtain the following results for two portfolios:

Portfolio	$\hat{\mu}$	$\hat{\sigma}$	ξ
#1	1%	3%	1
#2	10%	2%	1

   Calculate the stress scenario for each portfolio when the return period is equal to one year. Comment on these results.

14.3.2 Conditional expectation and linearity

We consider the bivariate Gaussian random vector (X, Y):

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho_{x,y}\sigma_x\sigma_y \\ \rho_{x,y}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\right)$$

1. Using the conditional distribution theorem, show that:

   $$Y = \beta_0 + \beta X + \sigma U$$

   where $U \sim \mathcal{N}(0, 1)$. Give the expressions of β_0, β and σ.

2. Deduce the conditional expectation function $m(x)$:

   $$m(x) = \mathbb{E}[Y \mid X = x]$$

3. Let $\left(\tilde{X}, \tilde{Y}\right)$ be the log-normal random vector such that $\tilde{X} = \exp(X)$ and $\tilde{Y} = \exp(Y)$. Find the conditional expectation function $\tilde{m}(x)$:

   $$\tilde{m}(x) = \mathbb{E}\left[\tilde{Y} \mid \tilde{X} = x\right]$$

4. Comment on these results.

14.3.3 Conditional quantile and linearity

Let X and Y be a $n \times 1$ random vector and a random variable. We assume that (X, Y) is Gaussian:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{x,x} & \Sigma_{x,y} \\ \Sigma_{y,x} & \Sigma_{y,y} \end{pmatrix}\right)$$

We note $\mathbf{F}_x(x)$ and $\mathbf{F}_y(x)$ the marginal distributions, and $\mathbf{F}(x, y)$ the joint distribution.

1. Calculate the conditional distribution $\mathbf{F}(y \mid X = x)$ of the random variable $Y(x) = Y \mid X = x$. Deduce the conditional quantile defined by:

$$q_\alpha(x) = \inf\{q : \Pr(Y(x) \leq q) \geq \alpha\}$$

2. Show that:

$$q_\alpha(x) = \beta_0(\alpha) + \beta^\top x$$

where $\beta_0(\alpha)$ is a function that depends on the confidence level α.

3. Compare $q_\alpha(x)$ with the conditional expectation $m(x)$. Deduce the main difference between linear regression and quantile regression.

4. We consider an exponential default time $\tau \sim \mathcal{E}(\lambda)$ that depends on the risk factors X. Moreover, we assume that X is Gaussian $\mathcal{N}(\mu_x, \Sigma_{x,x})$ and the dependence between the default time τ and the risk factors X is a Normal copula. Find the conditional quantile function $q_\alpha^\tau(x)$ of the random variable $\tau(x) = \tau \mid X = x$.

5. We now consider the probability of default PD associated to the default time $\tau \sim \mathcal{E}(\lambda)$. Calculate the conditional quantile function $q_\alpha^{\mathrm{PD}}(x)$ of the random variable $\mathrm{PD}(x) = \mathrm{PD} \mid X = x$.

6. We consider the single factor case where $X \sim \mathcal{N}\left(\mu_x, \sigma_x^2\right)$ and we assume that the parameter of the Normal copula between τ and X is equal to ρ. Show that:

$$q_\alpha^{\mathrm{PD}}(x) = \Phi\left(\Phi^{-1}(\alpha)\sqrt{1-\rho^2} + \rho\frac{(x-\mu_x)}{\sigma_x}\right)$$

7. Comment on these results and propose a quantile regression model to stress the probability of default.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 15

Credit Scoring Models

Credit scoring refers to statistical models to measure the creditworthiness of a person or a company. They have progressively replaced judgemental systems and are now widely used by financial and banking institutions that check the credit rating and capacity of the borrower before to approve a loan. Therefore, credit scoring is at the heart of the decision-making system for granting credit. This is particularly true for consumer credit (mortgage, credit card, personal loan, etc.). Credit scoring models are also used for commercial firms, but their final outputs are generally not sufficient for making a decision. For instance, they can be completed with the knowledge of the relationship manager on the company.

Credit scoring first emerged in the United States. For instance, one of the oldest credit scores is the FICO score that was introduced in 1989 by Fair Isaac Corporation. The FICO score is based on consumer credit files of consumer credit reporting agencies such as Experian, Equifax and TransUnion. It remains today the best-known and most-used external scoring system in the world. In thirty years, credit scoring models have evolved considerably, and financial institutions have generally built their own internal credit scoring system. In particular, the development of credit scoring techniques has speeded up in the 2000s with the introduction of the IRB formula in the Basel II Accord. For instance, they are now used for estimating the probability of default or the loss given default, while validation and backtesting procedures are better defined. The estimation of credit scores has also benefitted from the massive development of marketing scores, big data and machine learning.

15.1 The method of scoring

15.1.1 The emergence of credit scoring

15.1.1.1 Judgmental credit systems versus credit scoring systems

The underlying idea of credit valuation is to use the experience in order to approve or deny the credit of a (new) customer. In the case of judgmental credit analysis, the decision is made by a credit analyst or the relationship manager, and is based on the character, the capacity and the capital of the borrower. Past experience of the credit analyst is then fundamental, and two credit analysts may give two different answers. Moreover, it takes many years to build a track record, because it is not an industrial process. Indeed, the credit analyst can analyze only a limited number of requests per week. Because of the high costs, financial institutions have sought to automate credit decisions.

In 1941, Durand presented a statistical analysis of credit valuation. He showed that credit analysts uses similar factors, and proposed a credit rating formula based on nine factors: (1) age, (2) sex, (3) stability of residence, (4) occupation, (5) industry, (6) stability of employment, (7) bank account, (8) real estate and (9) life insurance. The score is additive and can take values between 0 and 3.46. For instance, 0.40 is added to the score if the

applicant is a woman, 0.30 if the applicant is 50 years old or more, etc. Durand's formula is the first credit scoring model that has been published. Such credit scoring models become more and more popular in financial institutions in the 1950s and 1960s, but the real turning point is the development of the credit card business in the 1970s (Thomas, 2000). From an industrial point of view, a credit scoring system has two main advantages compared to a judgmental credit system:

1. it is cost efficient, and can treat a huge number of applicants;
2. decision-making process is rapid and consistent across customers.

Generally, financial institutions also consider that credit scoring systems are more efficient and successful than judgmental credit systems. However, comparing track records is always a difficult exercise since it depends on many factors. Some credit analysts may have a very good track record, while the live performance of some statistical credit models may be worse than their backtest performance. Nevertheless, the case of credit cards has demonstrated that credit scoring models are far better than judgmental credit systems. The main reason is the large amount of data that can be analyzed by a statistical model. While experience is essential for a credit analyst, the efficiency of credit scoring depends on the quality and amount of data.

15.1.1.2 Scoring models for corporate bankruptcy

These models appear with the research of Tamari (1966), who proposed to combine several financial ratios for assessing the financial health of corporate firms. Nevertheless, the weight of each ratio was assumed to be fixed and has been arbitrary calibrated. The empirical work of Beaver (1966) was more interesting since he estimated the univariate statistical relationship between financial ratios and the failure. However, the seminal paper for the evaluation of creditworthiness is the publication of Altman (1968). Using a small dataset and the statistical method of discriminant analysis, he introduced the concept of z-score for predicting bankruptcy of commercial firms. The score was equal to:

$$Z = 1.2 \cdot X_1 + 1.4 \cdot X_2 + 3.3 \cdot X_3 + 0.6 \cdot X_4 + 1.0 \cdot X_5$$

where the variables X_j represent the following financial ratios:

X_j	Ratio
X_1	Working capital / Total assets
X_2	Retained earnings / Total assets
X_3	Earnings before interest and tax / Total assets
X_4	Market value of equity / Total liabilities
X_5	Sales / Total assets

If we note Z_i the score of the firm i, we can calculate the normalized score $Z_i^\star = (Z_i - m_z)/\sigma_z$ where m_z and σ_z are the mean and standard deviation of the observed scores. $Z_i^\star$ can then be compared to the quantiles of the Gaussian distribution or the empirical distribution. A low value of $Z_i^\star$ (for instance $Z_i^\star < 2.5$) indicates that the firm has a high probability of default. Today, the technique of z-score, which consists of normalizing a score, is very popular and may be found in many fields of economics, finance, marketing and statistics.

15.1.1.3 New developments

Since the publication of Durand (1941) and Altman (1968), the research on credit scoring can be split into three main categories:

- The first category concerns the default of corporate firms. It appears that the choice of financial ratios and relevant metrics as explanatory variables are more important than the model itself (Hand, 2006). Other factors such as the business cycle, economic conditions or market prices (Hillegeist *et al.*, 2004) may be taken into account. Moreover, the one-size-fits-all approach is not appropriate and credit scoring models are different for stock-listed companies, medium-sized companies, financial companies or industrial companies (Altman *et al.*, 2010).

- The second category focuses on consumer credit and retail debt management (credit cards, mortgages, etc.). Sample sizes are larger than for corporate credit (Thomas, 2000) and may justify the use of more sophisticated techniques that include the behavior of the customer (Thomas *et al.*, 2017).

- The third research direction concerns statistical methods. Besides discriminant analysis, new approaches have been proposed, in particular logit or probit models (Ohlson, 1980; Lennox, 1999) and survival models (Shumway, 2001). Moreover, with the availability of more personal data, machine learning techniques such as neural networks (West, 2000) are also used and tested in credit scoring and are not reserved for only marketing scores.

15.1.2 Variable selection

15.1.2.1 Choice of the risk factors

Variables used to determine the creditworthiness of a borrower are generally based on 5 risk factor categories, also called the five Cs:

1. **Capacity** measures the applicant's ability to meet the loan payments. For example, lenders may look at the debt-to-income or the job stability of the applicant. In the case of corporate firms, the cash flow dynamics is a key element.

2. **Capital** is the size of assets that are held by the borrower. In the case of consumer credit, it corresponds to the net wealth of the borrower. For a corporate firm, it can be machinery, equipment, buildings, investment portfolio, etc.

3. **Character** measures the willingness to repay the loan. For example, the lender can investigate the payment history of the applicant. If the applicant has children, the applicant may have more incentive than if he/she is single.

4. **Collateral** concerns additional forms of security that the borrower can provide to the lender. This item is particularly important in the case of corporate credit.

5. **Conditions** refer to the characteristics of the loan and the economic conditions that might affect the borrower. For example, the score is generally a decreasing function of the maturity and the interests paid by the borrower. For corporate firms, some sectors are more dependent on the economic cycle than others.

In Table 15.1, we report some variables that are used when building a consumer credit score. This type of score is generally used by banks, since they may include information that is related to the banking relationship.

Scores are developed by banks and financial institutions, but they can also be developed by consultancy companies. This is the case of the FICO® scores, which are the most widely

TABLE 15.1: An example of risk factors for consumer credit

Character	Age of applicant
	Marital status
	Number of children
	Educational background
	Time with bank
	Time at present address
Capacity	Annual income
	Current living expenses
	Current debts
	Time with employer
Capital	Purpose of the loan
	Home status
	Saving account
Condition	Maturity of the loan
	Paid interests

used credit scoring systems in the world[1]. They are based on 5 main categories: payment history (35%), amount of debt (30%), length of credit history (15%), new credit (10%) and credit mix (10%).They generally range from 300 to 850, while the average score of US consumers is 695. These scores are generally classified as follows: exceptional (800+), very good (740-799), good (670-739), fair (580-669) and poor (580−).

Corporate credit scoring systems use financial ratios:

1. **Profitability**: gross profit margin, operating profit margin, return-on-equity (ROE), etc.

2. **Solvency**: debt-to-assets ratio, debt-to-equity ratio, interest coverage ratio, etc.

3. **Leverage**: liabilities-to-assets ratio (financial leverage ratio), long-term debt/assets, etc.

4. **Liquidity**: current assets/current liabilities (current ratio), quick assets/current liabilities (quick or cash ratio), total net working capital, assets with maturities of less than one year, etc.

Liquidity and solvency ratios measure the company's ability to satisfy its short-term and long-term obligations, while profitability ratios measure its ability to generate profits from its resources. High profitability, high solvency and high liquidity reduces the probability of default, but a high leverage increases the credit risk of the company. The score may also include non-financial variables: firm age[2], size (number of employees), quality of accounting information, management quality, etc. For instance, we generally consider that large firms default less often than small firms. Like retail scores, corporate scores are built by banks but also by consulting firms and credit agencies. For example, Moody's proposes the RiskCalc model (Falkenstein *et al.*, 2000).

[1]The FICO scores are developed since 1989 by Fair Isaac Corporation, which is a Californian-based firm. There are more than 20 scores that are commonly used for auto lending, credit card decisioning, mortgage lending, etc. In the US, FICO scores are used in over 90% of lending decisions (source: https://www.myfico.com).

[2]Recent firms may be penalized.

15.1.2.2 Data preparation

Of course data quality is essential for building a robust credit scoring. However, data preparation is not limited to check the data and remove outliers or fill missing values. Indeed, a '*one-size-fits-all*' approach is generally not appropriate, because a scoring model is generally more a decision tree system than a parsimonious econometric model. This is why credit scoring is work-intensive on data mining. Once the data is clean, we can begin the phase of exploratory data analysis, which encompasses three concurrent steps: variable transformation, slicing-and-dicing segmentation and potential interaction research. The first step consists in applying a non-linear transformation, for example by computing the logarithm, while the second and third steps are the creation of categorical/piecewise and interaction variables.

Piecewise and dummy variables Let b be a $p \times 1$ vector of bounds. We assume that b is sorted in ascending order. We note $b^{(1)} = (-\infty, b)$, $b^{(2)} = (b, +\infty)$, $b^{(3)} = (b_1, b)$ and:

$$b^{(4)} = (0, b_2 - b_1, b_3 - b_2, \ldots, b_p - b_{p-1}, 0)$$

It follows that $b^{(1)}$, $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$ are four vector of dimension $(p+1) \times 1$. From the vector b, we can then create $(p+1)$ piecewise variables which are defined by:

$$PW_j = \left(X - b_j^{(3)}\right) \cdot \mathbb{1}\left\{X > b_j^{(1)}\right\} \cdot \mathbb{1}\left\{X \leq b_j^{(2)}\right\} + b_j^{(4)} \cdot \mathbb{1}\left\{X > b_j^{(2)}\right\}$$

The underlying idea is to have an affine function if the original variable takes its values in the interval $]b_{j-1}, b_j]$. For instance, Figure 15.1 represents the fourth piecewise variables which are obtained from $b = (-0.5, 0, 1)$. In a similar way, we define dummy variables as follows:

$$D_j = \mathbb{1}\left\{X > b_j^{(1)}\right\} \cdot \mathbb{1}\left\{X \leq b_j^{(2)}\right\}$$

In this case, D_j takes a value of 1 if $X \in]b_{j-1}, b_j]$. Using $b = (-0.5, 0, 1)$, we obtain Figure 15.2.

Optimal slicing An important point is the choice of the bound $b = (b_1, b_2, \ldots, b_K)$. It is obvious that the optimal values depend on the response variable Y. For that, we introduce the contingency table of the random vector (Y, X), which corresponds to a table of counts with p rows and q columns:

Y/X	$X \in \mathcal{I}_1^{(X)}$	$\cdots$	$X \in \mathcal{I}_j^{(X)}$	$\cdots$	$X \in \mathcal{I}_q^{(X)}$
$Y \in \mathcal{I}_1^{(Y)}$	$n_{1,1}$		$n_{1,j}$		$n_{1,q}$
$\vdots$					
$Y \in \mathcal{I}_i^{(Y)}$	$n_{i,1}$		$n_{i,j}$		$n_{i,q}$
$\vdots$					
$Y \in \mathcal{I}_p^{(Y)}$	$n_{p,1}$		$n_{p,j}$		$n_{p,q}$

where $n_{i,j}$ is the number of observations such that $Y \in \mathcal{I}_i^{(Y)}$ and $X \in \mathcal{I}_j^{(X)}$. We assume that the set are disjoints: $\mathcal{I}_{j_1}^{(X)} \cap \mathcal{I}_{j_2}^{(X)} = \emptyset$ for $j_1 \neq j_2$ and $\mathcal{I}_{i_1}^{(Y)} \cap \mathcal{I}_{i_2}^{(Y)} = \emptyset$ for $i_1 \neq i_2$. We introduce the following notations:

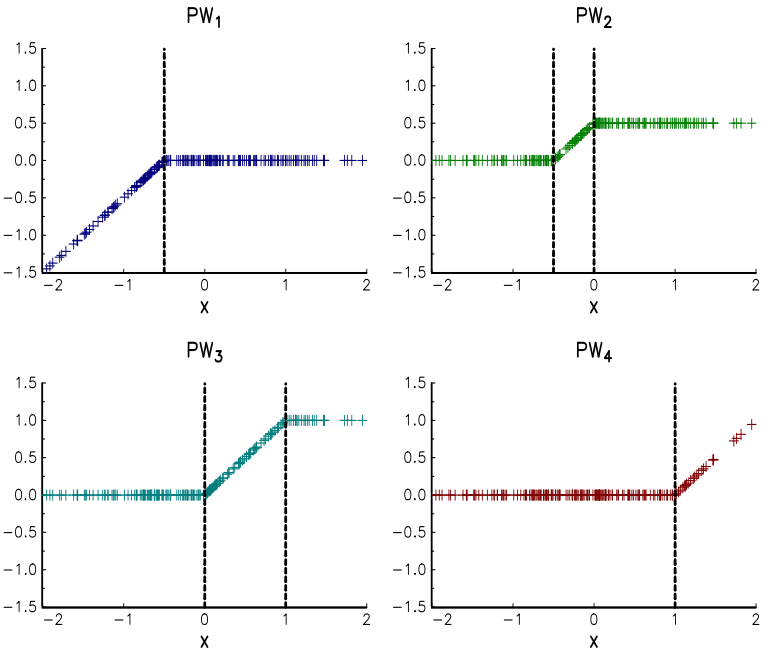


FIGURE 15.1: Piecewise variables

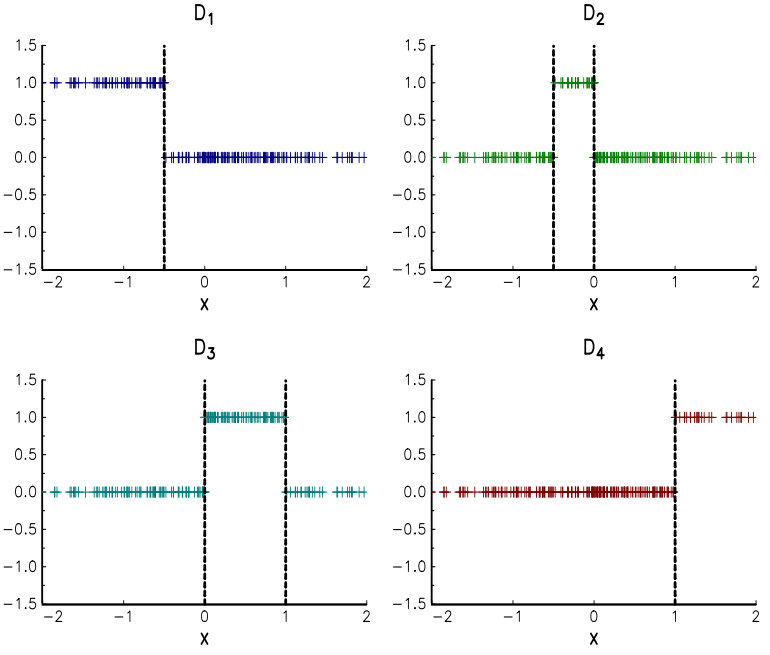


FIGURE 15.2: Dummy variables

- $n_{i,\cdot} = \sum_{j=1}^{q} n_{i,j}$ is the number of observations such that $Y \in \mathcal{I}_i^{(Y)}$;
- $n_{\cdot,j} = \sum_{i=1}^{p} n_{i,j}$ is the number of observations such that $X \in \mathcal{I}_j^{(X)}$;
- $n = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{i,j}$ is the total number of observations[3].

If we assume that X and Y are independent (null hypothesis $\mathcal{H}_0$), the expected number of observations such that $Y \in \mathcal{I}_i^{(Y)}$ and $X \in \mathcal{I}_j^{(X)}$ must be equal to:

$$\bar{n}_{i,j} = \frac{n_{i,\cdot} \times n_{\cdot,j}}{n}$$

Under $\mathcal{H}_0$, we can prove that the Pearson's statistic χ has a chi-squared limit distribution:

$$\chi = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(n_{i,j} - \bar{n}_{i,j})^2}{\bar{n}_{i,j}} \sim \chi^2(\nu)$$

where $\nu = (p-1)(q-1)$. If we apply the Pearson's chi-squared statistic to the previous scoring problem, the contingency table becomes:

X	$X \leq b_1$	$b_1 < X \leq b_2$	$\cdots$	$b_{p-1} < X \leq b_p$	$X > b_p$
$Y = 0$	$n_{0,1}$	$n_{0,2}$	$\cdots$	$n_{0,p}$	$n_{0,p+1}$
$Y = 1$	$n_{1,1}$	$n_{1,2}$	$\cdots$	$n_{1,p}$	$n_{1,p+1}$

We assume here that Y is a binary random variable: $Y = 0$ indicates a good credit and $Y = 1$ corresponds to a bad credit. We note $\chi(b)$ the value of the chi-squared statistic that depends on the slicing vector b:

$$\chi(b) = \sum_{i=0}^{1} \sum_{j=1}^{p+1} \frac{(n_{i,j} - \bar{n}_{i,j})^2}{\bar{n}_{i,j}}$$

The optimal value of b is defined by:

$$b^\star = \arg\max \chi(b) \tag{15.1}$$

Indeed, if X and Y are independent, we have $\chi(b) = 0$. In this case, the variable X does not help to predict the variable Y. Maximizing the chi-squared statistic is equivalent to finding the slicing setup that deviates the most from the independent case.

In order to solve the maximization problem (15.1), we may use the dynamic programming principle, whose objective function is to solve this problem:

$$\{c^\star(k)\}_{k=1}^{K-1} = \arg\max \sum_{k=1}^{K-1} f(k, s(k), c(k)) + f(K, s(K)) \tag{15.2}$$

$$\text{s.t.} \quad \begin{cases} s(k+1) = g(k, s(k), c(k)) \\ s(k) \in \mathcal{S}(k) \\ c(k) \in \mathcal{C}(k) \\ s(1) = s \end{cases}$$

The underlying idea is to initialize the algorithm[4] with a predetermined slice $\{b_1, b_2, \ldots, b_p\}$, to aggregate the knots in order to find the optimal slice $\{b_1^\star, b_2^\star, \ldots, b_{p^\star}^\star\}$ for a given value

[3] We also have:

$$n = \sum_{i=1}^{p} n_{i,\cdot} = \sum_{j=1}^{q} n_{\cdot,j}$$

[4] The algorithm is described on page 1049.

of $p^\star$. For that, we note $n_{i,j}(b_{j_1}, b_{j_2}) = \#(Y = i, b_{j_1} < X \leq b_{j_2})$. The chi-squared marginal contribution is defined by:

$$\chi(b_{j_1}, b_{j_2}) = \sum_{i=0}^{1} \frac{(n_{i,j}(b_{j_1}, b_{j_2}) - \bar{n}_{i,j}(b_{j_1}, b_{j_2}))^2}{\bar{n}_{i,j}(b_{j_1}, b_{j_2})} \qquad \text{for } j_1 < j_2$$

$\chi(b_{j_1}, b_{j_2})$ can be viewed as the Pearson's statistic when we only consider the observations such that $b_{j_1} < X \leq b_{j_2}$. The gain function is equal to:

$$f(k, s(k), c(k)) = \begin{cases} -\infty & \text{if } c(k) \leq s(k) \\ \chi\left(b_{s(k)+1}, b_{c(k)}\right) & \text{otherwise} \end{cases}$$

If $k = 1$, we have:

$$f(1, s(1), c(1)) = \begin{cases} -\infty & \text{if } c(1) \leq s(1) \\ \chi\left(b_0, b_{s(1)}\right) + \chi\left(b_{s(1)+1}, b_{c(1)}\right) & \text{otherwise} \end{cases}$$

The transfer function is defined as follows:

$$s(k+1) = g(k, s(k), c(k)) = c(k)$$

The state variable $s(k)$ and the control variable $c(k)$ take their values in the set $\{1, 2, \ldots, p\}$. The number K of iterations is exactly equal to $p^\star$ and we have:

$$f(K, s_j) = \chi\left(b_{s_j+1}, b_p\right)$$

In the case where $p^\star = 1$, the dynamic programming algorithm reduces to the brute force algorithm:

$$j^\star = \underset{j \in \{b_1, b_2, \ldots, b_p\}}{\arg\max}\ \chi(-\infty, b_j) + \chi(b_j, \infty)$$

In this case, the optimal slice is composed of two classes: $X \leq b_{j^\star}$ and $X > b_{j^\star}$.

Example 163 *We consider 40 observations of the random vector (Y, X). Below, we indicate the values taken by X when $Y = 0$ and $Y = 1$:*

- *$Y = 0$: -2.0, -1.1, -1.0, -0.7, -0.5, -0.5, -0.4, -0.3, -0.2, -0.2, 0.0, 0.7, 0.8, 0.9, 1.0, 1.4, 1.9, 2.8, 3.2, 3.7.*
- *$Y = 1$: -5.2, -4.3, -3.6, -2.7, -1.8, -1.5, -1.2, -1.0, -0.8, -0.1, 0.0, 0.2, 0.2, 0.3, 0.5, 0.5, 0.5, 0.7, 0.8, 1.9.*

If we consider the following grid $b = (-5, -4, -3, -2, -1, 0, 1, 2, 3)$, we obtain the following contingency table:

X	$\mathcal{I}_1^{(X)}$	$\mathcal{I}_2^{(X)}$	$\mathcal{I}_3^{(X)}$	$\mathcal{I}_4^{(X)}$	$\mathcal{I}_5^{(X)}$	$\mathcal{I}_6^{(X)}$	$\mathcal{I}_7^{(X)}$	$\mathcal{I}_8^{(X)}$	$\mathcal{I}_9^{(X)}$	$\mathcal{I}_{10}^{(X)}$
$Y = 0$	0	0	0	0	2	8	4	3	1	2
$Y = 1$	1	1	1	1	3	3	9	1	0	0

where $\mathcal{I}_1^{(X)} = \{X \leq -5\}$, $\mathcal{I}_2^{(X)} = \{-5 < X \leq -4\}$, $\ldots$, $\mathcal{I}_{10}^{(X)} = \{X > -4\}$. If we would like to slice X into two classes, we use the brute force algorithm. If we group the intervals $\left\{\mathcal{I}_2^{(X)}, \ldots, \mathcal{I}_{10}^{(X)}\right\}$, the contingency table becomes:

X	$X \leq -5$	$X > -5$	$n_{i,\cdot}$
$Y = 0$	0	20	20
$Y = 1$	1	19	20
$n_{\cdot,j}$	1	39	$n = 40$

We deduce that:

$$\begin{aligned}\chi &= \frac{(0-0.5)^2}{0.5}+\frac{(20-19.5)^2}{19.5}+\frac{(1-0.5)^2}{0.5}+\frac{(19-19.5)^2}{19.5}\\ &= 1.02564\end{aligned}$$

If we now consider the two groups $\left\{\mathcal{I}_1^{(X)},\mathcal{I}_2^{(X)}\right\}$ and $\left\{\mathcal{I}_3^{(X)},\ldots,\mathcal{I}_{10}^{(X)}\right\}$, we obtain the following contingency table:

X	$X\leq -4$	$X>-4$	$n_{i,\cdot}$
$Y=0$	0	20	20
$Y=1$	2	18	20
$n_{\cdot,j}$	2	38	$n=40$

The associated Pearson's chi-squared statistic is then equal to:

$$\begin{aligned}\chi &= \frac{(0-1.0)^2}{1.0}+\frac{(20-19.0)^2}{19.0}+\frac{(2-1.0)^2}{1.0}+\frac{(18-19.0)^2}{19.0}\\ &= 2.10526\end{aligned}$$

We can proceed in the same way with the other values of b and we obtain the following values of χ when the cut-off is b_j:

X	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
χ	1.03	2.11	3.24	4.44	3.58	0.00	4.33	3.24	2.11

Since the maximum is reached for b_4 ($\chi=4.44$), the optimal slicing is the following:

X	$X\leq -2$	$X>-2$
$Y=0$	0	20
$Y=1$	4	16

If we prefer to slice X into three classes, the dynamic programming algorithm finds that the optimal cut-offs are $b^\star=(-2,1)$. In the case of four classes, the optimal slicing is:

X	$X\leq -1$	$-1<X\leq 0$	$0<X\leq 1$	$X>1$
$Y=0$	2	8	4	6
$Y=1$	7	3	9	1

and the optimal value $\chi^\star$ is equal to 10.545. In order to understand how does the dynamic programming algorithm work, we report the $\boldsymbol{J}$ and $\boldsymbol{C}$ matrices in Table 15.2. We notice that the optimal value is $\mathcal{J}(1,s^\star(1))=10.545$ where $s^\star(1)$ is the 5^{th} state. Moreover, the optimal controls are $c^\star(1)=6$ and $c^\star(2)=7$ implying that $s^\star(2)=c^\star(1)$ and $s^\star(3)=c^\star(2)$ are the 6^{th} and 7^{th} states. This is why the optimal cut-offs are $b^\star=(-1,0,1)$, that is the 5^{th}, 6^{th} and 7^{th} elements of the initial vector b.

Remark 178 *We notice that the optimal slice $b^\star$ depends on the initial grid b. This implies that another grid b will not necessarily give the same optimal slice. For instance, we have used a step of 1 in the previous example. If we use a step of 0.2, we obtain the optimal solution $b^\star=(-0.8,-0.2,0.6)$. We have reported the corresponding slicing in Figure 15.3. In this case, the Pearson's chi-squared statistic is equal to 18.444, which is better than the value 10.545 obtained previously. This is why it is better to use a small step than a large step. The risk is that the dynamic programming algorithm produces some classes with a low number of observations. To prevent this possible overfitting, we can impose that $\chi(b_{j_1},b_{j_2})=-\infty$ when the number of observations is below a given threshold ($\#(b_{j_1}<X\leq b_{j_2})\leq n_{\min}$). This ensures that each optimized class has at least $n_{\min}$ observations.*

TABLE 15.2: Dynamic programming matrices $\boldsymbol{J}$ and $\boldsymbol{C}$

state	$\mathcal{J}(1, s(1))$	$\mathcal{J}(1, s(2))$	$\mathcal{J}(1, s(3))$	$c(1)$	$c(2)$
1	7.6059	4.0714	0.0256	4	7
2	7.7167	3.8618	0.1053	6	7
3	9.0239	3.7048	0.2432	6	7
4	10.4945	3.6059	0.4444	6	7
5	10.5450	3.5714	0.8065	6	7
6	5.9231	5.4945	0.0000	7	7
7	4.7576	4.0000	3.5714	8	8
8	$-\infty$	3.0000	3.0000	1	9
9	$-\infty$	$-\infty$	2.0000	1	1

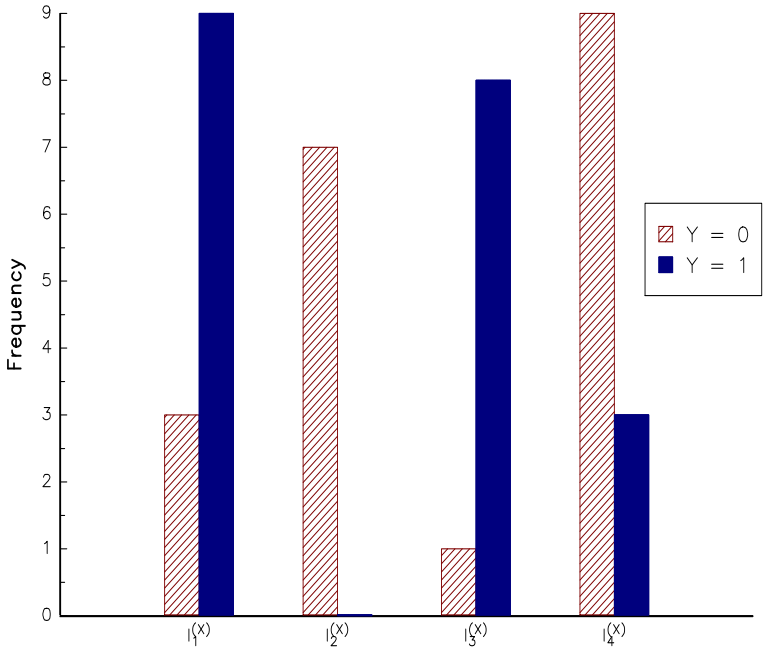


FIGURE 15.3: Optimal slicing with four classes

15.1.2.3 Variable selection

In practice, one may have many candidate variables $X = (X_1, \ldots, X_m)$ for explaining the variable Y. The variable selection problem consists in finding the best set of optimal variables. Let us assume the following statistical model:

$$Y = f(X) + u$$

where $u \sim \mathcal{N}\left(0, \sigma^2\right)$. We denote the prediction by $\hat{Y} = \hat{f}\left(X\right)$. By assuming the standard statistical hypotheses, we obtain:

$$\begin{aligned}
\mathbb{E}\left[\left(Y - \hat{Y}\right)^2\right] &= \mathbb{E}\left[\left(f\left(X\right) + u - \hat{f}\left(X\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(f\left(X\right) - \hat{f}\left(X\right)\right)^2\right] + \mathbb{E}\left[u^2\right] \\
&= \left(\mathbb{E}\left[\hat{f}\left(X\right)\right] - f\left(X\right)\right)^2 + \mathbb{E}\left[\left(\hat{f}\left(X\right) - \mathbb{E}\left[\hat{f}\left(X\right)\right]\right)^2\right] + \sigma^2 \\
&= \text{Bias}^2 + \text{Variance} + \text{Error}
\end{aligned}$$

Hastie *et al.* (2009) decompose the mean squared error of $\hat{f}\left(X\right)$ into three terms: a bias component, a variance component and an irreducible error. This bias-variance decomposition depends on the complexity of the model. When the model complexity is low (*i.e.* when there is a low number of regressors), the estimator $\hat{f}\left(X\right)$ generally presents a high bias but a low variance. When the model complexity is high (*i.e.* when there is a high number of regressors), the estimator $\hat{f}\left(X\right)$ generally presents a low bias but a high variance. The underlying idea of variable selection is then to optimize the bias-variance trade-off.

Best subset selection A first approach is to find the best subset of size k for $k \in \{1, \ldots, m\}$ that gives the smallest residual sum of squares. It follows that the search is performed through 2^m possible subsets, meaning that we rapidly face a combinatorial explosion. Moreover, minimizing the residual sum of squares is equivalent to consider the largest subset $(1, \ldots, m)$. This is why we prefer to consider an information criterion that penalizes the degree of freedom of the model. For instance, the Akaike criterion is defined as follows:

$$\text{AIC}\left(\alpha\right) = -2\boldsymbol{\ell}_{(k)}\left(\hat{\theta}\right) + \alpha \cdot \text{df}_{(k)}^{(\text{model})}$$

where $\boldsymbol{\ell}_{(k)}\left(\hat{\theta}\right)$ and $\text{df}_{(k)}^{(\text{model})}$ are the log-likelihood and the degree of freedom of the k^{th} model[5]. Therefore, the best model corresponds to the model that minimizes the Akaike criterion. In practice, the penalization parameter is generally set to $\alpha = 2$. In the case of the previous model, we deduce that:

$$\text{AIC}\left(2\right) = n \ln\left(\frac{\text{RSS}\left(\hat{\theta}\right)}{n}\right) + 2\,\text{df}_{(k)}^{(\text{model})}$$

Stepwise approach Another way for selecting variables is to use sequential approaches: forward selection, backward selection and forward/backward combined selection. In the case of forward selection, we start with the intercept and include one variable by one variable. At each step, we select the model of dimension $k+1$ with the most significant F-value with respect to the previous optimal model of dimension k:

$$F = \frac{\text{RSS}\left(\hat{\theta}_{(k)}\right) - \text{RSS}\left(\hat{\theta}_{(k+1)}\right)}{\text{RSS}\left(\hat{\theta}_{(k+1)}\right) / \text{df}_{(k+1)}^{(\text{residual})}}$$

[5] $\text{df}_{(k)}^{(\text{model})}$ is a complexity measure of the model, and corresponds to the number of estimated parameters. It is sometimes called the '*model degree of freedom*' whereas the classical measure used in linear regression t-statistics $\text{df}_{(k)}^{(\text{residual})}$ is called the '*residual degree of freedom*'. We have the following relationship $\text{df}_{(k)}^{(\text{residual})} = n - \text{df}_{(k)}^{(\text{model})}$ where n is the number of observations.

We stop when no model produces a significant F-value at the 95% confidence level. In the case of backward selection, we start with all the variables and remove one variable by one variable. At each step, we select the model of dimension k with the smallest significant F-value with respect to the previous optimal model of dimension $k+1$. The forward/backward combined procedure consists in using a forward step followed by a backward step, and to iterate this loop until the convergence criterion is reached. The convergence criterion can be expressed as a maximum number of loops[6].

Lasso approach The lasso method consists in adding a L_1 penalty function to the optimization function in order to obtain a sparse parameter vector θ:

$$L_1(\theta) = \|\theta\|_1 = \sum_{k=1}^{K} |\theta_k|$$

For example, the lasso regression model is specified as follows (Tibshirani, 1996):

$$y_i = \sum_{k=1}^{K} \beta_k x_{i,k} + u_i \quad \text{s.t.} \quad \sum_{k=1}^{K} |\beta_k| \leq \tau$$

where τ is a scalar to control the sparsity. Using the notations introduced on page 604, we have:

$$\begin{aligned} \hat{\beta}(\tau) &= \arg\min (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \\ &\text{s.t.} \quad \|\beta\|_1 \leq \tau \end{aligned} \tag{15.3}$$

This problem is equivalent to the Lagrange optimization program $\hat{\beta}(\lambda) = \arg\min \mathcal{L}(\beta; \lambda)$ where[7]:

$$\begin{aligned} \mathcal{L}(\beta; \lambda) &= \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + \lambda \|\beta\|_1 \\ &\propto \frac{1}{2} \beta^\top \left(\mathbf{X}^\top \mathbf{X}\right) \beta - \beta^\top \left(\mathbf{X}^\top \mathbf{Y}\right) + \lambda \|\beta\|_1 \end{aligned}$$

The solution $\hat{\beta}(\lambda)$ can be found by solving the augmented QP program where $\beta = \beta^+ - \beta^-$ under the constraints $\beta^+ \geq \mathbf{0}$ and $\beta^- \geq \mathbf{0}$. We deduce that:

$$\begin{aligned} \|\beta\|_1 &= \sum_{k=1}^{K} \left|\beta_k^+ - \beta_k^-\right| \\ &= \sum_{k=1}^{K} \left|\beta_k^+\right| + \sum_{k=1}^{K} \left|\beta_k^-\right| \\ &= \mathbf{1}^\top \beta^+ + \mathbf{1}^\top \beta^- \end{aligned}$$

Since we have:

$$\beta = \begin{pmatrix} I_K & -I_K \end{pmatrix} \begin{pmatrix} \beta^+ \\ \beta^- \end{pmatrix}$$

the augmented QP program is specified as follows:

$$\begin{aligned} \hat{\theta} &= \arg\min \frac{1}{2} \theta^\top Q \theta - \theta^\top R \\ &\text{s.t.} \quad \theta \geq \mathbf{0} \end{aligned}$$

[6] The algorithm also stops when the variable to be added is the same as the last deleted variable.

[7] τ and λ are related by the relationship $\tau = \left\|\hat{\beta}(\lambda)\right\|_1$.

where $\theta = (\beta^+, \beta^-)$, $\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X} & -\mathbf{X} \end{pmatrix}$, $Q = \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$ and $R = \tilde{\mathbf{X}}^\top \mathbf{Y} - \lambda \cdot \mathbf{1}$. If we denote $A = \begin{pmatrix} I_K & -I_K \end{pmatrix}$, we obtain:

$$\hat{\beta}(\lambda) = A\hat{\theta}$$

Remark 179 *If we consider Problem (15.3), we can also solve it using another augmented QP program:*

$$\begin{aligned} \hat{\theta} &= \arg\min \frac{1}{2}\theta^\top Q\theta - \theta^\top R \\ &s.t. \quad \begin{cases} C\theta \geq D \\ \theta \geq \mathbf{0} \end{cases} \end{aligned}$$

where $Q = \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$, $R = \tilde{\mathbf{X}}^\top \mathbf{Y}$, $C = -\mathbf{1}^\top$ *and* $D = -\tau$. *We again have* $\hat{\beta}(\tau) = A\hat{\theta}$.

We have:

$$\begin{aligned} \mathrm{RSS}(\beta) &= (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \\ &= \left(\mathbf{Y} - \mathbf{X}\left(\hat{\beta}^{\mathrm{ols}} + \beta - \hat{\beta}^{\mathrm{ols}}\right)\right)^\top \left(\mathbf{Y} - \mathbf{X}\left(\hat{\beta}^{\mathrm{ols}} + \beta - \hat{\beta}^{\mathrm{ols}}\right)\right) \\ &= \left(\mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathrm{ols}}\right)^\top \left(\mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathrm{ols}}\right) + 2\left(\mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathrm{ols}}\right)^\top \mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) + \\ &\quad \left(\beta - \hat{\beta}^{\mathrm{ols}}\right)^\top \mathbf{X}^\top \mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \end{aligned}$$

We notice that:

$$\begin{aligned} (*) &= \left(\mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathrm{ols}}\right)^\top \mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \\ &= \left(\mathbf{Y}^\top - \left(\hat{\beta}^{\mathrm{ols}}\right)^\top \mathbf{X}^\top\right) \mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \\ &= \left(\mathbf{Y}^\top - \left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\right)^\top \mathbf{X}^\top\right) \mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \\ &= \left(\mathbf{Y}^\top\mathbf{X} - \left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\right)^\top \mathbf{X}^\top\mathbf{X}\right) \left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \\ &= \left(\mathbf{Y}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{X}\right) \left(\beta - \hat{\beta}^{\mathrm{ols}}\right) \\ &= 0 \end{aligned}$$

Finally, we obtain:

$$\mathrm{RSS}(\beta) = \mathrm{RSS}\left(\hat{\beta}^{\mathrm{ols}}\right) + \left(\beta - \hat{\beta}^{\mathrm{ols}}\right)^\top \mathbf{X}^\top\mathbf{X}\left(\beta - \hat{\beta}^{\mathrm{ols}}\right)$$

If we consider the equation $\mathrm{RSS}(\beta) = c$, we distinguish three cases:

1. if $c < \mathrm{RSS}\left(\hat{\beta}^{\mathrm{ols}}\right)$, there is no solution;
2. if $c = \mathrm{RSS}\left(\hat{\beta}^{\mathrm{ols}}\right)$, there is one solution $\beta^\star = \hat{\beta}^{\mathrm{ols}}$;
3. if $c > \mathrm{RSS}\left(\hat{\beta}^{\mathrm{ols}}\right)$, we have:

$$\left(\beta - \hat{\beta}^{\mathrm{ols}}\right)^\top A\left(\beta - \hat{\beta}^{\mathrm{ols}}\right) = 1$$

where:

$$A = \frac{\mathbf{X}^\top \mathbf{X}}{c - \text{RSS}\left(\hat{\beta}^{\text{ols}}\right)}$$

The solution $\beta^\star$ is an ellipsoid, whose center is $\hat{\beta}^{\text{ols}}$ and principal axes are the eigenvectors of the matrix A.

If we add the lasso constraint $\sum_{k=1}^{K} |\beta_k| \leq \tau$, the lasso estimator $\hat{\beta}(\tau)$ corresponds to the tangency between the diamond shaped region and the ellipsoid that corresponds to the possible maximum value of c. The diamond shape region due to the lasso constraint ensures that the lasso estimator is sparse:

$$\exists \eta > 0 : \forall \tau < \eta,\ \min\left(\hat{\beta}_1(\tau), \ldots, \hat{\beta}_K(\tau)\right) = 0$$

For example, the two-dimensional case is represented in Figure 15.4. We notice that $\hat{\beta}_1(\tau)$ is equal to zero if $\tau < \eta$. This sparsity property is central for understanding the variable selection procedure.

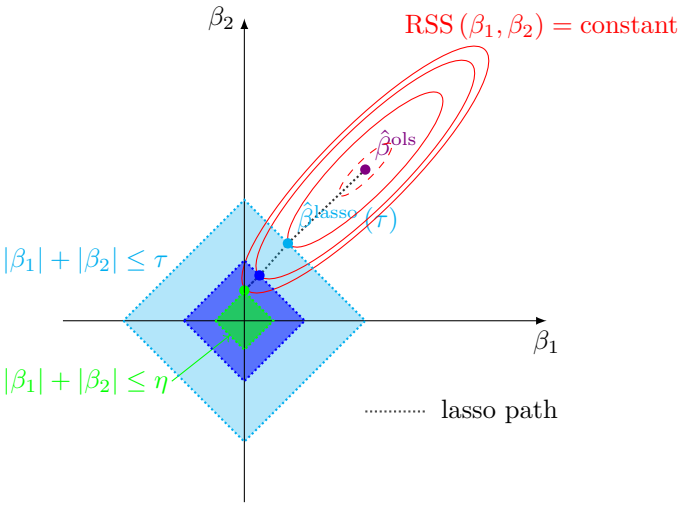


FIGURE 15.4: Interpretation of the lasso regression

Example 164 *Using the data given in Table 15.3, we consider the linear regression model:*

$$y_i = \beta_0' + \sum_{k=1}^{5} \beta_k' x_{i,k} + u_i \tag{15.4}$$

The objective is to determine the importance of each variable.

The lasso method can be used for ranking the variables. For that, we consider the following linear regression:

$$\tilde{y}_i = \sum_{k=1}^{5} \beta_k \tilde{x}_{i,k} + u_i$$

TABLE 15.3: Data of the lasso regression problem

i	y	x_1	x_2	x_3	x_4	x_5
1	3.1	2.8	4.3	0.3	2.2	3.5
2	24.9	5.9	3.6	3.2	0.7	6.4
3	27.3	6.0	9.6	7.6	9.5	0.9
4	25.4	8.4	5.4	1.8	1.0	7.1
5	46.1	5.2	7.6	8.3	0.6	4.5
6	45.7	6.0	7.0	9.6	0.6	0.6
7	47.4	6.1	1.0	8.5	9.6	8.6
8	−1.8	1.2	9.6	2.7	4.8	5.8
9	20.8	3.2	5.0	4.2	2.7	3.6
10	6.8	0.5	9.2	6.9	9.3	0.7
11	12.9	7.9	9.1	1.0	5.9	5.4
12	37.0	1.8	1.3	9.2	6.1	8.3
13	14.7	7.4	5.6	0.9	5.6	3.9
14	−3.2	2.3	6.6	0.0	3.6	6.4
15	44.3	7.7	2.2	6.5	1.3	0.7

where $\tilde{y}_i$ and $\tilde{x}_{i,k}$ are the standardized data[8]:

$$\frac{y_i - \bar{y}}{s_y} = \sum_{k=1}^{5} \beta_k \left(\frac{x_{i,k} - \bar{x}_k}{s_{x_k}} \right) + u_i \tag{15.5}$$

Linear regressions (15.4) and (15.5) are related by the following equation:

$$y_i = \left(\bar{y} - \sum_{k=1}^{5} \frac{s_y \beta_k}{s_{x_k}} \bar{x}_k \right) + \sum_{k=1}^{5} \frac{s_y \beta_k}{s_{x_k}} x_{i,k} + s_y u_i$$

We deduce that $\beta_0' = \bar{y} - \sum_{k=1}^{5} (s_y / s_{x_k}) \beta_k \bar{x}_k$ and $\beta_k' = (s_y / s_{x_k}) \beta_k$. When performing lasso regression, we always standardize the data in order to obtain comparable beta's. Otherwise, the penalty function $\|\beta\|_1$ does not make a lot of sense. In Table 15.4, we have estimated the lasso coefficients $\beta_k(\lambda)$ for different values of the shrinkage parameter λ. When $\lambda = 0$, we obtain the OLS estimate, and the lasso regression selects all the available variables. When $\lambda \to \infty$, the solution is $\hat{\beta}(\infty) = \mathbf{0}$, and the lasso regression selects no explanatory variables. In Table 15.4, we verify that the number of selected variables is a decreasing function of λ. For instance, the lasso regression selects respectively four and three variables when λ is equal to 0.9 and 2.5. It follows that the most important variable is the third one, followed by the first, second, fourth and fifth variables.

In Figure 15.5, we have reported the path of the lasso estimate $\hat{\beta}(\lambda)$ with respect to the scaling factor $\tau^\star \in [0, 1]$, which is defined as follows:

$$\tau^\star = \frac{\tau}{\tau_{\max}} = \frac{\left\| \hat{\beta}(\lambda) \right\|_1}{\left\| \hat{\beta}(0) \right\|_1}$$

[8] The notations $\bar{x}_k$ and s_{x_k} represent the mean and the standard deviation of the data $\left\{ x_{i,k}, i = 1, \ldots, n \right\}$.

$\tau^\star$ is equal to zero when $\lambda \to \infty$ (no selected variable) and one when $\lambda = 0$, which corresponds to the OLS case. From this path, we verify the lasso ordering:

$$x_3 \succ x_1 \succ x_2 \succ x_4 \succ x_5$$

TABLE 15.4: Results of the lasso regression

λ	0.0	0.9	2.5	5.5	7.5
$\hat{\beta}_1(\lambda)$	0.4586	0.4022	0.3163	0.1130	
$\hat{\beta}_2(\lambda)$	−0.1849	−0.2005	−0.1411		
$\hat{\beta}_3(\lambda)$	0.8336	0.7265	0.5953	0.3951	0.2462
$\hat{\beta}_4(\lambda)$	−0.1893	−0.1102			
$\hat{\beta}_5(\lambda)$	0.0931				
$\left\|\hat{\beta}(\lambda)\right\|_1$	1.7595	1.4395	1.0527	0.5081	0.2462
$\text{RSS}\left(\hat{\beta}(\lambda)\right)$	0.0118	0.0304	0.1180	0.4076	0.6306
R_c^2	0.9874	0.9674	0.8735	0.5633	0.3244
$\text{df}^{(\text{model})}$	5	4	3	2	1

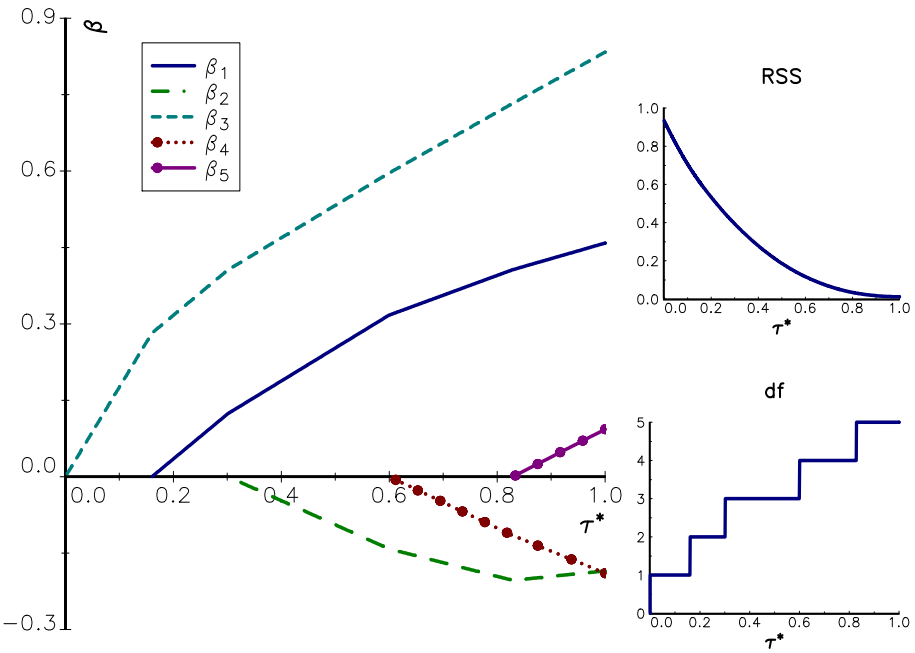


FIGURE 15.5: Variable selection with the lasso regression

15.1.3 Score modeling, validation and follow-up

15.1.3.1 Cross-validation approach

In order to avoid overfitting, we can also split the dataset into a training set and a validation set. The training set is used to estimate the model, for example the vector θ in

the case of a parametric model, while the validation set is used to compute the prediction error and the residual sum of squares. This approach can be generalized for model selection. In this case, the training set is used to fit the several models, while the validation set is used to select the right model (Hastie *et al.*, 2009). We generally distinguish two types of cross-validation.

1. In exhaustive cross-validation methods, learning and testing are based on all possible ways to divide the original sample into a training set and a validation set. For example, leave-p-out cross-validation (LpOCV) assumes that the validation set is composed of p observations, while the training set corresponds to the remaining observations. Since the number of training and validation sets is equal to C_p^n, this approach may be computationally intensive. In order to reduce the complexity, we can choose $p = 1$. This approach is called the leave-one-out cross-validation (LOOCV).

2. Non-exhaustive cross-validation methods split the original sample into training and validation sets. For instance, the k-fold approach randomly divides the dataset into k (almost) equally sized subsamples. At each iteration, one subsample is choosen as a validation set, while the $k-1$ remaining subsamples form the training set. This means that the model is fitted using all but the j^{th} group of data, and the j^{th} group of data is used for the test set. We repeat the procedure k times, in such a way that each subsample is tested exactly once. In the case of a linear regression, the k-fold cross validation error is generally computed as:

$$\mathcal{E}_{\text{cv}} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i\in\mathcal{G}_j}\left(y_i - x_i^\top \hat{\beta}\,(j)\right)^2$$

where $i \in \mathcal{G}_j$ denotes the observations of the j^{th} subsample and $\hat{\beta}\,(j)$ the estimate of β obtained by leaving out the j^{th} subsample. Even in simple cases, it cannot be guaranteed that the function $\mathcal{E}_{\text{cv}}$ has a unique minimum. The simple grid search approach is probably the best approach. The exhaustive leave-one-out cross validation (LOOCV) is a particular case when k is equal to the size of the dataset. Moreover, we can show that LOOCV is asymptotically equivalent to the AIC criterion (Stone, 1977).

In order to illustrate the principle of cross-validation, we consider the ridge estimator:

$$\hat{\beta} = \arg\min \frac{1}{2}\left(\mathbf{Y} - \mathbf{X}\beta\right)^\top \left(\mathbf{Y} - \mathbf{X}\beta\right) + \frac{\lambda}{2}\beta^\top\beta$$

where $\mathbf{Y}$ is a $n \times 1$ vector, $\mathbf{X}$ is a $n \times K$ matrix and β is a $K \times 1$ vector. The ridge model is then a regularized linear regression model with a L_2-norm penalty (Hoerl and Kennard, 1970). It follows that the expression of $\hat{\beta}$ is equal to:

$$\hat{\beta} = \left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1}\mathbf{X}^\top\mathbf{Y}$$

In the case of the leave-one-out cross validation, Allen (1971, 1974) showed that the function $\mathcal{E}_{\text{cv}}$ has an explicit expression known as the predicted residual error sum of squares (or PRESS) statistic:

$$\mathcal{P}\text{ress} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,-i}\right)^2$$

where $\hat{y}_{i,-i}$ is the estimate of y_i based on the ridge model when leaving out the i^{th} observation. Indeed, we have[9]:

$$\mathcal{P}\text{ress} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{u}_i^2}{(1-h_i)^2}$$

where $\hat{u}_i = y_i - x_i^\top \hat{\beta}$ and $h_i = x_i^\top \left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1} x_i$. With this formula, we don't need to estimate the n estimators $\hat{\beta}_{-i}$, where $\hat{\beta}_{-i}$ is the ridge estimator when leaving out the i^{th} observation.

TABLE 15.5: Data of the ridge regression problem

i	y	x_1	x_2	x_3	x_4	x_5
1	−23.0	−8.0	6.0	−12.7	9.5	−7.5
2	−21.0	−6.5	11.1	5.4	6.6	6.7
3	−5.0	−14.4	−13.3	−3.2	0.8	1.0
4	−39.6	−6.7	26.0	11.5	15.5	6.5
5	5.8	2.3	−7.1	−4.6	7.0	−0.6
6	13.6	2.0	−13.0	−13.3	−0.9	−8.6
7	14.0	10.7	−4.9	−23.1	2.5	19.0
8	−5.2	−8.5	1.0	4.2	−11.5	12.9
9	6.9	3.4	4.9	9.5	−12.8	11.0
10	−5.2	0.0	5.1	−14.3	−3.8	−10.0
11	0.0	1.0	4.0	14.1	−3.5	−23.6
12	3.0	2.4	1.6	−1.2	−4.8	−9.2
13	9.2	−0.1	−10.6	16.0	7.5	5.8
14	26.1	15.2	2.5	5.3	−18.0	10.4
15	−6.3	−19.2	−20.7	−5.1	3.9	−13.8
16	11.5	10.1	1.7	−12.1	−2.7	13.9
17	4.8	3.8	0.8	2.7	1.0	14.4
18	35.2	23.1	1.2	−5.0	−16.1	3.3
19	14.0	13.1	6.6	1.6	−7.4	−3.5
20	−21.4	−19.0	0.7	0.8	−2.7	11.3

Example 165 *Using the data given in Table 15.5, we consider the linear regression model:*

$$y_i = \sum_{k=1}^{5}\beta_k x_{i,k} + u_i$$

The objective is to determine the ridge parameter λ by cross-validation.

In order to estimate the optimal value of λ, we calculate the PRESS function and find its minimum:

$$\lambda^\star = \arg\min \mathcal{P}\text{ress}\,(\lambda)$$

In Figure 15.6, we have represented the PRESS function for several values of λ. Using a bisection, we deduce that the optimal value is $\lambda^\star = 3.36$.

Remark 180 *The ridge regression is a good example where we can obtain an analytical formula for the cross-validation error $\mathcal{E}_{\text{cv}}$. In most of statistical models, this is not the case*

[9] See Exercise 15.4.2 on page 1022.

and we have to use a grid search for selecting the optimal model. This approach may be time-consuming. However, since the calibration of credit scoring models is done once per year, it is not an issue. Nevertheless, the computational time of cross-validation may be prohibitive with on-the-fly or real time statistical models.

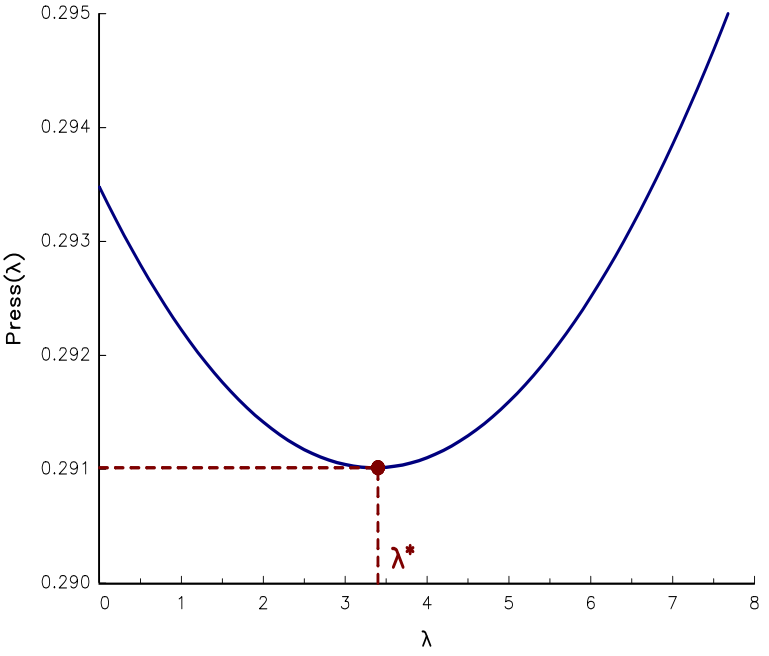


FIGURE 15.6: Selection of the ridge parameter using the PRESS statistic

15.1.3.2 Score modeling

Score modeling is the backbone of credit scoring. This is why it is extensively studied in the next two sections of this chapter. However, we present here some elements in order to understand the main challenges. The score is generally a (non-linear) function of exogenous variables X and parameters θ: $S = f(X;\theta)$. We assume that θ has been already estimated by a statistical inference method and the model $S = f\left(X;\hat{\theta}\right)$ has been validated. For example, $f\left(X;\hat{\theta}\right)$ may be a ridge regression model where $\hat{\theta} = \left(\hat{\beta},\hat{\lambda}\right)$ and $\hat{\lambda}$ has been calibrated by a cross-validation method. The score estimation $S = f\left(X;\hat{\theta}\right)$ is the preliminary part of the decision rule. Indeed, we have now to decide if we select or not the applicant. It can be done using the following rule:

$$\begin{cases} S < s \Longrightarrow Y = 0 \Longrightarrow \text{reject} \\ S \geq s \Longrightarrow Y = 1 \Longrightarrow \text{accept} \end{cases}$$

The difficulty lies in the choice of the cut-off s. For instance, if the model is a logit model, the score is a probability between 0 and 1:

$$\Pr\{Y = 1\} = \Pr\{\text{having a good risk}\} = f\left(X;\hat{\theta}\right)$$

At first sight, a natural cut-off is $s = 50\%$:

$$\begin{cases} S < 50\% \Longrightarrow Y = 0 \Longrightarrow \text{reject} \\ S \geq 50\% \Longrightarrow Y = 1 \Longrightarrow \text{accept} \end{cases}$$

However, we will see in Section 15.3 on page 1008 that $s = 50\%$ is not necessarily the optimal cut-off, in particular when the population is heterogenous[10]. Moreover, the decision rule may be influenced by other factors that are not driven by s statistical point of view. For example, if the loss associated to the selection of a bad risk is larger than the gain associated to the selection of a bad risk, the optimal cut-off may be larger than 50%.

15.1.3.3 Score follow-up

Once we have built a scoring system, we begin to collect new information about the selected applicants. We can then backtest the score in order to check its robustness. Let us consider a rating system, whose annual probability of default is given by the following table:

Rating	A	B	C	D	E	F
Probability of default	0.5%	1%	2%	5%	15%	25%

Each year, we can calculate for each grade the default frequency and adjust the decision rule in order to obtain a coherent scoring system. Below, we have reported two examples of default frequencies:

Rating	A	B	C	D	E	F
Year 1	0.05%	2.3%	2.8%	7.5%	22.6%	35.1%
Year 2	0.5%	2.7%	1.3%	2.0%	15.1%	25.1%

It is obvious that Year 2 produces closer figures to the expected result than Year 1. However, Year 2 raises more concerns than Year 1 in terms of coherency. Indeed, the default frequencies are not increasing between ratings B, C and D. On the contrary, we observe a coherent ranking for Year 1, which faces an average default rate larger than predicted.

Besides the coherency issue, the stability of the scoring system is another important key element of the follow-up. Two axes of analysis can be conducted. The first one concerns the structure of the population with respect to the score. In the table below, we report the observed frequencies of each class:

Rating	A	B	C	D	E	F
Year 0	25%	20%	20%	20%	10%	5%
Year 1	15%	20%	25%	17%	13%	10%
Year 2	15%	15%	30%	15%	15%	10%

We notice a change in the population distribution, implying that the original scoring system may be no longer valid. The second axis of analysis concerns the exogenous variables that compose the score. In this case, the analysis consists in comparing the structure of the population with respect to each variable.

Another issue is the status of the rejected applicants. Indeed, there is an asymmetry between applicants that are accepted for credit and the others. We know what accepted applicants will become in terms of good/bad risk, but we don't know what the good/bad status of rejected applicants would have been (Hand and Henley, 1997):

[10] For example when the number of good risks is larger than the number of bad risks.

> "*The behavior of those who have been rejected, if instead they had been accepted, is unknown. If one estimates a model using data only on accepted applicants, those estimated parameters may be biased when applied to all applicants. In addition, if cut-off scores are chosen to equalize the actual and predicted number of defaulting applicants then a sample of accepted applicants is likely to yield inappropriate cut-offs for the population of all applicants*" (Crook and Banasik, 2004, page 857).

The statistical study of these rejected applicants is called '*reject inference*' and can be viewed as a missing data problem[11] (Little and Rubin, 2014). Except when selected and rejected populations are perfectly coherent with the scoring decision rule, the fact that we do not observe the rejected population introduces a bias. Let us consider the example of a tight decision rule implying that we never observe a bad risk. It is obvious that the calibrated statistical model does not reflect the entire population, but only the selected population. The issue is even high because a credit scoring model does not reduce to a statistical problem, but it is used from a business point of view. Questions about the market share and the other competitors are also essential. We have reported below an illustration:

Choice	Number of selected applicants	Default rate	Total profit (in $ mn)	Per-unit profit (in $)
#1	1 000 000	5%	100	100
#2	2 000 000	7%	150	75
#3	5 000 000	10%	180	36

What is the optimal choice? If the goal is to minimize the default rate, the best choice is #1. If the goal is to maximize the total profit, the third choice is optimal. There are several statistical approaches to perform reject inference (extrapolation, augmentation, reweighting, reclassification, etc.). However, they are not satisfactory because they focus on the default rate and ignore business issues. Nevertheless, they can help to test if the credit scoring model is biased (Banasik and Crook, 2007).

15.2 Statistical methods

Unsupervised learning is a branch of statistical learning, where test data does not include a response variable. It is opposed to supervised learning, whose goal is to predict the value of the response variable Y given a set of explanatory variables X. In the case of unsupervised learning, we only know the X-values, because the Y-values do not exist or are not observed. Supervised and unsupervised learning are also called '*learning with/without a teacher*' (Hastie *et al.*, 2009). This metaphor means that we have access to the correct answer provided by the supervisor (or the teacher) in supervised learning. In the case of unsupervised learning, we have no feedback on the correct answer. For instance, the linear regression is a typical supervised learning model, whereas the principal component analysis is an approach of unsupervised learning.

[11] We generally distinguish three types of missing value problems: missing completely at random or MCAR, missing at random or MAR, and missing not at random or MNAR. Credit scoring models generally face MAR or MNAR situation.

15.2.1 Unsupervised learning

In the following paragraphs, we focus on cluster analysis and dimension reduction, which are two unsupervised approaches for detecting commonalities in data.

15.2.1.1 Clustering

Cluster analysis is a method for the assignment of observations into groups or clusters. It is then an exploratory data analysis which allows to group similar observations together. As a result, the objective of clustering methods is to maximize the proximity between observations of a same cluster and to maximize the dissimilarity between observations which belong to different clusters. In what follows, we consider two popular cluster methods: K-means and hierarchical clustering.

K-means clustering It is a special case of combinatorial algorithms. This kind of algorithm does not use a probability distribution but works directly on observed data. We consider n observations with K attributes $x_{i,k}$ ($i = 1, \ldots, n$ and $k = 1, \ldots, K$). We note x_i the $K \times 1$ vector $(x_{i,1}, \ldots, x_{i,K})$. We would like to build n_C clusters $\mathcal{C}_j$ defined by the index j where $j = 1, \ldots, n_C$ with the following properties:

1. clusters must be disjoint: $\mathcal{C}_j \cap \mathcal{C}_{j'} = \emptyset$ for $j \neq j'$;
2. clusters must describe the entire dataset: $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_{n_C} = \{1, \ldots, n\}$;
3. observations assigned to a cluster are statistically similar.

Let $\mathcal{C}$ be the mapping function which permits to assign an observation to a cluster, meaning that $\mathcal{C}(i) = j$ assigns the i^{th} observation to the j^{th} cluster $\mathcal{C}_j$ – j is also called the corresponding label. The principle of combinatorial algorithms is to adjust the mapping function $\mathcal{C}$ in order to minimize the following loss function (Hastie *et al.*, 2009):

$$\mathcal{L}(\mathcal{C}) = \frac{1}{2} \sum_{j=1}^{n_C} \sum_{\mathcal{C}(i)=j} \sum_{\mathcal{C}(i')=j} d(x_i, x_{i'})$$

where $d(x_i, x_{i'})$ is the dissimilarity measure between the observations i and i'. As a result, the optimal mapping function is denoted $\mathcal{C}^\star = \arg\min \mathcal{L}(\mathcal{C})$.

In the case of the K-means algorithm, the dissimilarity measure is the Frobenius distance (or Euclidean norm):

$$d(x_i, x_{i'}) = \sum_{k=1}^{K} (x_{i,k} - x_{i',k})^2 = \|x_i - x_{i'}\|^2$$

Therefore, the loss function becomes[12]:

$$\mathcal{L}(\mathcal{C}) = \sum_{j=1}^{n_C} n_j \sum_{\mathcal{C}(i)=j} \|x_i - \bar{x}_j\|^2$$

[12] In Exercise 15.4.3 on page 1023, we show that:

$$\sum_{\mathcal{C}(i)=j} \frac{1}{2} \sum_{\mathcal{C}(i')=j} \|x_i - x_{i'}\|^2 = \sum_{\mathcal{C}(i)=j} n_j \|x_i - \bar{x}_j\|^2$$

where $\bar{x}_j = (\bar{x}_{1,j}, ..., \bar{x}_{K,j})$ is the $(K \times 1)$ mean vector associated with the j^{th} cluster and $n_j = \sum_{i=1}^{n} \mathbb{1}\{\mathcal{C}(i) = j\}$ is the corresponding number of observations. If we note $\mu_j^\star = \arg\min \sum_{\mathcal{C}(i)=j} \|x_i - \mu_j\|^2$, the previous minimization problem is equivalent to:

$$\left\{\mathcal{C}^\star, \mu_1^\star, \ldots, \mu_{n_C}^\star\right\} = \arg\min \sum_{j=1}^{n_C} n_j \sum_{\mathcal{C}(i)=j} \|x_i - \mu_j\|^2$$

where μ_j is called the centroid of cluster $\mathcal{C}_j$. This minimization problem may be solved by the Lloyd's iterative algorithm:

1. we initialize cluster centroids $\mu_1^{(0)}, \ldots, \mu_{n_C}^{(0)}$;
2. at the iteration s, we update the mapping function $\mathcal{C}^{(s)}$ using the following rule:

$$\mathcal{C}^{(s)}(i) = \arg\min_j \left\|x_i - \mu_j^{(s-1)}\right\|^2$$

3. we then compute the optimal centroids of the clusters $\left\{\mu_1^{(s)}, \ldots, \mu_{n_C}^{(s)}\right\}$:

$$\mu_j^{(s)} = \frac{1}{n_j} \sum_{\mathcal{C}^{(s)}(i)=j} x_i$$

4. we repeat steps 2 and 3 until convergence, that is when the assignments do not change: $\mathcal{C}^\star = \mathcal{C}^{(s)} = \mathcal{C}^{(s-1)}$.

We can show that the algorithm converges to a local minimum, implying that the main issue is to determine if the solution is also a global minimum. The answer depends on the initial choice of centroids. Generally, the algorithm is initialized with random centroids. In this case, we can run the algorithm many times and choose the clusters that give the smallest value of the function $\mathcal{L}(\mathcal{C})$. We also notice that the number of clusters is an hyperparameter of the clustering model[13]. This implies that we have to test different values of n_C in order to find the '*optimal*' partition.

TABLE 15.6: Data of the clustering problem

i	X_1	X_2	X_3	X_4	X_5
1	17.6	19.6	19.8	20.4	28.8
2	13.2	17.5	17.5	17.4	24.2
3	35.9	25.4	32.4	25.0	40.7
4	28.1	24.0	25.1	28.7	26.7
5	23.5	23.6	23.7	14.3	18.1
6	36.5	30.3	29.5	32.0	29.5
7	14.0	23.9	18.3	19.2	17.2
8	36.7	29.0	30.3	21.1	28.7
9	31.2	19.4	29.9	33.3	23.8
10	17.0	20.5	23.8	16.0	19.7

[13] Originally, the K-means method defines K clusters by their means (or centroids). In this book, K is the number of explanatory variables. This is why we prefer to use the notation j for cluster labeling, while n_C represents the number of classes.

TABLE 15.7: Optimal centroids $\mu_j^\star$ for 2 and 3 clusters

$\mu_j^\star$	X_1	X_2	X_3	X_4	X_5
			$n_C = 2$		
$\mu_1^\star$	17.06	21.02	20.62	17.46	21.60
$\mu_2^\star$	33.68	25.62	29.44	28.02	29.88
			$n_C = 3$		
$\mu_1^\star$	17.06	21.02	20.62	17.46	21.60
$\mu_2^\star$	36.37	28.23	30.73	26.03	32.97
$\mu_3^\star$	29.65	21.70	27.50	31.00	25.25

Example 166 *We consider the clustering problem of* 10 *observations with five variables* X_1 *to* X_5*. The data are reported in Table 15.6. We would like to know if two clusters are sufficient or if we need more clusters to analyze the similarity.*

By setting n_C equal to 2, we obtain the following optimal clustering: $\mathcal{C}_1^\star = \{1, 2, 5, 7, 10\}$ and $\mathcal{C}_2^\star = \{3, 4, 6, 8, 9\}$. Optimal centroids are reported in Table 15.7. It follows that $\mathcal{L}(\mathcal{C}_1^\star, \mathcal{C}_2^\star) = 3\,390.32$. In the case $n_C = 3$, the optimal clustering becomes $\mathcal{C}_1^\star = \{1, 2, 5, 7, 10\}$, $\mathcal{C}_2^\star = \{3, 6, 8\}$ and $\mathcal{C}_3^\star = \{4, 9\}$ while the loss function $\mathcal{L}(\mathcal{C}_1^\star, \mathcal{C}_2^\star, \mathcal{C}_3^\star)$ is equal to $1\,832.94$. We notice that the K-means algorithm has split the second cluster into two new clusters. The loss function does not help to determine the optimal number of clusters, because $\mathcal{L}(\mathcal{C})$ tends to zero when n_C increases. The most popular approach is the Elbow method, which consists in drawing the percentage of variance explained as a function of the number of clusters and detecting when the marginal gain is small. However, there is no good solution to estimate n_C, because they generally overestimate the number of clusters[14]. This is why it is better to fix the minimum number of observations by cluster. It is obvious that two clusters are sufficient in our example, because $n_C = 3$ leads to having a cluster with only two observations.

Hierarchical clustering The K-means clustering method presents several weak points. First, it requires many iterations when the number of observations and the number of clusters are large. Second, the solution highly depends on the cluster initialization, implying that we need to run many times the Lloyd's algorithm in order to find the optimal clustering. Third, the number of clusters is definitively an issue.

The idea of hierarchical clustering is to create a tree structure in order to model the relationships between the different clusters. Unlike the K-means algorithm, this algorithm does not depend on the number of clusters or the initialization assignment. However, it depends on the dissimilarity measure between two clusters. In Figure 15.7, we have represented an example of tree structure (or dendrogram) obtained by hierarchical clustering. The 1^{st} and 3^{rd} observations are grouped in order to obtain a first cluster. This cluster is then merged with the 5^{th} observation in order to define a new cluster. In a similar way, the 6^{th} and 7^{th} observations are grouped to obtain a first cluster. This cluster is then merged with the 10^{th} observation in order to define a new cluster. The tree structure indicates how two clusters are merged into a new single cluster. The lowest level of the tree corresponds to the individual observations. In this case, each cluster contains one observation. The highest level of the tree corresponds to the entire dataset. In this case, there is only one cluster that contains all the observations.

[14] For instance, we obtain $\mathcal{C}_1^\star = \{1, 2\}$, $\mathcal{C}_2^\star = \{5, 7, 10\}$, $\mathcal{C}_3^\star = \{3, 6, 8\}$ and $\mathcal{C}_4^\star = \{4, 9\}$ when $n_C = 4$.

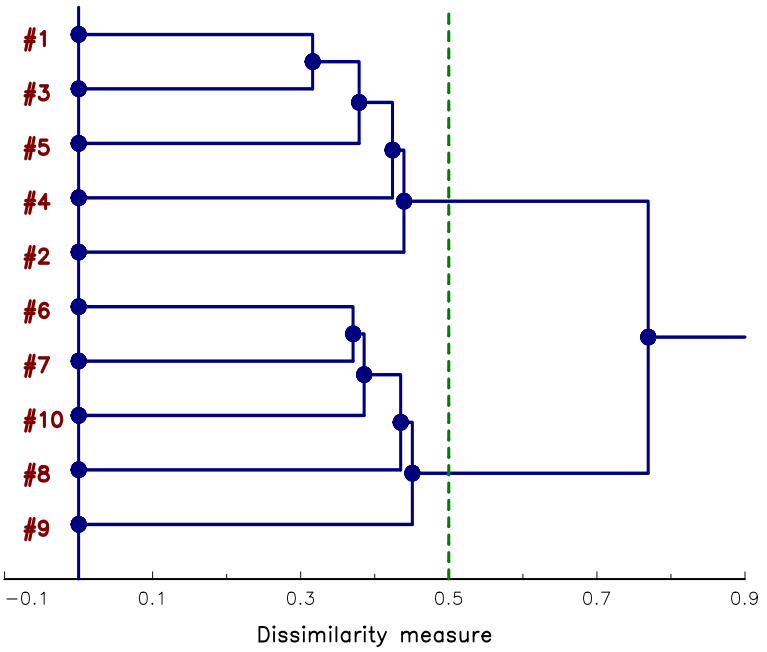


FIGURE 15.7: An example of dendrogram

Remark 181 *There is a difference between a basic tree and a dendrogram. Indeed, the x-axis of the dendrogram corresponds to the dissimilarity measure. Therefore, we can easily see which merge creates small or large dissimilarity.*

We generally distinguish two approaches for hierarchical clustering:

- in the agglomerative method (also called bottom-up clustering), the algorithm starts with the individual clusters and recursively merge the closest pairs of clusters into one single cluster;
- in the divisive method (also called the top-down clustering), the algorithm starts with the single cluster containing all the observations and recursively splits a cluster into two new clusters, which present the maximum dissimilarity.

Let $\mathcal{C}_j$ and $\mathcal{C}_{j'}$ be two clusters. The objective function of the agglomerative method is to minimize the dissimilarity measure $D(\mathcal{C}_j, \mathcal{C}_{j'})$ while we maximize the dissimilarity measure $D(\mathcal{C}_j, \mathcal{C}_{j'})$ in the divisive method. In what follows, we only consider the agglomerative method, because it is more efficient in terms of computational time and it is more widespread used.

The dissimilarity measure $D(\mathcal{C}_j, \mathcal{C}_{j'})$ is defined as a linkage function of pairwise dissimilarities $d(x_i, x_{i'})$ where $\mathcal{C}(i) = j$ and $\mathcal{C}(i') = j'$. Therefore, the agglomerative method requires defining two dissimilarity measures: the linkage function between two clusters $D(\mathcal{C}_j, \mathcal{C}_{j'})$ and the distance between two observations $d(x_i, x_{i'})$. For this last one, we generally consider the Mahalanobis distance:

$$d(x_i, x_{i'}) = \sqrt{(x_i - x_{i'})^\top \hat{\Sigma} (x_i - x_{i'})}$$

where $\hat{\Sigma}$ is the sample covariance matrix or the Minkowski distance:

$$d(x_i, x_{i'}) = \left(\sum_{k=1}^{K} |x_{i,k} - x_{i',k}|^p \right)^{1/p}$$

where $p > 1$. The case $p = 2$ corresponds to the Euclidean distance. For the linkage function, we generally consider three approaches. The single linkage (or nearest neighbor) is the smallest distance between the clusters:

$$D(\mathcal{C}_j, \mathcal{C}_{j'}) = \min_{\{\mathcal{C}(i)=j, \mathcal{C}(i')=j'\}} d(x_i, x_{i'})$$

The complete linkage (or furthest neighbor) is the largest distance between the clusters:

$$D(\mathcal{C}_j, \mathcal{C}_{j'}) = \max_{\{\mathcal{C}(i)=j, \mathcal{C}(i')=j'\}} d(x_i, x_{i'})$$

Finally, the average linkage is the average distance between the clusters:

$$D(\mathcal{C}_j, \mathcal{C}_{j'}) = \frac{1}{n_j n_{j'}} \sum_{\mathcal{C}(i)=j} \sum_{\mathcal{C}(i')=j'} d(x_i, x_{i'})$$

At each iteration, we search the clusters j and j' which minimize the dissimilarity measure and we merge them into one single cluster. When we have merged all the observations into one single cluster, the algorithm is stopped. It is also easy to perform a segmentation by considering a particular level of the tree. Indeed, we notice that the algorithm exactly requires $n - 1$ iterations. The level $L^{(s)} = s$ is then associated to the s^{th} iteration and we note $D^{(s)} = D(\mathcal{C}_{j^\star}, \mathcal{C}_{j'^\star})$ the minimum value of $D(\mathcal{C}_j, \mathcal{C}_{j'})$.

In Figure 15.7, the dendrogram was based on simulated data using the single linkage rule and the Euclidean distance. We have considered 10 observations divided into two groups. The attributes of the first (resp. second) one correspond to simulated Gaussian variates with a mean of 20% (resp. 30%) and a standard deviation of 5% (resp. 5%). The intra-group cross-correlation is set to 80% whereas the inter-group correlation is equal to 0%. We obtain satisfactory results. Indeed, if we would like to consider two clusters, the first cluster is composed of the first five observations, whereas the second cluster is composed of the last five observations. In practice, hierarchical clustering may produce concentrated segmentation as illustrated in Figure 15.8. We use the same simulated data as previously except that the standard deviation for the second group is set to 25%. In this case, if we would like to consider two clusters, we obtain a cluster with 9 elements and another cluster with only one element (the 6^{th} observation).

Let us consider Example 166 on page 946. By using the Euclidean distance, we obtain the dendrograms in Figure 15.9. If we would like to split the data into two clusters, we find for the three methods the solution $\{1, 2, 5, 7, 10\}$ and $\{3, 4, 6, 8, 9\}$, which also corresponds to the solution given by the K-means analysis. In the case of the single linkage method, we have reported in Table 15.8 for each level $L^{(s)}$ the distance $D^{(s)}$, the two nearest neighbours $i^\star$ and $i'^\star$ and the created cluster $\mathcal{C}^{(s)}$. We notice that the solution for 3 clusters differs from the K-means solution. Indeed, we find $\{1, 2, 5, 7, 10\}$, $\{4, 6, 8, 9\}$ and $\{3\}$ for the single linkage method.

15.2.1.2 Dimension reduction

We now turn to the concept of dimension reduction, which consists in finding some common patterns in order to better explain the data. For instance, we might want to reduce

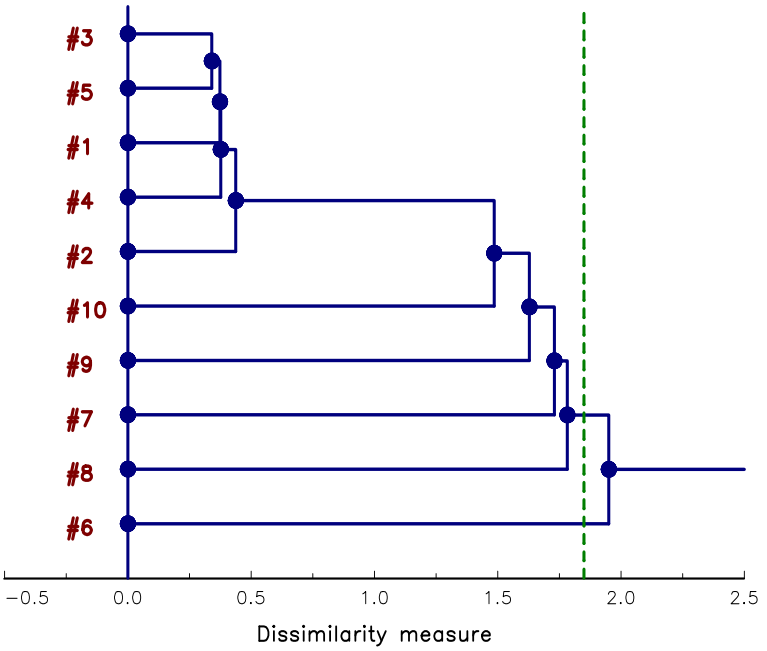


FIGURE 15.8: Unbalanced clustering

TABLE 15.8: Agglomerative hierarchical clustering (single linkage)

$L^{(s)}$	$D^{(s)}$	$(i^\star, i'^\star)$	$\mathcal{C}^{(s)}$
1	7.571	$(5, 10)$	$\{5, 10\}$
2	7.695	$(1, 2)$	$\{1, 2\}$
3	8.204	$(5, 7)$	$\{5, 7, 10\}$
4	9.131	$(4, 9)$	$\{4, 9\}$
5	9.238	$(1, 5)$	$\{1, 2, 5, 7, 10\}$
6	11.037	$(6, 8)$	$\{6, 8\}$
7	12.179	$(4, 6)$	$\{4, 6, 8, 9\}$
8	13.312	$(3, 4)$	$\{3, 4, 6, 8, 9\}$
9	15.199	$(1, 3)$	$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

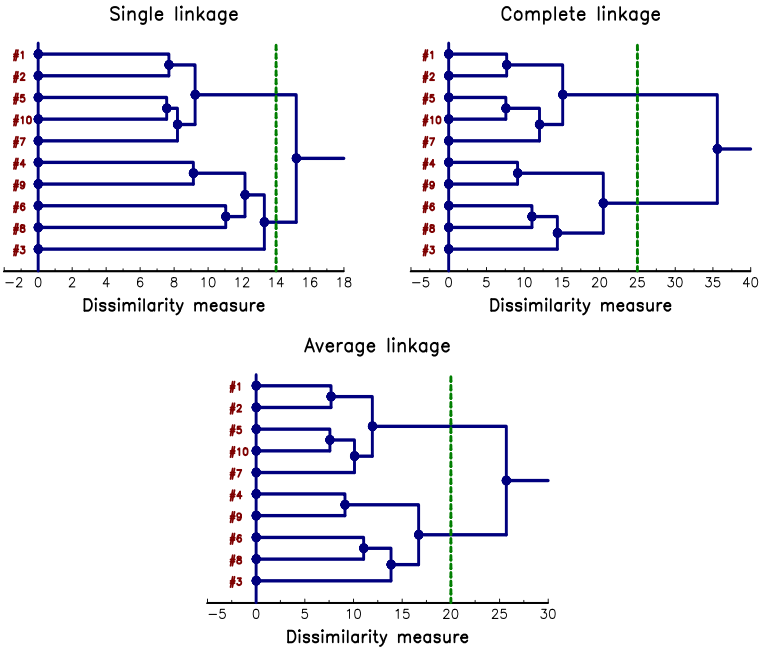


FIGURE 15.9: Comparison of the three dendrograms

a dataset with 1 000 variables to the two or three most important patterns. In machine learning, dimension reduction is also known as feature extraction, which is defined as the process to build new variables or features that are more informative and less redundant than the original variables.

Principal component analysis Let X be a $K \times 1$ random vector, whose covariance matrix is equal to Σ. We consider the linear transform $Z = B^\top X$ where $B = (\beta_1, \ldots, \beta_K)$ is a $K \times K$ matrix and the β's are $K \times 1$ vectors. The j^{th} element of Z is denoted by Z_j and we have $Z_j = \beta_j^\top X = \sum_{k=1}^K \beta_{k,j} X_k$. Z_j is also called the j^{th} principal component. The idea of PCA is to find a first linear function $\beta_1^\top X$ such that the variance of Z_1 is maximum and then a j^{th} linear function $\beta_j^\top X$ such that the variance of Z_j is maximum and Z_j is uncorrelated with $Z_1, \ldots, Z_{j-1}$ for all $j \geq 2$ (Jolliffe, 2002). We can show that B is the matrix of eigenvectors[15] of the covariance matrix Σ:

$$\Sigma B = B\Lambda$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_K)$ is the diagonal matrix of eigenvalues with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$. Since Σ is a symmetric and positive define matrix, we also have:

$$\Sigma = B\Lambda B^{-1} = B\Lambda B^\top$$

[15] See Exercise 15.4.4 on page 1024 for the derivation of this result.

and B is an orthonormal matrix. By construction, we have[16]:

$$\begin{aligned}\text{var}\,(Z_j) &= \text{var}\left(\beta_j^\top X\right)\\ &= \beta_j^\top \Sigma \beta_j\\ &= \Lambda_{j,j}\\ &= \lambda_j\end{aligned}$$

and:

$$\begin{aligned}\text{cov}\,(Z_j, Z_{j'}) &= \beta_j^\top \Sigma \beta_{j'}\\ &= \Lambda_{j,j'}\\ &= 0\end{aligned}$$

We deduce the spectral decomposition of the covariance matrix:

$$\Sigma = B\Lambda B^\top = \sum_{j=1}^{K} \lambda_j \beta_j \beta_j^\top$$

We note $B_{(1:j)}$ and $B_{(K-j+1:K)}$ the matrices that contains the first and last j columns of B. We consider the random vector $\tilde{Z} = \left(\tilde{Z}_1, \ldots, \tilde{Z}_j\right) = \tilde{B}^\top X$ of dimension j. Here are some properties of the PCA (Jolliffe, 2002):

1. the trace of cov $\left(\tilde{Z}\right)$ is maximized if $\tilde{B} = B_{(1:j)}$ corresponds to the j first eigenvectors;
2. the trace of cov $\left(\tilde{Z}\right)$ is minimized if $\tilde{B} = B_{(K-j+1:K)}$ corresponds to the j last eigenvectors;
3. the covariance of X given $\tilde{Z}$ is:

$$\begin{aligned}\text{cov}\left(X \mid \tilde{Z}_1, \ldots, \tilde{Z}_j\right) &= \Sigma_{X,X} - \Sigma_{X,\tilde{Z}} \Sigma_{\tilde{Z},\tilde{Z}}^{-1} \Sigma_{\tilde{Z},X}\\ &= \sum_{k=j+1}^{K} \lambda_k \beta_k \beta_k^\top\end{aligned}$$

4. we consider the following linear regression model:

$$X = A\tilde{Z} + U$$

where A is a $K \times j$ matrix and $U = (U_1, \ldots, U_K)$ is the vector of residuals; if we note $\Omega = \text{diag}\left(\sigma_1^2, \ldots, \sigma_K^2\right)$ the covariance matrix of U, the trace of Ω is minimized if $\tilde{B} = B_{(1:j)}$.

Remark 182 *The principal component analysis can be performed with correlation matrices instead of covariance matrices. Jolliffe (2002) presents different arguments for justifying this choice. Indeed, PCA makes more sense when the variables are comparable. Otherwise, the principal components are dominated by the variables with the largest variances.*

[16] Because of the following equality:

$$\Lambda = B^{-1}\Sigma B = B^\top \Sigma B$$

Example 167 *We consider the random vector* $X = (X_1, X_2, X_3, X_4)$*, whose individual variances are equal to* 1*,* 2*,* 3 *and* 4*. The correlation matrix is:*

$$\rho = \begin{pmatrix} 1.00 & & & \\ 0.30 & 1.00 & & \\ 0.50 & 0.10 & 1.00 & \\ 0.20 & 0.50 & -0.50 & 1.00 \end{pmatrix}$$

We have reported the eigendecomposition of Σ and ρ in Tables 15.9 and 15.10. We observe some differences. For instance, the first principal component of the covariance matrix Σ is:

$$Z_1 = 0.18 \cdot X_1 + 0.33 \cdot X_2 + 0.53 \cdot X_3 + 0.76 \cdot X_4$$

whereas the first principal component of the correlation matrix ρ is:

$$Z_2 = 0.48 \cdot X_1 + 0.44 \cdot X_2 + 0.53 \cdot X_3 + 0.55 \cdot X_4$$

We verify that the sum of eigenvalues is equal to the sum of variances for the covariance matrix, and the number of variables for the correlation matrix[17].

TABLE 15.9: Eigendecomposition of the covariance matrix

	β_1	β_2	β_3	β_4
X_1	0.18	−0.20	−0.57	0.77
X_2	0.33	0.58	−0.63	−0.40
X_3	0.53	−0.73	−0.13	−0.41
X_4	0.76	0.31	0.50	0.27
λ_j	5.92	2.31	1.31	0.46

TABLE 15.10: Eigendecomposition of the correlation matrix

	β_1	β_2	β_3	β_4
X_1	0.48	−0.44	−0.65	−0.40
X_2	0.44	0.67	−0.40	0.45
X_3	0.53	−0.51	0.38	0.57
X_4	0.55	0.33	0.53	−0.56
λ_j	2.06	0.97	0.73	0.23

We now develop the interpretation tools of PCA. The quality of representation is defined as the percentage of total variance that is explained by the j^{th} principal component (or PC):

$$\mathbb{Q}_j = \frac{\lambda_j}{\sum_{k=1}^K \lambda_k}$$

We have $0 \leq \mathbb{Q}_j \leq 1$. The cumulative quality of representation is just the cumulative sum of the quality values:

$$\mathbb{Q}_j^\star = \sum_{k=1}^j \mathbb{Q}_k = \frac{\sum_{k=1}^j \lambda_k}{\sum_{k=1}^K \lambda_k}$$

[17] We have $\text{trace}(\Sigma) = \sum_{k=1}^K \sigma_k^2$ and $\text{trace}\left(B\Lambda B^{-1}\right) = \text{trace}\left(\Lambda B^{-1} B\right) = \text{trace}(\Lambda) = \sum_{j=1}^K \lambda_j$. For the correlation matrix, we deduce that $\sum_{j=1}^K \lambda_j = K$.

$\mathbb{Q}_j^\star$ is also called the quality of representation of the j^{th} principal plane[18]. The correlation between the variable X_k and the factor Z_j is given by:

$$\text{cor}\,(X_k, Z_j) = \beta_{k,j}\sqrt{\lambda_j}$$

It follows that the quality of representation of the variable X_k with respect to the j^{th} PC is[19]:

$$\mathbb{Q}_{k,j} = \text{cor}^2\,(X_k, Z_j) = \beta_{k,j}^2\lambda_j$$

We can also define the contribution of the variable X_k to the j^{th} PC[20]:

$$\mathbb{C}_{k,j} = \beta_{k,j}^2$$

In order to the understand the association between variables, we generally plot the correlation circle between two principal components that corresponds to the scatterplot of $\text{cor}\,(X_k, Z_j)$ and $\text{cor}\,(X_k, Z_{j'})$.

Remark 183 *In practice, we estimate the covariance or the correlation matrix using a sample. Let $x_i = (x_{i,1}, \ldots, x_{i,K})$ be the i^{th} observation. We note $z_{i,j} = \beta_j^\top x_i$ the projection of x_i onto the j^{th} principal component. The quality of representation and the contribution of an observation to a principal component are then equal to:*

$$\mathbb{Q}_{i,j} = \frac{z_{i,j}^2}{\sum_{k=1}^K z_{i,k}^2}$$

and:

$$\mathbb{C}_{i,j} = \frac{z_{i,j}^2}{\sum_{i=1}^n z_{i,j}^2}$$

We consider again Example 166 on page 946. In Table 15.11, we have reported the results of the PCA applied to the correlation matrix of data. The first PC explains 68.35% of the variance, while the quality of representation of the second PC is equal to 14.54%. This means that we can explain 82.89% with only two factors. For the first factor, each variable has a positive loading. This is not the case of the second factor, where the factor loadings of X_1, X_2 and X_3 are negative. We notice that X_1 and X_3 are well represented by Z_1 (95.81% and 86.95%). For the second PC, the second variable X_2 is the most represented (41.40%). If we consider the last PC, the quality of representation is poor (less than 1%). This indicates that the last PC has a very low explanation power. We notice that the rationale of the fourth PC is to model X_3 because the second and third PCs do not explain this variable. The contribution values $\mathbb{C}_{k,j}$ are also interesting to confirm the previous results. For instance, X_1 does not contribute to Z_2. It follows that the second PC represents the opposition of X_2

[18] That is the plane composed of the first j principal components.

[19] We verify that the sum of $\mathbb{Q}_{k,j}$ is equal to the variance of the j^{th} PC:

$$\sum_{k=1}^K \mathbb{Q}_{k,j} = \sum_{k=1}^K \beta_{k,j}^2\lambda_j = \lambda_j \sum_{k=1}^K \beta_{k,j}^2 = \lambda_j$$

[20] We verify that the sum of $\mathbb{C}_{k,j}$ is equal to 100%:

$$\sum_{k=1}^K \mathbb{C}_{k,j} = \sum_{k=1}^K \beta_{k,j}^2 = 1$$

TABLE 15.11: Principal component analysis of Example 166

Factor	Z_1	Z_2	Z_3	Z_4	Z_5
λ_j	3.4173	0.7271	0.5548	0.2783	0.0226
$\mathbb{Q}_j$	68.35%	14.54%	11.10%	5.57%	0.45%
$\mathbb{Q}_j^\star$	68.35%	82.89%	93.98%	99.55%	100.00%
Matrix B of eigenvectors					
X_1	0.5295	−0.1015	−0.0567	−0.2554	0.8006
X_2	0.3894	−0.7546	−0.0500	0.4855	−0.2019
X_3	0.5044	−0.0188	−0.0247	−0.6650	−0.5499
X_4	0.3952	0.5318	−0.6238	0.3995	−0.1107
X_5	0.3967	0.3702	0.7775	0.3120	−0.0609
Correlation between X_k and Z_j					
X_1	97.88%	−8.66%	−4.22%	−13.47%	12.03%
X_2	71.98%	−64.35%	−3.72%	25.61%	−3.03%
X_3	93.25%	−1.60%	−1.84%	−35.08%	−8.27%
X_4	73.06%	45.35%	−46.46%	21.07%	−1.66%
X_5	73.34%	31.57%	57.91%	16.46%	−0.92%
Quality of representation of each variable $\mathbb{Q}_{k,j}$					
X_1	95.81%	0.75%	0.18%	1.82%	1.45%
X_2	51.81%	41.40%	0.14%	6.56%	0.09%
X_3	86.95%	0.03%	0.03%	12.31%	0.68%
X_4	53.38%	20.57%	21.59%	4.44%	0.03%
X_5	53.78%	9.96%	33.54%	2.71%	0.01%
Contribution of each variable $\mathbb{C}_{k,j}$					
X_1	28.04%	1.03%	0.32%	6.52%	64.09%
X_2	15.16%	56.94%	0.25%	23.57%	4.08%
X_3	25.44%	0.04%	0.06%	44.22%	30.24%
X_4	15.62%	28.29%	38.91%	15.96%	1.23%
X_5	15.74%	13.70%	60.46%	9.73%	0.37%

with respect to X_4 and X_5. Clearly, the third PC mainly concerns X_4 and X_5. Figure 15.10 represents the scatterplot of the factor values $z_{i,j}$ for the first two principal components. We notice that the second component classifies the observations in the same way than the K-means algorithm or the agglomerative hierarchical clustering. Indeed, we retrieve the two clusters $\{1, 2, 5, 7, 10\}$ and $\{3, 4, 6, 8, 9\}$. This is not the case of the first component, which operates the following classification $\{1, 2, 3, 4, 9\}$ and $\{5, 6, 7, 8, 10\}$. In Figure 15.11, we have reported the correlation circle between different PCs. If we consider the first two PCs, the variables X_1 and X_2 are clearly opposed to the variables X_4 and X_5. The second panel confirms the competition between X_4 and X_5 due to the third PC.

Non-negative matrix factorization There are several alternative approaches to principal component analysis. For instance, independent component analysis (ICA) estimates additive factors that are maximally independent. Another popular method is the non-negative matrix factorization (NMF). Let A be a non-negative matrix $m \times p$. We define the NMF decomposition of A as follows:

$$A \approx BC$$

where B and C are two non-negative matrices with respective dimensions $m \times n$ and $n \times p$. Compared to classic decomposition algorithms, we remark that BC is an approximation

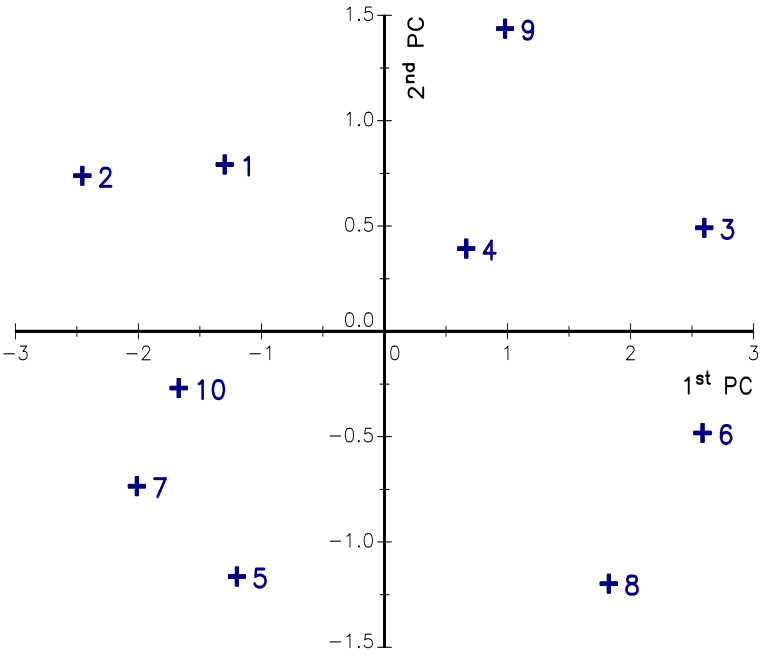


FIGURE 15.10: Scatterplot of the factor values $z_{i,1}$ and $z_{i,2}$

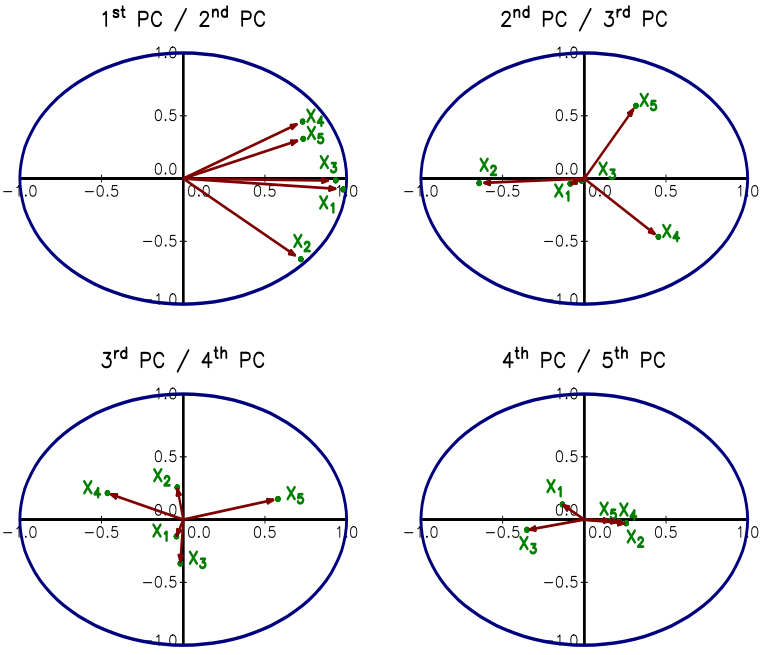


FIGURE 15.11: PCA correlation circle

of A. There are also different ways to obtain this approximation meaning that B and C are not necessarily unique[21]. We also notice that the decomposition $A \approx BC$ is equivalent to $A^\top \approx C^\top B^\top$. It means that the storage of the data is not important. Rows of A may represent either the observations or the variables, but the interpretation of the B and C matrices depend on the choice of the storage. We remark that:

$$A_{i,j} = \sum_{k=1}^{n} B_{i,k} C_{k,j}$$

Suppose that we consider a variable/observation storage. Therefore, $B_{i,k}$ depends on the variable i whereas $C_{k,j}$ depends on the observation j. In this case, we may interpret B as a matrix of weights. In factor analysis, B is called the loading matrix and C is the factor matrix. $B_{i,k}$ is then the weight of factor k for variable i and $C_{k,j}$ is the value taken by factor k for observation j. If we use an observation/variable storage which is the common way to store data in statistics, B and C become the factor matrix and the loading matrix.

Because the dimensions m, n and p may be very large, one of the difficulties with NMF is to derive a numerical algorithm with a reasonable computational time. Lee and Seung (1999) developed a simple algorithm with strong performance and applied it to pattern recognition with success. Since this seminal work, this algorithm has been improved and there are today several ways to obtain a non-negative matrix factorization. In order to find the approximate factorization, we need to define the loss function $\mathcal{L}$ which measures the quality of the factorization. The optimization program is then:

$$\begin{aligned} \{B^\star, C^\star\} &= \arg\min \mathcal{L}(A, BC) \\ &\text{u.c.} \begin{cases} B \succ \mathbf{0} \\ C \succ \mathbf{0} \end{cases} \end{aligned} \tag{15.6}$$

Lee and Seung (2001) considered two loss functions. The first one is the Frobenious norm:

$$\mathcal{L}(A, BC) = \sum_{i=1}^{m} \sum_{j=1}^{p} \left(A_{i,j} - (BC)_{i,j} \right)^2$$

whereas the second one is Kullback-Leibler divergence:

$$\mathcal{L}(A, BC) = \sum_{i=1}^{m} \sum_{j=1}^{p} \left(A_{i,j} \ln \frac{A_{i,j}}{(BC)_{i,j}} - A_{i,j} + (BC)_{i,j} \right)$$

To solve Problem (15.6), Lee and Seung (2001) proposed to use the multiplicative update algorithm. Let $B_{(s)}$ and $C_{(s)}$ be the matrices at iteration s. For the Frobenious norm[22], we have:

$$\begin{cases} B_{(s+1)} = B_{(s)} \odot \left(A C_{(s)}^\top \right) \oslash \left(B_{(s)} C_{(s)} C_{(s)}^\top \right) \\ C_{(s+1)} = C_{(s)} \odot \left(B_{(s+1)}^\top A \right) \oslash \left(B_{(s+1)}^\top B_{(s+1)} C_{(s)} \right) \end{cases}$$

[21] Let D be a nonnegative matrix such that D^{-1} is nonnegative too. For example, D may be a permutation of a diagonal matrix. In this case, we have:

$$A \approx BD^{-1}DC \approx B'C'$$

where $B' = BD^{-1}$ and $C' = DC$ are two nonnegative matrices. This shows that the decomposition is not unique.

[22] A similar algorithm may be derived for the Kullback-Leibler divergence.

where $\odot$ and $\oslash$ are respectively the element-wise multiplication and division operators. Under some assumption, we may show that $B^\star = B_{(\infty)}$ and $C^\star = C_{(\infty)}$, meaning that the multiplicative update algorithm converges to the optimal solution.

For large datasets, the computational time to find the optimal solution may be large with the previous algorithm. Since the seminal work of Lee and Seung, a lot of methods have also been proposed to improve the multiplicative update algorithm and speed the converge. Among these methods, we may mention the algorithm developed by Lin (2007), which is based on the alternating non-negative least squares:

$$\begin{cases} B_{(s+1)} = \arg\min \mathcal{L}\left(A, BC_{(s)}\right) \\ C_{(s+1)} = \arg\min \mathcal{L}\left(A, B_{(s+1)}C\right) \end{cases} \tag{15.7}$$

with the constraints $B_{(s+1)} \succ \mathbf{0}$ and $C_{(s+1)} \succ \mathbf{0}$. We notice that the two optimization problems (15.7) are symmetric because we may cast the first problem in the form of the second problem: $B_{(s+1)}^\top = \arg\min \mathcal{L}\left(A^\top, C_{(s)}^\top B^\top\right)$. So, we may only focus on the following optimization problem:

$$\begin{aligned} C^\star &= \arg\min \mathcal{L}(A, BC) \\ &\text{u.c.} \quad C \succ \mathbf{0} \end{aligned}$$

In the case of the Frobenious norm, we have $\partial_C \mathcal{L}(A, BC) = 2B^\top(BC - A)$. The projected gradient method consists in the following iterating scheme:

$$C \leftarrow C - \alpha \cdot \frac{\partial \mathcal{L}(A, BC)}{\partial C}$$

where α is the descent length. Let (β, γ) be two scalars in $]0, 1[$. Instead of finding the optimal value of α at each iteration, Lin (2007) proposed to update α in a very simple way depending on the inequality equation:

$$(1-\gamma)\frac{\partial \mathcal{L}(A, BC)}{\partial C}^\top \left(\tilde{C} - C\right) + \frac{1}{2}\left(\tilde{C} - C\right)^\top \frac{\partial^2 \mathcal{L}(A, BC)}{\partial C\, \partial C^\top}\left(\tilde{C} - C\right) \leq 0$$

where $\tilde{C}$ is the update of C. If this inequality equation is verified, α is increased ($\alpha \leftarrow \alpha/\beta$), otherwise α is decreased ($\alpha \leftarrow \alpha\beta$).

Remark 184 *The choice of $B_{(0)}$ and $C_{(0)}$ for initializing NMF algorithms is important. The random method consists in generating matrices with positive random numbers*[23]*. Another popular approach is the non-negative double singular value decomposition, which is a modification of the singular value decomposition by considering only the non-negative part of the singular values (Boutsidis and Gallopoulos, 2008).*

In order to understand why NMF is different from other factor methods, we consider a simulation study. We consider a basket of four financial assets. The asset prices are driven by a multidimensional geometric Brownian motion. The drift parameter is equal to 5% whereas the diffusion parameter is 20%. The cross-correlation $\rho_{i,j}$ between assets i and j is equal to 20%, but $\rho_{1,2} = 70\%$ and $\rho_{3,4} = 50\%$. In order to preserve the time homogeneity, the data correspond to $x_{i,t} = \ln S_{i,t}$ where $S_{i,t}$ is the price of the asset i at time t. In Figure 15.12, we report the time series $x_{i,t}$ for the four assets (panel 1) and, the first factor estimated

[23] For example, we can use the probability distributions $\mathcal{U}_{[0,1]}$ or $|\mathcal{N}(0,1)|$.

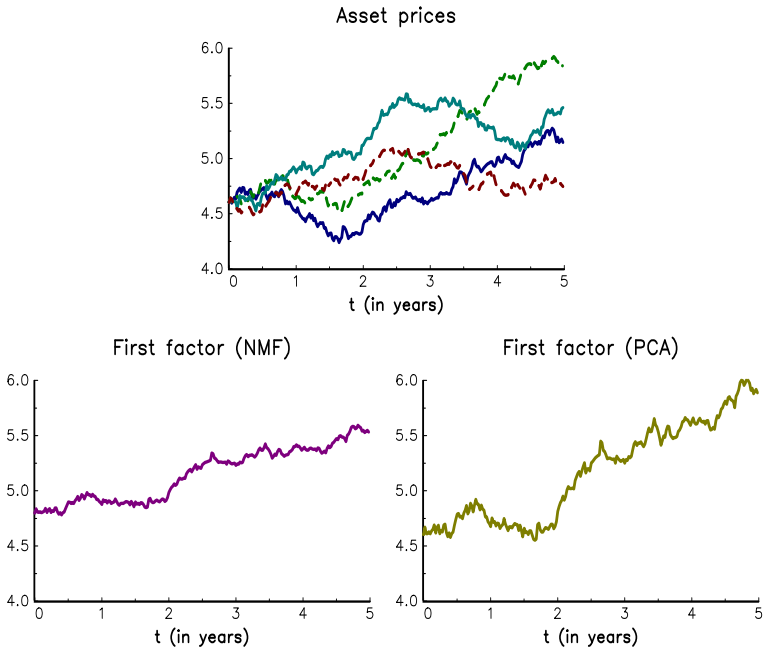


FIGURE 15.12: Estimating the first factor of a basket of financial assets

by NMF[24] (panel 2) and PCA (panel 3). We notice that the NMF factor[25] is not scaled in the same way than the PCA factor. However, the correlation between the first differences is equal to 98.8%. In the first panel in Figure 15.13, we compare the decomposition of the variance according to the factors. We notice that PCA explains more variance than NMF for a given number of factors. We obtain this result because NMF may be viewed as a constrained principal component analysis with nonnegative matrices. However, it does not mean that each PCA factor explains more variance than the corresponding NMF factor. For example, the second NMF factor explains more variance than the second PCA factor in Figure 15.13. In the other panels, we compare the dynamics of the first asset with the dynamics given by the NMF factors[26]. With three risk factors, the reconstructed signal has a correlation of 93.7% with the original signal.

15.2.2 Parametric supervised methods

15.2.2.1 Discriminant analysis

Discriminant analysis was first developed by Fisher (1936). This approach is close to the principal component analysis (PCA) and is used to predict class membership for independent variables. For that, we assume that we have n_C disjoint classes $\mathcal{C}_j$ where $j = 1, \ldots, J$. Discriminant analysis consists then in assigning an observation to one and only one class.

[24]The NMF decomposition corresponds to:

$$\underbrace{\ln X}_{n_X \times n_T} \approx \underbrace{B}_{n_X \times n_F} \cdot \underbrace{C}_{n_F \times n_T}$$

where n_X is the number of time-series, n_T is the number of dates and n_F is the number of NMF factors.

[25]In this example, B is the loading matrix while C is the matrix of time-series factors.

[26]The reconstructed multidimensional signal is just the matrix product BC for different values of n_F.

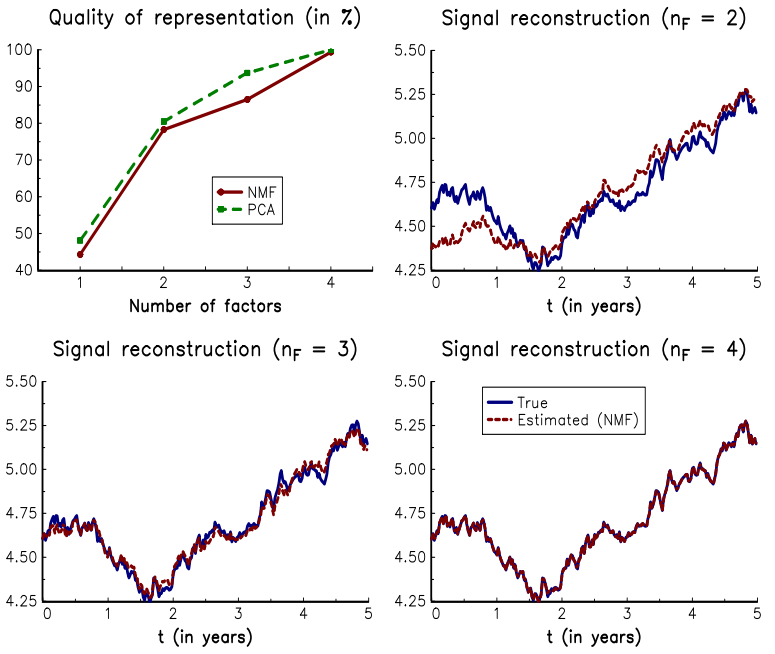


FIGURE 15.13: Variance decomposition and signal reconstruction

We consider an input vector x and we divide the input space into n_C decision regions, whose boundaries are called decision boundaries (Bishop, 2006). Classification methods can then be seen as a supervised clustering methods, where the categorical response variable is directly the class. For example, Figure 15.14 corresponds to a classification problem with seven classes and two explanatory variables X_1 and X_2. The goal is then to predict for each observation its class. For instance, we would like that the model predicts that the first observation belongs to the first class, the second observation belongs to the fifth class, etc.

The two-dimensional case Using the Bayes theorem, we have:

$$\begin{aligned}\Pr\{A \cap B\} &= \Pr\{A \mid B\} \cdot \Pr\{B\} \\ &= \Pr\{B \mid A\} \cdot \Pr\{A\}\end{aligned}$$

It follows that:

$$\Pr\{A \mid B\} = \Pr\{B \mid A\} \cdot \frac{\Pr\{A\}}{\Pr\{B\}}$$

If we apply this result to the conditional probability $\Pr\{i \in \mathcal{C}_1 \mid X = x\}$, we obtain:

$$\Pr\{i \in \mathcal{C}_1 \mid X = x\} = \Pr\{X = x \mid i \in \mathcal{C}_1\} \cdot \frac{\Pr\{i \in \mathcal{C}_1\}}{\Pr\{X = x\}}$$

The log-probability ratio is then equal to:

$$\begin{aligned}\ln \frac{\Pr\{i \in \mathcal{C}_1 \mid X = x\}}{\Pr\{i \in \mathcal{C}_2 \mid X = x\}} &= \ln \left(\frac{\Pr\{X = x \mid i \in \mathcal{C}_1\}}{\Pr\{X = x \mid i \in \mathcal{C}_2\}} \cdot \frac{\Pr\{i \in \mathcal{C}_1\}}{\Pr\{i \in \mathcal{C}_2\}} \right) \\ &= \ln \frac{\Pr\{X = x \mid i \in \mathcal{C}_1\}}{\Pr\{X = x \mid i \in \mathcal{C}_2\}} + \ln \frac{\Pr\{i \in \mathcal{C}_1\}}{\Pr\{i \in \mathcal{C}_2\}} \\ &= \ln \frac{f_1(x)}{f_2(x)} + \ln \frac{\pi_1}{\pi_2}\end{aligned}$$

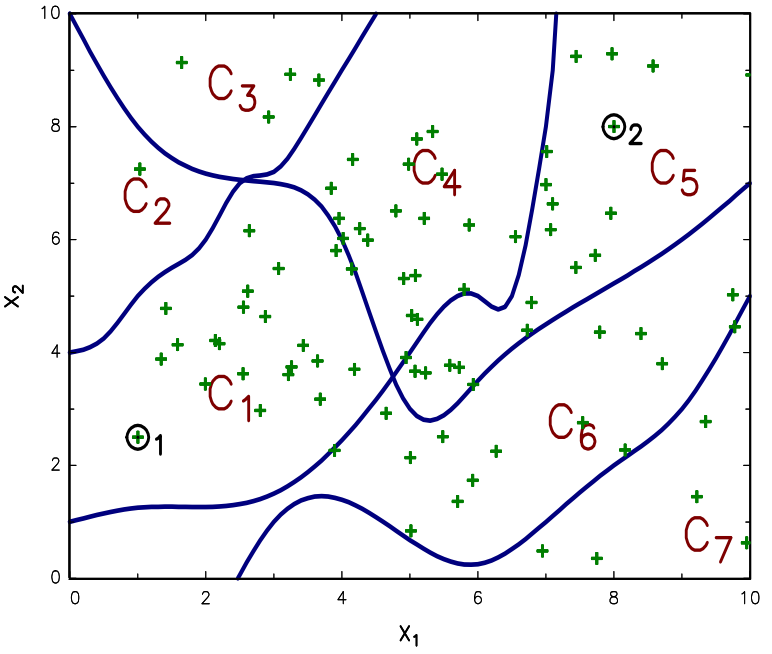


FIGURE 15.14: Classification statistical problem

where $\pi_j = \Pr\{i \in \mathcal{C}_j\}$ is the probability of the j^{th} class and $f_j(x) = \Pr\{X = x \mid i \in \mathcal{C}_j\}$ is the conditional probability density function of X. By construction, the decision boundary is defined such that we are indifferent to an assignment rule ($i \in \mathcal{C}_1$ and $i \in \mathcal{C}_2$), implying that:

$$\Pr\{i \in \mathcal{C}_1 \mid X = x\} = \Pr\{i \in \mathcal{C}_2 \mid X = x\} = \frac{1}{2}$$

Finally, we deduce that the decision boundary satisfies the following equation:

$$\ln \frac{f_1(x)}{f_2(x)} + \ln \frac{\pi_1}{\pi_2} = 0$$

If we model each class density as a multivariate normal distribution:

$$X \mid i \in \mathcal{C}_j \sim \mathcal{N}(\mu_j, \Sigma_j)$$

we have:

$$f_j(x) = \frac{1}{(2\pi)^{K/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_j)^\top \Sigma_j^{-1}(x-\mu_j)\right)$$

We deduce that:

$$\begin{aligned}\ln \frac{f_1(x)}{f_2(x)} &= \frac{1}{2}\ln\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(x-\mu_1)^\top \Sigma_1^{-1}(x-\mu_1) + \\ &\quad \frac{1}{2}(x-\mu_2)^\top \Sigma_2^{-1}(x-\mu_2)\end{aligned}$$

The decision boundary is then given by:

$$\begin{aligned}&\frac{1}{2}\ln\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(x-\mu_1)^\top \Sigma_1^{-1}(x-\mu_1) + \\ &\quad \frac{1}{2}(x-\mu_2)^\top \Sigma_2^{-1}(x-\mu_2) + \ln\frac{\pi_1}{\pi_2} = 0 \qquad (15.8)\end{aligned}$$

Since the decision boundary is quadratic in x, such approach is called quadratic discriminant analysis (QDA).

If we assume that $\Sigma_1 = \Sigma_2 = \Sigma$, Equation (15.8) becomes:

$$\frac{1}{2}(x-\mu_2)^\top \Sigma^{-1}(x-\mu_2) - \frac{1}{2}(x-\mu_1)^\top \Sigma^{-1}(x-\mu_1) + \ln\frac{\pi_1}{\pi_2} = 0$$

or:

$$(\mu_2-\mu_1)^\top \Sigma^{-1} x = \frac{1}{2}\left(\mu_2^\top \Sigma^{-1}\mu_2 - \mu_1^\top \Sigma^{-1}\mu_1\right) + \ln\frac{\pi_2}{\pi_1} \tag{15.9}$$

It follows that the decision boundary is then linear in x. This is why we called this approach the linear discriminant analysis (LDA).

Example 168 *We consider two classes and two explanatory variables* $X = (X_1, X_2)$ *where* $\pi_1 = 50\%$, $\pi_2 = 1 - \pi_1 = 50\%$, $\mu_1 = (1, 3)$, $\mu_2 = (4, 1)$, $\Sigma_1 = I_2$ *and* $\Sigma_2 = \gamma I_2$ *where* $\gamma = 1.5$.

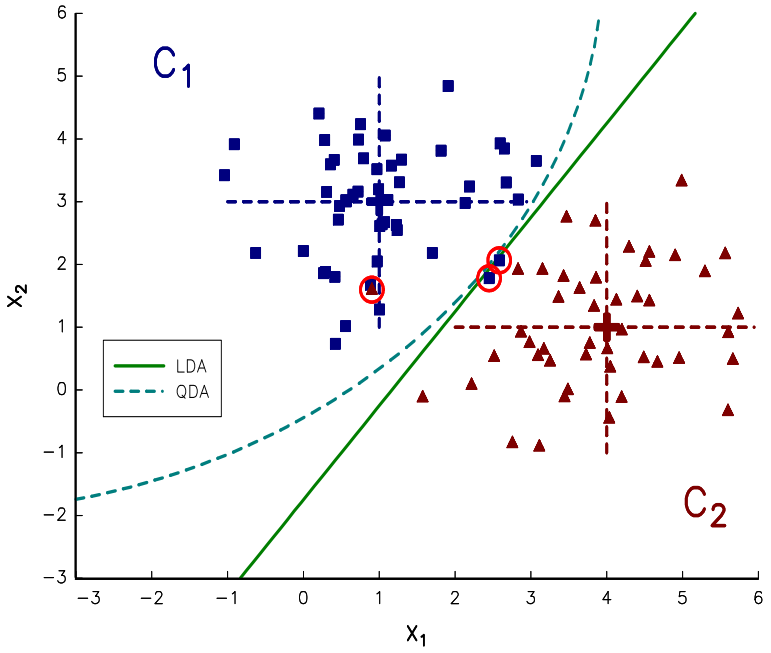


FIGURE 15.15: Boundary decision of discriminant analysis

By solving Equations (15.8) and (15.9), we obtain the QDA and LDA decision boundaries[27] reported in Figure 15.15. We verify that the LDA decision boundary is linear while the QDA decision region is convex. For each class, we have also simulated 50 realizations. We observe that the discriminant analysis performs the right classification most of the times. However, we notice that two observations from class $\mathcal{C}_1$ and one observation from class $\mathcal{C}_2$ are not properly classified. In Figure 15.16, we analyze the impact of the parameters on the decision boundary. The top/left panel corresponds to the previous example, whereas we only change one parameter for each other panel. For instance, we increase the variance of the second variable in the top/right panel. We observe that the impact on the LDA decision

[27] For the linear discriminant analysis, we have used $\Sigma = (\Sigma_1 + \Sigma_2)/2$.

boundary is minor, but this is not the case for the QDA decision boundary. Indeed, the convexity is stronger because X_2 can take more larger values than X_1. This is why for the extreme values, the QDA decision boundary can be approximated by a vertical line when $x_1 \to -\infty$ and an horizontal line when $x_2 \to +\infty$. Let us now introduce a correlation ρ between X_1 and X_2. It follows that the QDA decision boundary becomes more and more linear when we increase ρ (bottom/left panel). Finally, the impact of the probabilities (π_1, π_2) is crucial as shown in the bottom/right panel. It is obvious that the boundary decision moves to the right when π_1 increases, because the decision region concerning $i \in \mathcal{C}_1$ must be larger. For instance, we must always accept $i \in \mathcal{C}_1$ at the limit case $\pi_1 = 100\%$.

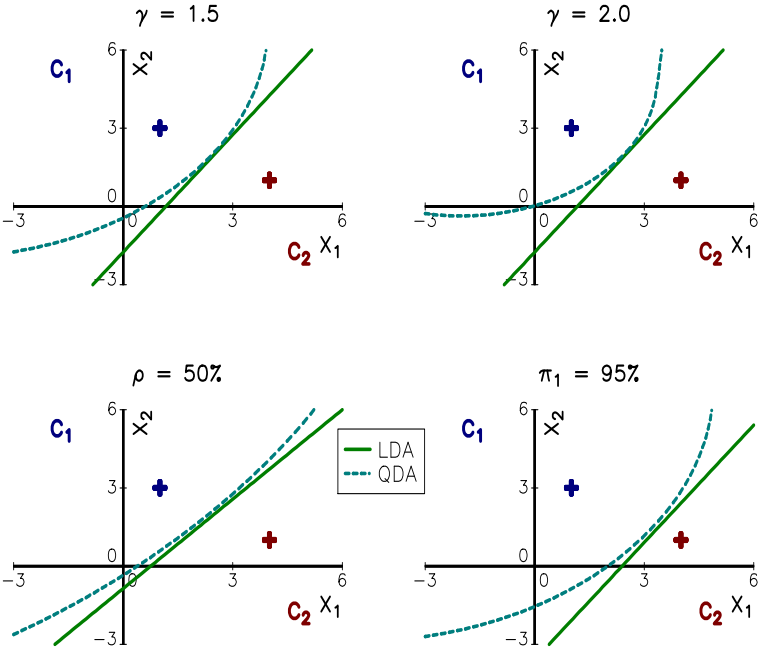


FIGURE 15.16: Impact of the parameters on LDA/QDA boundary decisions

The general case We can generalize the previous analysis to J classes. In this case, the Bayes formula gives:

$$\begin{aligned} \Pr\{i \in \mathcal{C}_j \mid X = x\} &= \Pr\{X = x \mid i \in \mathcal{C}_j\} \cdot \frac{\Pr\{i \in \mathcal{C}_j\}}{\Pr\{X = x\}} \\ &= c \cdot f_j(x) \cdot \pi_j \end{aligned}$$

where $c = 1/\Pr\{X = x\}$ is a normalization constant that does not depend on j. We note $S_j(x) = \ln \Pr\{i \in \mathcal{C}_j \mid X = x\}$ the discriminant score function for the j^{th} class. We have:

$$S_j(x) = \ln c + \ln f_j(x) + \ln \pi_j$$

If we again assume that $X \mid i \in \mathcal{C}_j \sim \mathcal{N}(\mu_j, \Sigma_j)$, we obtain:

$$\begin{aligned} S_j(x) &= \ln c' + \ln \pi_j - \frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (x - \mu_j)^\top \Sigma_j^{-1} (x - \mu_j) \\ &\propto \ln \pi_j - \frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (x - \mu_j)^\top \Sigma_j^{-1} (x - \mu_j) \end{aligned} \tag{15.10}$$

where $\ln c' = \ln c - \frac{K}{2} \ln 2\pi$. Given an input x, we calculate the scores $S_j(x)$ for $j = 1, \ldots, J$ and we choose the label $j^\star$ with the highest score value. As in the two-class case, we can assume an homoscedastic model ($\Sigma_j = \Sigma$), implying that the discriminant score function becomes:

$$\begin{aligned} S_j(x) &= \ln c'' + \ln \pi_j - \frac{1}{2}(x - \mu_j)^\top \Sigma_j^{-1}(x - \mu_j) \\ &\propto \ln \pi_j + \mu_j^\top \Sigma^{-1} x - \frac{1}{2}\mu_j^\top \Sigma^{-1} \mu_j \end{aligned} \tag{15.11}$$

where $\ln c'' = \ln c' - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} x^\top \Sigma^{-1} x$. Equation (15.11) defines the LDA score function, whereas Equation (15.10) defines the QDA score function.

Remark 185 *In practice, the parameters π_j, μ_j and Σ_j are unknown. We replace them by the corresponding estimates $\hat{\pi}_j$, $\hat{\mu}_j$ and $\hat{\Sigma}_j$. For the linear discriminant analysis, $\hat{\Sigma}$ is estimated by pooling all the classes.*

Example 169 *We consider the classification problem of* 33 *observations with two explanatory variables X_1 and X_2, and three classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$. The data are reported in Table 15.12.*

TABLE 15.12: Data of the classification problem

i	$\mathcal{C}_j$	X_1	X_2	i	$\mathcal{C}_j$	X_1	X_2	i	$\mathcal{C}_j$	X_1	X_2
1	1	1.03	2.85	12	2	3.70	5.08	23	3	3.55	0.58
2	1	0.20	3.30	13	2	2.81	1.99	24	3	3.86	1.83
3	1	1.69	3.73	14	2	3.66	2.61	25	3	5.39	0.47
4	1	0.98	3.52	15	2	5.63	4.19	26	3	3.15	−0.18
5	1	0.98	5.15	16	2	3.35	3.64	27	3	4.93	1.91
6	1	3.47	6.56	17	2	2.97	3.55	28	3	3.87	2.61
7	1	3.94	4.68	18	2	3.16	2.92	29	3	4.09	1.43
8	1	1.55	5.99	19	3	3.00	0.98	30	3	3.80	2.11
9	1	1.15	3.60	20	3	3.09	1.99	31	3	2.79	2.10
10	2	1.20	2.27	21	3	5.45	0.60	32	3	4.49	2.71
11	2	3.66	5.49	22	3	3.59	−0.46	33	3	3.51	1.82

The first step is to estimate the parameters π_j, μ_j and Σ_j, whose values[28] are reported in Table 15.13. The second step consists in calculating the score function $S_j(x)$ for each class j using Equations (15.10) and (15.11). Results are given in Table 15.14. Besides the QDA and LDA methods, we have also considered a third approach LDA^2, which corresponds to a linear discriminant analysis by including the squared values of variables. This means that the explanatory variables are X_1, X_2, X_1^2 and X_2^2 in LDA^2. By including polynomials, the LDA^2 method is more convex than the original LDA method, and can be seen as an approximation of the QDA method.

If we consider the first observation, the maximum score is reached for the first class (-2.28 for QDA, 0.21 for LDA and 6.93 for LDA^2). If we consider the 14$^{\text{th}}$ observation,

[28] For the LDA method, we have:

$$\hat{\Sigma} = \begin{pmatrix} 1.91355 & -0.71720 \\ -0.71720 & 3.01577 \end{pmatrix}$$

TABLE 15.13: Parameter estimation of the discriminant analysis

Class	$\mathcal{C}_1$		$\mathcal{C}_2$		$\mathcal{C}_3$	
$\hat{\pi}_j$	0.273		0.273		0.455	
$\hat{\mu}_j$	1.666	4.376	3.349	3.527	3.904	1.367
$\hat{\Sigma}_j$	1.525	0.929	1.326	0.752	0.694	−0.031
	0.929	1.663	0.752	1.484	−0.031	0.960

QDA and LDA predict the third class, whereas LDA^2 predicts the second class, which is the true value. In Figure 15.17, we have reported the class assignment performed by the three approaches, and we have indicated the bad class predictions by a circle. In order to understand these results, we have also calculated the decision regions in Figure 15.18. According to QDA, the decision boundary is almost linear between $\mathcal{C}_1$ and $\mathcal{C}_2$, whereas it is quadratic between $\mathcal{C}_2$ and $\mathcal{C}_3$. LDA produces linear decision boundaries, but the decision surface for $\mathcal{C}_1$ has changed. Finally, LDA^2 can produce complex decision surfaces, even more complex than those produced by QDA.

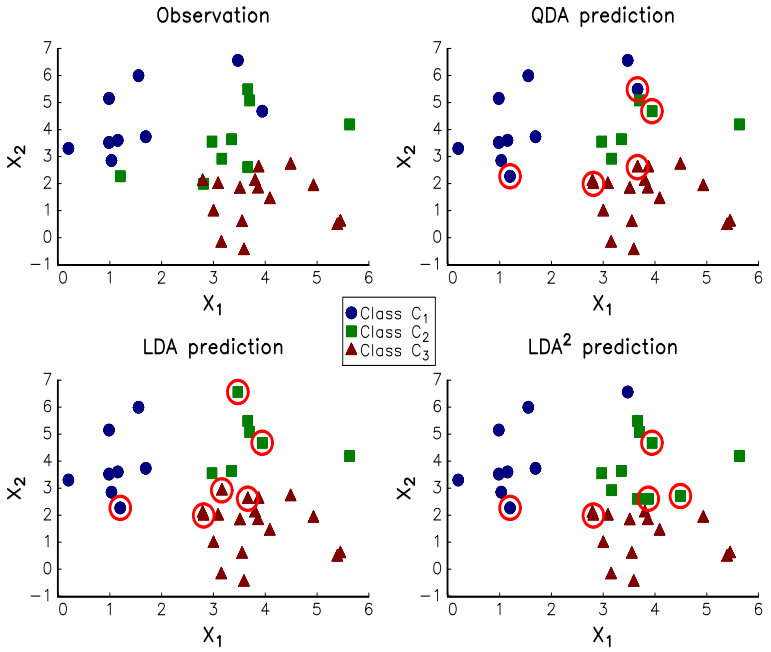


FIGURE 15.17: Comparing QDA, LDA and LDA^2 predictions

Class separation maximization In the following, we show that the linear discriminant analysis is equivalent to maximize class separability and is also related to the principal component analysis. We note $x_i = (x_{i,1}, \ldots, x_{i,K})$ the $K \times 1$ vector of exogenous variables X for the i^{th} observation. The mean vector and the variance (or scatter) matrix of Class $\mathcal{C}_j$ is equal to:

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} x_i$$

TABLE 15.14: Computation of the discriminant scores $S_j(x)$

i	QDA			LDA			LDA^2		
	$S_1(x)$	$S_2(x)$	$S_3(x)$	$S_1(x)$	$S_2(x)$	$S_3(x)$	$S_1(x)$	$S_2(x)$	$S_3(x)$
1	−2.28	−3.69	−7.49	0.21	−0.96	−0.79	6.93	5.60	5.76
2	−2.28	−6.36	−12.10	−0.26	−2.17	−2.34	1.38	−2.13	−1.89
3	−1.76	−3.13	−6.79	2.84	2.16	1.71	12.13	12.01	11.38
4	−1.80	−4.43	−8.88	1.35	0.09	−0.22	7.73	6.20	5.93
5	−2.36	−7.75	−13.70	4.32	2.93	1.45	8.12	5.54	4.76
6	−3.16	−5.63	−14.68	10.75	11.36	8.95	14.82	13.99	12.96
7	−3.79	−1.92	−6.32	8.06	9.22	8.15	17.36	19.03	17.89
8	−2.85	−8.43	−15.23	6.73	5.76	3.70	10.47	8.09	7.15
9	−1.74	−4.12	−8.37	1.76	0.64	0.27	8.94	7.77	7.39
10	−3.14	−3.21	−6.17	−0.58	−1.56	−0.98	6.59	5.55	6.15
11	−2.87	−3.01	−9.45	9.10	9.96	8.31	16.89	17.65	16.42
12	−3.04	−2.38	−7.77	8.42	9.34	7.98	17.28	18.50	17.28
13	−6.32	−2.29	−1.62	1.41	1.82	2.64	12.48	13.94	14.46
14	−6.91	−2.07	−1.42	3.86	4.94	5.34	15.15	17.41	17.34
15	−9.79	−3.62	−7.12	9.79	12.43	11.75	12.58	14.01	13.50
16	−3.90	−1.47	−3.44	5.25	5.99	5.65	16.84	18.82	18.03
17	−3.31	−1.55	−3.61	4.50	4.92	4.63	16.25	17.95	17.21
18	−4.84	−1.60	−2.19	3.65	4.28	4.45	15.51	17.48	17.14
19	−10.21	−4.12	−1.27	−0.13	0.52	2.06	8.98	9.99	11.70
20	−7.05	−2.41	−1.24	1.85	2.50	3.32	12.99	14.72	15.22
21	−23.11	−11.16	−2.56	2.98	5.75	7.61	3.79	4.57	7.26
22	−19.22	−9.53	−2.42	−1.84	−0.57	2.01	1.81	1.53	5.51
23	−13.86	−5.92	−1.01	−0.01	1.15	2.98	7.65	8.67	10.95
24	−10.01	−3.43	−0.70	2.75	4.07	5.02	12.84	14.95	15.65
25	−23.48	−11.44	−2.54	2.65	5.38	7.33	3.40	4.09	6.95
26	−15.87	−7.59	−2.30	−2.01	−1.14	1.23	3.19	3.02	6.50
27	−14.09	−5.40	−1.52	4.56	6.78	7.70	11.17	13.24	14.08
28	−7.55	−2.27	−1.39	4.18	5.45	5.85	15.10	17.44	17.40
29	−12.40	−4.67	−0.61	2.38	3.92	5.17	11.21	13.14	14.33
30	−8.85	−2.87	−0.88	3.17	4.41	5.17	13.77	15.97	16.37
31	−5.97	−2.17	−1.72	1.58	1.97	2.70	12.78	14.26	14.67
32	−9.40	−2.97	−1.81	5.33	7.11	7.46	14.55	16.95	16.93
33	−8.84	−3.01	−0.80	2.19	3.21	4.16	12.82	14.77	15.45

and[29]:

$$\mathbf{S}_j = n\hat{\Sigma}_j = \sum_{i \in \mathcal{C}_j} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^\top$$

where n_j is the number of observations in the j^{th} class. If consider the total population, we also have:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

[29]The variance matrix is equal to the unscaled covariance matrix and is also called the scatter matrix.

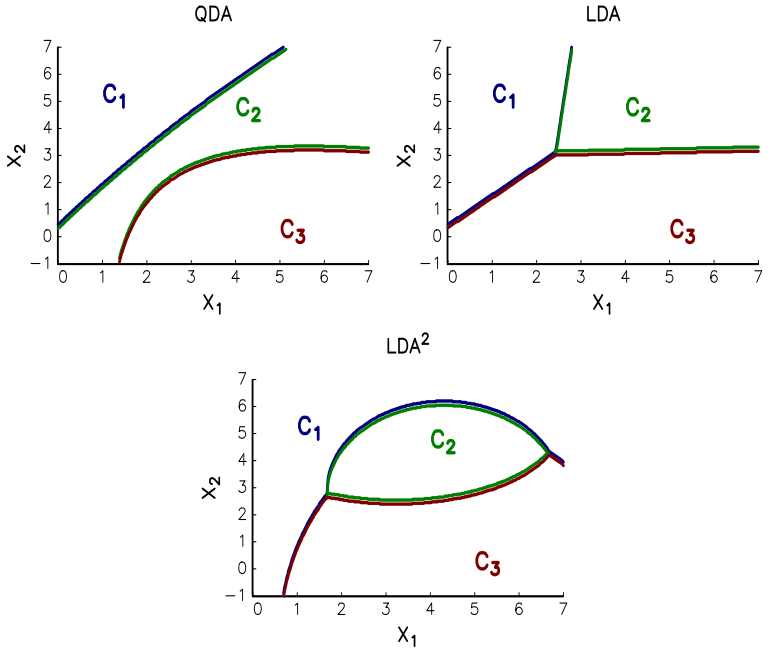


FIGURE 15.18: QDA, LDA and LDA^2 decision regions

and:

$$\mathbf{S} = n\hat{\Sigma} = \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^\top$$

We notice that:

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{J} n_j \hat{\mu}_j$$

We define the between-class variance matrix as:

$$\mathbf{S}_B = \sum_{j=1}^{J} n_j (\hat{\mu}_j - \hat{\mu})(\hat{\mu}_j - \hat{\mu})^\top$$

and the within-class variance matrix as:

$$\mathbf{S}_W = \sum_{j=1}^{J} \mathbf{S}_j$$

We can show that the total variance matrix can be decomposed into the sum of the within-class and between-class variance matrices[30]:

$$\mathbf{S} = \mathbf{S}_W + \mathbf{S}_B$$

The discriminant analysis defined by Fisher (1936) consists in finding the discriminant linear combination $\beta^\top X$ that has the maximum between-class variance relative to the within-class variance: $\beta^\star = \arg\max J(\beta)$ where $J(\beta)$ is the Fisher criterion:

$$J(\beta) = \frac{\beta^\top \mathbf{S}_B \beta}{\beta^\top \mathbf{S}_W \beta}$$

[30] See Exercise 15.4.5 on page 1024.

Since the objective function is invariant if we rescale the vector $\beta - J(\beta') = J(\beta)$ if $\beta' = c\beta$, we can impose that $\beta^\top \mathbf{S}_W \beta = 1$. It follows that:

$$\begin{aligned} \hat{\beta} &= \arg\max \beta^\top \mathbf{S}_B \beta \\ \text{s.t.} &\quad \beta^\top \mathbf{S}_W \beta = 1 \end{aligned} \tag{15.12}$$

The Lagrange function is:

$$\mathcal{L}(\beta; \lambda) = \beta^\top \mathbf{S}_B \beta - \lambda \left(\beta^\top \mathbf{S}_W \beta - 1\right)$$

We deduce that the first-order condition is equal to:

$$\frac{\partial \mathcal{L}(\beta; \lambda)}{\partial \beta^\top} = 2\mathbf{S}_B \beta - 2\lambda \mathbf{S}_W \beta = \mathbf{0} \tag{15.13}$$

It is remarkable that we obtain a generalized eigenvalue problem[31] $\mathbf{S}_B \beta = \lambda \mathbf{S}_W \beta$ or equivalently:

$$\mathbf{S}_W^{-1} \mathbf{S}_B \beta = \lambda \beta \tag{15.14}$$

Even if $\mathbf{S}_W$ and $\mathbf{S}_B$ are two symmetric matrices, it is not necessarily the case for the product $\mathbf{S}_W^{-1} \mathbf{S}_B$. Using the eigendecomposition $\mathbf{S}_B = V \Lambda V^\top$, we have $\mathbf{S}_B^{1/2} = V \Lambda^{1/2} V^\top$. With the parametrization $\alpha = \mathbf{S}_B^{1/2} \beta$, Equation (15.14) becomes:

$$\mathbf{S}_B^{1/2} \mathbf{S}_W^{-1} \mathbf{S}_B^{1/2} \alpha = \lambda \alpha \tag{15.15}$$

because $\beta = \mathbf{S}_B^{-1/2} \alpha$. Equation (15.15) defines a right regular eigenvalue problem. Let λ_k and v_k be the k^{th} eigenvalue and eigenvector of the symmetric matrix $\mathbf{S}_B^{1/2} \mathbf{S}_W^{-1} \mathbf{S}_B^{1/2}$. It is obvious that the optimal solution $\alpha^\star$ is the first eigenvector v_1 corresponding to the largest eigenvalue λ_1. We conclude that the estimator is $\hat{\beta} = \mathbf{S}_B^{-1/2} v_1$ and the discriminant linear relationship is $Y^c = v_1^\top \mathbf{S}_B^{-1/2} X$. Moreover, we have[32]:

$$\lambda_1 = J\left(\hat{\beta}\right) = \frac{\hat{\beta}^\top \mathbf{S}_B \hat{\beta}}{\hat{\beta}^\top \mathbf{S}_W \hat{\beta}}$$

In Exercise 15.4.5 on page 1024, we show that the Fisher discriminant analysis is equivalent to the linear discriminant analysis in the case of two classes. This result can be extended to multiple classes and explains why this approach is also called Fisher linear discriminant analysis.

Example 170 *We consider a problem with two classes* $\mathcal{C}_1$ *and* $\mathcal{C}_2$*, and two explanatory variables* (X_1, X_2)*. Class* $\mathcal{C}_1$ *is composed of 7 observations:* $(1,2)$*,* $(1,4)$*,* $(3,6)$*,* $(3,3)$*,* $(4,2)$*,* $(5,6)$*,* $(5,5)$*, whereas class* $\mathcal{C}_2$ *is composed of 6 observations:* $(1,0)$*,* $(2,1)$*,* $(4,1)$*,* $(3,2)$*,* $(6,4)$ *and* $(6,5)$*.*

In Figure 15.19, we have reported these 13 observations in the plane (x_1, x_2). The computation of the first generalized eigenvector gives $\beta = (0.7547, -0.9361)$. We deduce that the slope of the optimal line direction is $\beta_1/\beta_2 = -0.8062$. Computing the Fisher score $s_i = \beta^\top x_i$ for the i^{th} observation is then equivalent to perform the orthogonal projection of

[31] See Appendix A.1.1.2 on page 1034 for the definition of the generalized eigendecomposition.
[32] Thanks to Equation (15.13), we have $\mathbf{S}_B \beta = \lambda \mathbf{S}_W \beta$ and $\beta^\top \mathbf{S}_B \beta = \lambda \beta^\top \mathbf{S}_W \beta$.

the points on this optimal line (Bishop, 2006). Concerning the assignment decision, we can consider the midpoint rule:

$$\begin{cases} s_i < \bar{\mu} \Rightarrow i \in \mathcal{C}_1 \\ s_i > \bar{\mu} \Rightarrow i \in \mathcal{C}_2 \end{cases}$$

where $\bar{\mu} = (\bar{\mu}_1 + \bar{\mu}_2)/2$, $\bar{\mu}_1 = \beta^\top \hat{\mu}_1$ and $\bar{\mu}_2 = \beta^\top \hat{\mu}_2$. However, this rule is not always optimal because it does not depend on the variance $\bar{s}_1^2$ and $\bar{s}_2^2$ of each class. In Figure 15.20, we have reported the Gaussian density of the scores for the two classes. Since we observe that the first class has a larger variance, the previous rule is not adapted. This is why we can use the tools presented in Section 15.3 in order to calibrate the optimal decision rule.

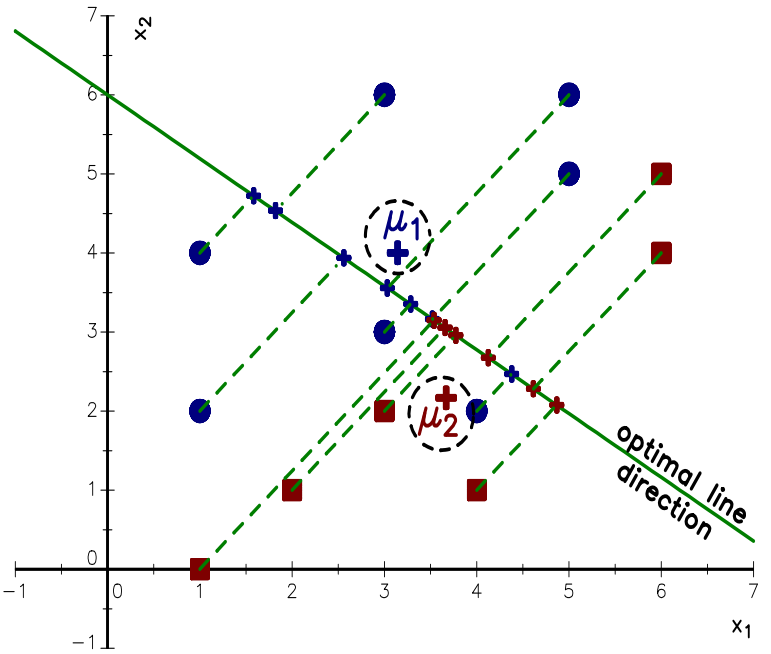


FIGURE 15.19: Linear projection and the Fisher solution

Remark 186 *$v_1^\top \mathbf{S}_B^{-1/2} X$ is called the first canonical or discriminant variable (Hastie et al., 2009) and we denote it by $Y_{(1)}^c$. The previous analysis can be used to find the second canonical variable $Y_{(2)}^c = \beta_{(2)}^\top X$ that is not correlated to $Y_{(1)}^c$ such that $J\left(\beta_{(2)}\right)$ is maximum. The solution is $\hat{\beta}_{(2)} = \mathbf{S}_B^{-1/2} v_2$ where v_2 is the eigenvector associated to the second largest eigenvalue λ_2. This method can be extended to the general problem of finding the k^{th} canonical variable $Y_{(k)}^c = \beta_{(k)}^\top X$ that is not correlated to $\left(Y_{(1)}^c, \ldots, Y_{(k-1)}^c\right)$ such that $J\left(\beta_{(k)}\right)$ is maximum. Again, we can show that the solution is $\hat{\beta}_{(k)} = \mathbf{S}_B^{-1/2} v_k$ where v_k is the eigenvector associated to the k^{th} largest eigenvalue λ_k. The computation of the K linear relationships $Y_{(k)}^c = \beta_{(k)}^\top X$ is called the multiple discriminant analysis (MDA). MDA can be seen as a generalized PCA method by taking into account a categorical response variable. Indeed, PCA performs an eigendecomposition of $\mathbf{S}$ (or $\hat{\Sigma}$) whereas MDA performs an eigendecomposition of $\mathbf{S}_W^{-1}\mathbf{S}_B$.*

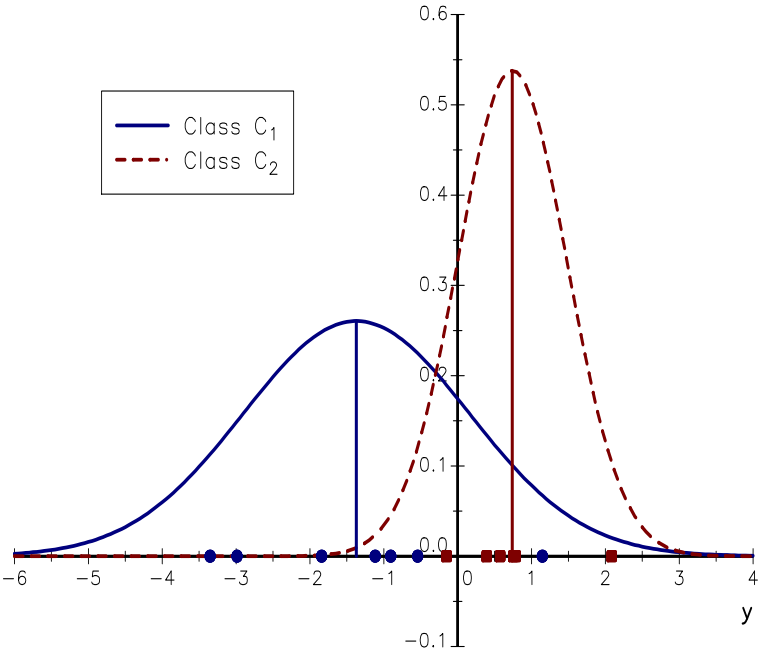


FIGURE 15.20: Class separation and the cut-off criterion

15.2.2.2 Binary choice models

The underlying idea of such models is to estimate the probability of a binary response based on several explanatory variables. They have been developed in several fields of research (biology, epidemiology, economy, etc.). In statistics, the two seminal papers are again written by Fisher (1935) and Cox (1958). Since these publications, these models have been extended and now represent a major field of study in statistics and econometrics[33].

General framework In this section, we assume that Y can take two values 0 and 1. We consider models that link the outcome to a set of factors X:

$$\Pr\{Y = 1 \mid X = x\} = \mathbf{F}\left(x^\top \beta\right)$$

By construction, $\mathbf{F}$ must be a cumulative distribution function in order to ensure that $\mathbf{F}(z) \in [0, 1]$. We also assume that the model is symmetric, implying that $\mathbf{F}(z) + \mathbf{F}(-z) = 1$. Given a sample $\{(x_i, y_i), i = 1, \ldots, n\}$, the log-likelihood function is equal to:

$$\ell(\theta) = \sum_{i=1}^{n} \ln \Pr\{Y_i = y_i\}$$

where y_i takes the values 0 or 1. We have:

$$\Pr\{Y_i = y_i\} = p_i^{y_i} \cdot (1 - p_i)^{1 - y_i}$$

[33]The materials presented below is based on surveys by Amemiya (1981, 1985) and McFadden (1984).

where $p_i = \Pr\{Y_i = 1 \mid X_i = x_i\}$. We deduce that:

$$\begin{aligned}\boldsymbol{\ell}(\theta) &= \sum_{i=1}^{n} y_i \ln p_i + (1-y_i)\ln(1-p_i) \\ &= \sum_{i=1}^{n} y_i \ln \mathbf{F}\left(x_i^\top \beta\right) + (1-y_i)\ln\left(1-\mathbf{F}\left(x_i^\top \beta\right)\right)\end{aligned}$$

We notice that the vector θ includes only the parameters β. By noting $f(z)$ the probability density function, it follows that the associated score vector and Hessian matrix of the log-likelihood function are:

$$\begin{aligned}\mathcal{S}(\beta) &= \frac{\partial \boldsymbol{\ell}(\beta)}{\partial \beta} \\ &= \sum_{i=1}^{n}\left(y_i \frac{f\left(x_i^\top \beta\right)}{\mathbf{F}\left(x_i^\top \beta\right)} - (1-y_i)\frac{f\left(x_i^\top \beta\right)}{1-\mathbf{F}\left(x_i^\top \beta\right)}\right) x_i \\ &= \sum_{i=1}^{n} \frac{f\left(x_i^\top \beta\right)}{\mathbf{F}\left(x_i^\top \beta\right)\mathbf{F}\left(-x_i^\top \beta\right)}\left(y_i - \mathbf{F}\left(x_i^\top \beta\right)\right) x_i\end{aligned}$$

and:

$$\begin{aligned}H(\beta) &= \frac{\partial^2 \boldsymbol{\ell}(\beta)}{\partial \beta\, \partial \beta^\top} \\ &= -\sum_{i=1}^{n} H_i \cdot \left(x_i x_i^\top\right)\end{aligned}$$

where:

$$\begin{aligned}H_i &= \frac{f\left(x_i^\top \beta\right)^2}{\mathbf{F}\left(x_i^\top \beta\right)\mathbf{F}\left(-x_i^\top \beta\right)} - \left(y_i - \mathbf{F}\left(x_i^\top \beta\right)\right)\cdot \\ &\quad \left(\frac{f'\left(x_i^\top \beta\right)}{\mathbf{F}\left(x_i^\top \beta\right)\mathbf{F}\left(-x_i^\top \beta\right)} - \frac{f\left(x_i^\top \beta\right)^2\left(1-2\mathbf{F}\left(x_i^\top \beta\right)\right)}{\mathbf{F}\left(x_i^\top \beta\right)^2\mathbf{F}\left(-x_i^\top \beta\right)^2}\right)\end{aligned}$$

Once $\hat{\beta}$ is estimated by the method of maximum likelihood, we can calculated the predicted probability for the i^{th} observation:

$$\hat{p}_i = \mathbf{F}\left(x_i^\top \hat{\beta}\right)$$

Like a linear regression model, we can define the residual as the difference between the observation y_i and the predicted value $\hat{p}_i$. We can also exploit the property that the conditional distribution of Y_i is a Bernoulli distribution $\mathcal{B}(p_i)$. This is why it is better to use the standardized (or Pearson) residuals:

$$\hat{u}_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i\left(1-\hat{p}_i\right)}}$$

These residuals are related to the Pearson's chi-squared statistic:

$$\begin{aligned}\chi^2_{\text{Pearson}} &= \sum_{i=1}^{n} \hat{u}_i^2 \\ &= \sum_{i=1}^{n} \frac{\left(y_i - \hat{p}_i\right)^2}{\hat{p}_i\left(1-\hat{p}_i\right)}\end{aligned}$$

This statistic may used to measure the goodness-of-fit of the model. Under the assumption $\mathcal{H}_0$ that there is no lack-of-fit, we have $\chi^2_{\text{Pearson}} \sim \chi^2_{n-K}$ where K is the number of exogenous variables. Another goodness-of-fit statistic is the likelihood ratio. For the '*saturated*' model, the estimated probability $\hat{p}_i$ is exactly equal to y_i. We deduce that the likelihood ratio is equal to:

$$\begin{aligned} -2\ln\Lambda &= 2\sum_{i=1}^{n} y_i \ln y_i + (1-y_i)\ln(1-y_i) - \\ &\quad 2\sum_{i=1}^{n} y_i \ln \hat{p}_i + (1-y_i)\ln(1-\hat{p}_i) \\ &= 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\hat{p}_i}\right) + (1-y_i)\ln\left(\frac{1-y_i}{1-\hat{p}_i}\right) \end{aligned}$$

In binomial choice models, $D = -2\ln\Lambda$ is also called the deviance and we have $D \sim \chi^2_{n-K}$. In a perfect fit $\hat{p}_i = y_i$, the likelihood ratio is exactly equal to zero. The forecasting procedure consists of estimating the probability $\hat{p} = \mathbf{F}\left(x^\top \hat{\beta}\right)$ for a given set of variables x and to use the following decision criterion:

$$Y = 1 \Leftrightarrow \hat{p} \geq \frac{1}{2}$$

Remark 187 *It could also be interesting to compute the marginal effects. We have:*

$$\mathbb{E}[Y \mid X = x] = \mathbf{F}\left(x^\top \hat{\beta}\right)$$

and:

$$\frac{\partial\, \mathbb{E}[Y \mid X = x]}{\partial\, x} = f\left(x^\top \hat{\beta}\right) \cdot \hat{\beta}$$

The marginal effects depend on the vector x and are then not easy to understand. This is why we generally compute them by using the mean of the regressors or averaging them across all the observations of the sample.

Logit analysis The logit model uses the following cumulative distribution function:

$$\mathbf{F}(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{e^z+1}$$

The probability density function is then equal to:

$$f(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$

We verify the property $\mathbf{F}(z) + \mathbf{F}(-z) = 1$. The log-likelihood function is equal to:

$$\begin{aligned} \boldsymbol{\ell}(\beta) &= \sum_{i=1}^{n} (1-y_i)\ln\left(1-\mathbf{F}\left(x_i^\top\beta\right)\right) + y_i \ln \mathbf{F}\left(x_i^\top\beta\right) \\ &= \sum_{i=1}^{n} (1-y_i)\ln\left(\frac{e^{-x_i^\top\beta}}{1+e^{-x_i^\top\beta}}\right) - y_i \ln\left(1+e^{-x_i^\top\beta}\right) \\ &= -\sum_{i=1}^{n} \ln\left(1+e^{-x_i^\top\beta}\right) + (1-y_i)\left(x_i^\top\beta\right) \end{aligned}$$

We also have[34]:

$$\mathcal{S}(\beta) = \sum_{i=1}^{n} \left(y_i - \mathbf{F}\left(x_i^\top \beta\right)\right) x_i$$

and[35]:

$$H(\beta) = -\sum_{i=1}^{n} f\left(x_i^\top \beta\right) \cdot \left(x_i x_i^\top\right)$$

Probit analysis The probit model assumes that $\mathbf{F}(z)$ is the Gaussian distribution. The log-likelihood function is then:

$$\boldsymbol{\ell}(\beta) = \sum_{i=1}^{n} (1 - y_i) \ln\left(1 - \Phi\left(x_i^\top \beta\right)\right) + y_i \ln \Phi\left(x_i^\top \beta\right)$$

The probit model can be seen as a latent variable model. Let us consider the linear model $Y^\star = \beta^\top X + U$ where $U \sim \mathcal{N}\left(0, \sigma^2\right)$. We assume that we do not observe $Y^\star$ but $Y = g(Y^\star)$. For example, if $g(z) = \mathbb{1}\{z > 0\}$, we obtain:

$$\begin{aligned}
\Pr\{Y = 1 \mid X = x\} &= \Pr\left\{\beta^\top X + U > 0 \mid X = x\right\} \\
&= \Phi\left(\frac{\beta^\top x}{\sigma}\right)
\end{aligned}$$

We notice that only the ratio β/σ is identifiable. Since we can set $\sigma = 1$, we obtain the probit model.

Regularization Let $\boldsymbol{\ell}(\theta)$ be the log-likelihood function. The regularized log-likelihood function is equal to:

$$\boldsymbol{\ell}(\theta; \lambda) = \boldsymbol{\ell}(\theta) - \frac{\lambda}{p} \|\theta\|_p^p$$

The case $p = 1$ is equivalent to consider a lasso penalization, whereas $p = 2$ corresponds to the ridge regularization. The optimal value $\theta^\star$ is obtained by maximizing the regularized log-likelihood function:

$$\theta^\star(\lambda) = \arg\max \boldsymbol{\ell}(\theta; \lambda)$$

In this problem, we consider λ as an hyperparameter, meaning that λ is not directly estimated by maximizing the penalized log-likelihood function with respect to $(\theta; \lambda)$. For instance, in the case of the lasso regularization, λ can be calibrated in order to obtain a sparse model or using cross-validation techniques.

[34] We use the property $f(z) = \mathbf{F}(z)(1 - \mathbf{F}(z))$, implying that:

$$\frac{f(z)}{\mathbf{F}(z)\,\mathbf{F}(-z)} = \frac{f(z)}{\mathbf{F}(z)(1 - \mathbf{F}(z))} = 1$$

[35] We use the property $f'(z) = -f(z)\,\mathbf{F}(z)\left(1 - e^{-z}\right)$, implying that:

$$\frac{f'(z)}{f(z)} - (1 - 2\mathbf{F}(z)) = 0$$

Extension to multinomial logistic regression We assume that Y can take J labels $(\mathfrak{L}_1, \ldots, \mathfrak{L}_J)$ or belongs to J disjoint classes $(\mathcal{C}_1, \ldots, \mathcal{C}_J)$. We define the conditional probability as follows:

$$\begin{aligned} p_j(x) &= \Pr\{Y = \mathfrak{L}_j \mid X = x\} \\ &= \Pr\{Y \in \mathcal{C}_j \mid X = x\} \\ &= \frac{e^{\beta_j^\top x}}{1 + \sum_{j=1}^{J-1} e^{\beta_j^\top x}} \end{aligned}$$

for $j = 1, \ldots, J-1$. The probability of the last label is then equal to:

$$\begin{aligned} p_J(x) &= 1 - \sum_{j=1}^{J-1} p_j(x) \\ &= \frac{1}{1 + \sum_{j=1}^{J-1} e^{\beta_j^\top x}} \end{aligned}$$

We verify that $0 \leq p_j(x) \leq 1$ for all $j = 1, \ldots, J$. The log-likelihood function becomes:

$$\boldsymbol{\ell}(\theta) = \sum_{i=1}^{n} \ln \left(\prod_{j=1}^{J} p_j(x_i)^{i \in \mathcal{C}_j} \right)$$

where θ is the vector of parameters $(\beta_1, \ldots, \beta_{J-1})$.

The multinomial logistic model can be formulated as a log-linear model. We note:

$$\ln p_j(x) = \beta_0 + \beta_j^\top x$$

Since we have $\sum_{j=1}^{J} p_j(x) = 1$, we deduce that the constant β_0 is given by:

$$\sum_{j=1}^{J} e^{\beta_0 + \beta_j^\top x} = 1 \Leftrightarrow \beta_0 = \ln \frac{1}{\sum_{j=1}^{J} e^{\beta_j^\top x}}$$

It follows that:

$$p_j(x) = \frac{e^{\beta_j^\top x}}{\sum_{j=1}^{J} e^{\beta_j^\top x}}$$

This function is known as the softmax function and plays an important role in neural networks. We also notice that the model is overidentified because the sum of probabilities is equal to 1. However, if we use the parametrization $\breve{\beta}_j = \beta_j - \beta_J$, we obtain the previous model[36], which is just identified.

[36] Indeed, we have:

$$\begin{aligned} p_j(x) &= \frac{e^{\breve{\beta}_j^\top x} e^{-\beta_J^\top x}}{e^{-\beta_J^\top x} \sum_{j=1}^{J} e^{\breve{\beta}_j^\top x}} \\ &= \frac{e^{\breve{\beta}_j^\top x}}{1 + \sum_{j=1}^{J-1} e^{\breve{\beta}_j^\top x}} \end{aligned}$$

because $e^{\breve{\beta}_J^\top x} = e^{(\beta_J - \beta_J)^\top x} = 1$.

15.2.3 Non-parametric supervised methods

We have named this section '*non-parametric supervised methods*' in order to group some approaches that share some of the same characteristics. First, even if some of them are parametric, these models are highly non-linear, meaning that it is extremely difficult to interpret these models. In this case, '*forecasting*' is the main motivation and is more important than '*modeling*'. Second, it would be illusory to consider or to do statistical inference. Most of the time, it is impossible to calculate the variance of the parameters and the associated t-statistics. Therefore, the term '*model calibration*' is more appropriate than the term '*model estimation*'. Finally, the number of parameters or unknowns can be large.

If we consider the linear regression model, we have $Y = \beta^\top X + u$ where (Y, X) forms a random vector. If we consider an observation i, we have $y_i = f(x_i) + u_i$ where $f(x_i) = \sum_{k=1}^K \beta_k x_{i,k}$. Let us now consider some non-linear features. We can replace the linear function by:

$$f(x_i) = \sum_{k=1}^K \beta_k \phi_k(x_i) = \beta^\top \phi(x_i)$$

For example, we can use quadratic, cubic or piecewise features. We abandon the framework of Gaussian conditional distribution, which is the basis of linear regression, and the reference to the random variables X and Y is not necessary. This means that the calibrated parameters $\left(\hat{\beta}_1, \ldots, \hat{\beta}_K\right)$ are less relevant. Only the calibrated function $\hat{f}(x)$ is important. For instance, if we use radial basis functions:

$$\phi_k(x) = \exp\left(-\frac{1}{2}\|x - c_k\|^2\right)$$

where c_k is the centering parameter, we obtain:

$$\hat{f}(x_i) = \sum_{k=1}^K \hat{\beta}_k e^{-\frac{1}{2}\|x_i - c_k\|^2}$$

Even if $\hat{f}(x)$ is a parametric function, it can be considered as a non-parametric model. Indeed, the functional form is the relevant quantity, not the parameters.

15.2.3.1 k-nearest neighbor classifier

The k-NN algorithm is one of the simplest non-parametric models. Let $\{(x_i, y_i)\}$ be the training sample of dimension n. We assume that the labels y_i can be assigned to J classes $(\mathcal{C}_1, \ldots, \mathcal{C}_J)$. The goal is to assign a label y for a given unlabeled observation x. For that, we select the k closest labeled observations in the training sample and we find the label $\hat{y}$ that appears most frequently within the k-subset. Said differently, the k-NN classifier uses the majority vote of the k closest neighbors and the classification rule depends on k, which is the hyperparameter. It is obvious that a high value of k helps to smooth the decision regions, but it increases the computational complexity. Moreover, there is a trade-off between bias and variance. If $k = 1$, we assign to x the label of the input x_i that is the closest. If $k = n$, we assign to x the most frequent label of the training sample. In the first case, we see that $\hat{y}$ is an unbiased estimator of y, but its variance is large. In the second case, the estimator is biased but it has a small variance.

The implementation of the k-NN algorithm requires defining the distance between the points x_i and x_j. Generally, we use the Euclidean distance, but we can consider the Minkowski distance. To find the k closest labeled observations, the simplest way is the

brute-force approach. When the number of observations n is large, we can use more efficient methods based on tree-based partition[37].

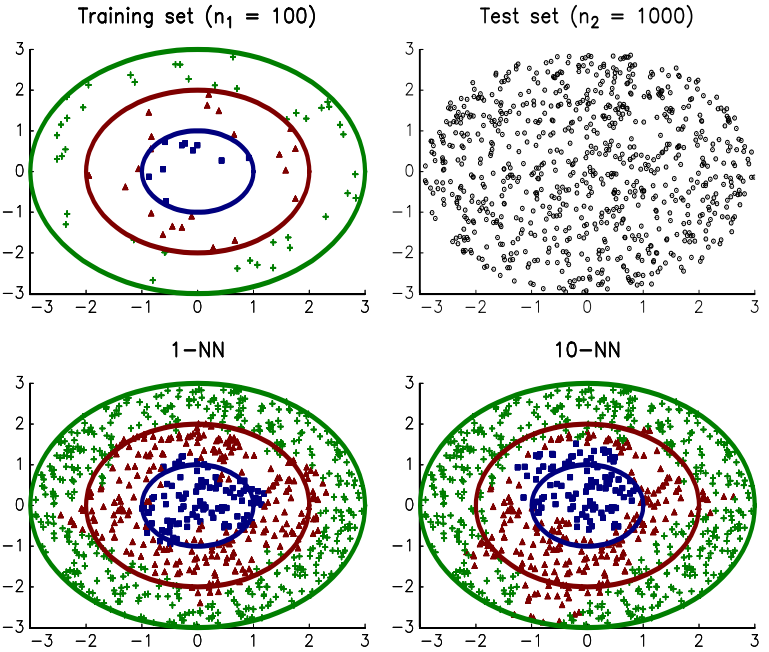


FIGURE 15.21: Illustration of the k-NN classifier

We consider the non-linearly separable classification problem, where the classes are distributed in rings around the point (0,0):

$$\mathcal{C}_j = \left\{(x_{i,1}, x_{i,2}) \in \mathbb{R}^2 : r_{j-1}^2 < x_{i,1}^2 + x_{i,2}^2 \leq r_j^2\right\}$$

where $j = \{1, 2, 3\}$ and $r_j = j$ is the radius of the ring. In the first panel in Figure 15.21, we have represented the three rings, and we have reported 100 simulated observations $(x_{i,1}, x_{i,2})$ that form the training set. In the second panel, we consider 1 000 observations. Solutions provided by 1-NN and 10-NN classifiers are given in the third and fourth panels. We notice that the 10-NN classifier is less efficient than the 1-NN classifier, because the number of closest neighbors is large compared to the number of observations in the training set.

Remark 188 *We can apply the k-NN algorithm to the regression. In this case, the predicted value $\hat{y}$ is the (weighted) average of the values y_i of the k closest neighbors*[38].

15.2.3.2 Neural networks

Neural networks as non-linear models We have seen that we can extend the linear model as follows[39]:

$$y_i = \beta_0 + \beta^\top \phi(x_i) + \varepsilon_i$$

In this case, we transform the input data $(x_{i,1}, \ldots, x_{i,K})$ into the auxiliary data $(z_{i,1}, \ldots, z_{i,K})$ where $z_{i,k} = \phi_k(x_{i,k})$. Here, the non-linearity property is introduced thanks

[37] The two most famous methods are the K-D tree and ball tree algorithms.
[38] The weight is generally inversely proportional to the distance.
[39] Here, we include a constant in the model.

to the non-linear function ϕ_k. However, there are many other ways to build a non-linear model. For instance, we can assume that:

$$y_i = \phi\left(\beta_0 + \beta^\top x_i\right) + \varepsilon_i$$

We first create the auxiliary data $z_i = \beta_0 + \beta^\top x_i$ from the inputs and then apply the non-linear function $\phi(x)$. If we use several non-linear functions, we obtain:

$$y_i = \sum_{j=1}^{J} \gamma_j \phi_j\left(\beta_0 + \beta^\top x_i\right) + \varepsilon_i$$

or:

$$\begin{aligned} y_i &= \varphi\left(\gamma_0 + \sum_{j=1}^{J} \gamma_j \phi_j\left(\beta_0 + \beta^\top x_i\right)\right) + \varepsilon_i \\ &= f(x_i) + \varepsilon_i \end{aligned}$$

The underlying idea of neural networks is to define a non-linear function $f(x)$, which is sufficiently flexible to fit complex relationships.

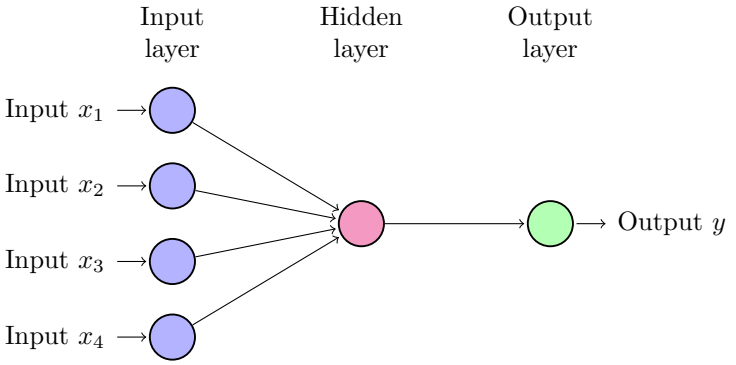


FIGURE 15.22: The perceptron

Neural networks as a mathematical representation of biological systems The term '*neural network*' makes reference to biological systems, in particular the brain. For instance, Rosenblatt (1958) proposed a self-organizing and adaptive model called the perceptron. It is no coincidence that the title of this publication is "*The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*". We have represented the perceptron in Figure 15.22. The input data are combined in order to produce an hidden variable $z = \sum_{k=1}^{K} \beta_k x_k$. Then, we apply the function $f(z)$ in order to obtain the output y:

$$y = f(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

In the context of neural networks, the function $f(z)$ is called the activation function and z is the hidden unit. If we generalize the perceptron by considering different hidden units, we obtain the artificial neural network described in Figure 15.23. In this example, we have four input units, five hidden units and one output unit. This model is also called a feed-forward neural network with one hidden layer. It can be extended in two directions.

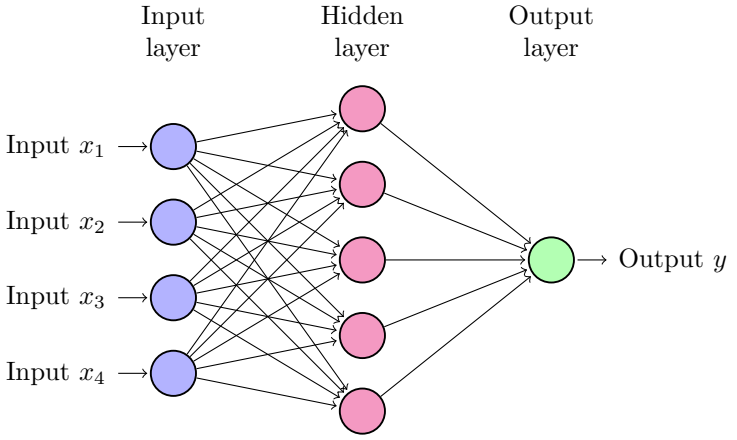


FIGURE 15.23: Feed-forward neural network with a single hidden layer

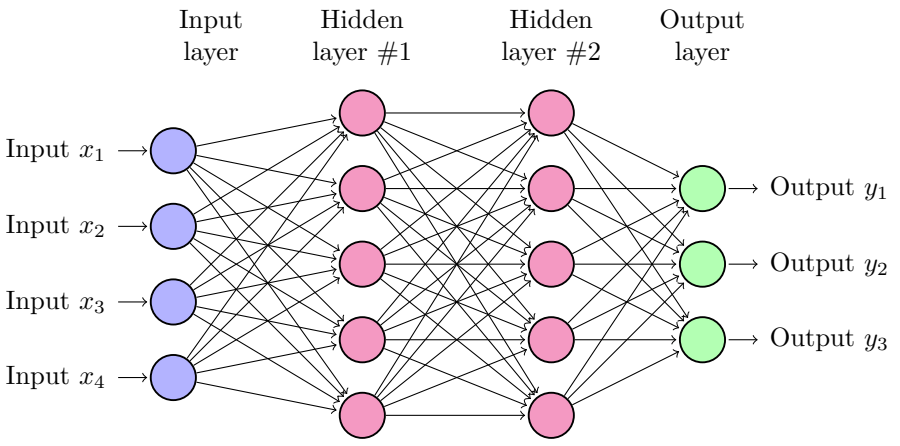


FIGURE 15.24: Feed-forward neural network with two hidden layers and three output units

First, we can consider several output units. Second, we can use several hidden layers. In this case, we speak about multi-layer neural networks (Figure 15.24). For example, deep learning refers to a neural network with a large number of hidden layers.

The term neural network does not only refer to the structure input layer – hidden layer – output layer. Activation functions generally map the resulting values into the range $[0,1]$ or $[-1,1]$ and correspond to sigmoidal functions. For example, the perceptron uses the Heaviside step function $f(z) = \mathbb{1}\{z > 0\}$, because it indicates if the neuron is activated or not. We can also use the sign function $f(z) = \text{sign}(z)$ in order to indicate a '*positive*' or '*negative*' potential. However, the most popular activation functions are continuous:

1. the logistic function is equal to:

$$f(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \tag{15.16}$$

   we have $f(z) \in [0,1]$, meaning that we can interpret $f(z)$ as a probability function; moreover, it is symmetric about 0.5;

2. the hyperbolic tangent function is defined by:

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{2z}-1}{e^{2z}+1} \tag{15.17}$$

   and we have $f(z) \in [-1,1]$;

3. the rectified linear unit (ReLU) function corresponds to:

$$f(z) = \max(0, z) \tag{15.18}$$

   and we have $f(z) \in [0,\infty)$.

Furthermore, neural networks are also characterized by the concept of learning algorithms. Neural networks can be seen as non-linear functions with some unknown parameters. The first idea is then to estimate the parameters by minimizing the residual sum of squares, meaning that neural networks are just a particulate case of non-linear least squares. However, neural networks generally use other techniques for identifying the parameters. Statistical learning implicitly refers to human brains or natural neural networks. The concept of learning is then central and shall contrast with the concept of optimization. The latter implies that there is one solution. In an artificial neural network, each node represents a neuron and each connection can be seen as a synapse. Since these connections transmit a signal from one neuron to another, the underlying idea is that they learn like in a human brain. This is why the parameters that control these connections are updated until the artificial neural network has learnt. In fact, the difference between optimization and learning is somewhat forced. Indeed, optimization also uses iterative algorithms that can be interpreted as learning algorithms. However, the learning algorithms that are used in artificial neural networks try to imitate the learning process of human brains[40]. They are also called adaptive learning rules in order to say that they are adaptive and they try to learn.

Remark 189 *According to Bishop (2006), the term neural network "has been used very broadly to cover a wide range of different models, many of which have been the subject of exaggerated claims regarding their biological plausibility". In fact, we use neural networks as non-linear regression models in the sequel.*

[40] It is particularly true for the first generation of algorithms that were discovered before 1990s.

Structure of the canonical neural network The notations used in neural networks and machine learning are generally different than those used in statistics. In this book, we have tried to use similar homogenous notations in order to make the reading easier:

- the observation (also called the example) is denoted by i;
- the input variable uses the index k and $x_{i,k}$ is the k^{th} input variable of the i^{th} observation;
- $z_{i,h}$ is the value taken by the h^{th} hidden variable and the i^{th} observation;
- for the output variables (also called the patterns), we introduce the notation $y_j(x_i)$ to name the model output taken by the j^{th} output variable and the i^{th} observation; sometimes, we use the alternative notation $\hat{y}_{i,j}$, which is more traditional in statistical inference theory.

The number of input, hidden and output variables are respectively equal n_x, n_z and n_y. The activation functions $f_{x,z}$ and $f_{z,y}$ links respectively the x's to the z's, and the z's to the y's. In order to distinguish them, $f_{z,y}$ is also called the output scaling function. We have[41]:

$$z_{i,h} = f_{x,z}(u_{i,h}) = f_{x,z}\left(\sum_{k=1}^{n_x} \beta_{h,k} x_{i,k}\right)$$

and:

$$y_j(x_i) = f_{z,y}(v_{i,j}) = f_{z,y}\left(\sum_{h=1}^{n_z} \gamma_{j,h} z_{i,h}\right)$$

where $u_{i,h}$ and $v_{i,j}$ are the intermediary variables before the activation of the functions $f_{x,z}$ and $f_{z,y}$. Finally, we have:

$$y_j(x_i) = f_{z,y}\left(\sum_{h=1}^{n_z} \gamma_{j,h} f_{x,z}\left(\sum_{k=1}^{n_x} \beta_{h,k} x_{i,k}\right)\right) \tag{15.19}$$

Figure 15.25 summarizes the structure and the notations of this neural network.

Remark 190 *Including a constant is equivalent to consider that $x_{i,1} = 1$. A variant model is to define $y_j(x_i)$ as follows:*

$$y_j(x_i) = f_{z,y}\left(\gamma_{j,0} + \sum_{h=1}^{n_z} \gamma_{j,h} f_{x,z}\left(\beta_{h,0} + \sum_{k=1}^{n_x} \beta_{h,k} x_{i,k}\right)\right) \tag{15.20}$$

In this case, we add a constant as an input variable ($\beta_{h,0}$) and a constant as a hidden variable ($\gamma_{j,0}$). Bishop (2006) shows that this model can be written as:

$$y_j(x_i) = f_{z,y}\left(\sum_{h=0}^{n_z} \gamma_{j,h} f_{x,z}\left(\sum_{k=0}^{n_x} \beta_{h,k} x_{i,k}\right)\right)$$

where $x_{i,0} = 1$. The other possibility is to have a direct link between the x's to the y's or skip-layer connections:

$$y_j(x_i) = f_{z,y}\left(\gamma_{j,0} + \sum_{h=1}^{n_z} \gamma_{j,h} f_{x,z}\left(\beta_{h,0} + \sum_{k=1}^{n_x} \beta_{h,k} x_{i,k}\right) + \sum_{k=1}^{n_x} \gamma_{j,n_z+k} x_{i,k}\right) \tag{15.21}$$

[41] Most of the time, we use the same activation function $f_{x,z}(u) = f_{z,y}(u) = f(u)$.

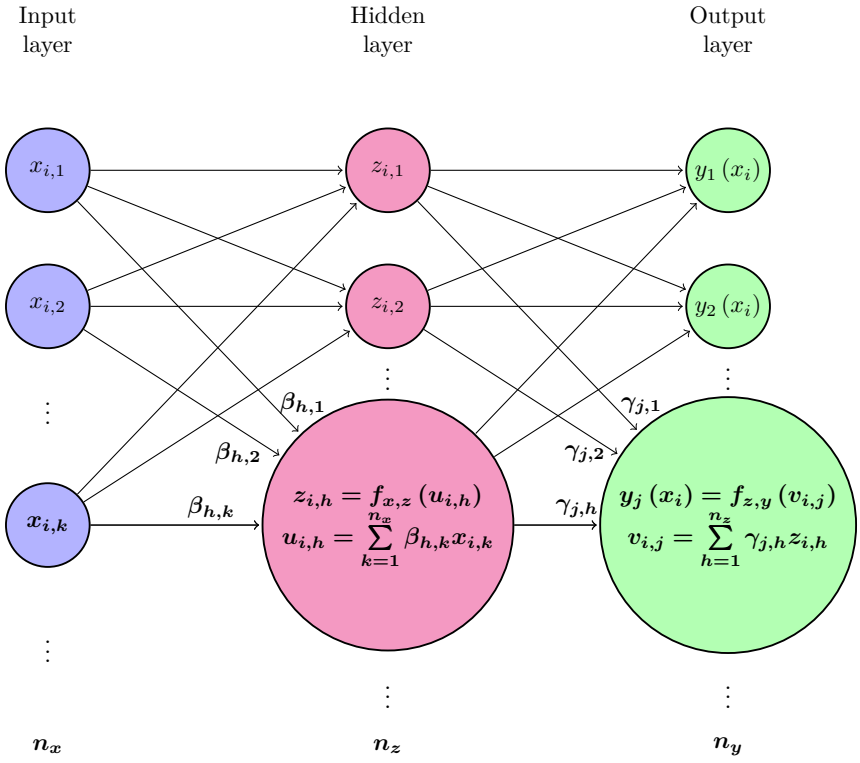


FIGURE 15.25: Canonical neural network

Loss function If we note $y_{i,j}$ the value of the output variable that is observed[42], we would like to verify:

$$y_j(x_i) = y_{i,j}$$

It follows that a natural loss function is the sum of squared errors:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n_y} \frac{1}{2} (y_j(x_i) - y_{i,j})^2 \tag{15.22}$$

where θ is the vector of parameters and n is the number of observations. Minimizing this loss function is also equivalent to maximize the log-likelihood function associated to the non-linear regression model:

$$y_{i,j} = y_j(x_i) + \varepsilon_{i,j}$$

where $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$ and $\varepsilon_{i,j} \perp \varepsilon_{i',j'}$ if $i \neq i'$ or $j \neq j'$.

The previous loss function is natural when considering a non-linear regression. In the case of binary classification ($y_i = 0$ or $y_i = 1$) and if the output $y(x_i)$ represents a probability, it is better to use the cross-entropy error loss[43]:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{n} (y_i \ln y(x_i) + (1 - y_i) \ln (1 - y(x_i))) \tag{15.23}$$

[42] It is called the target value or the pattern.

[43] We skip the subscript j because we assume that $j = 1$. We have then $y_i = y_{i,1}$ and $y(x_i) = y_1(x_i)$.

The choice of the loss function depends then on the output variable, but also on the activation function. For example, the cross-entropy error loss is adapted if $f_{z,y}$ corresponds to the logistic function, but not to the hyperbolic tangent function. In the case of the multi-class classification problem, Bishop (2006) proposes to consider the following loss function:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{n_y} y_{i,j} \ln y_j(x_i) \tag{15.24}$$

where n_y is equal to the number of classes n_C and $f_{z,y}$ corresponds to the softmax function that was previously defined in the case of the multi-logistic model:

$$\begin{aligned} y_j(x_i) &= f_{z,y}(v_{i,j}) \\ &= \frac{e^{v_{i,j}}}{\sum_{j'=1}^{n_y} e^{v_{i,j'}}} \end{aligned}$$

The loss function is then the opposite of the log-likelihood function.

Learning rules In order to minimize the loss function, we can use classical optimization algorithm[44] (Newton-Raphson, conjugate gradient, BFGS, DFP, Levenberg-Marquardt, etc.). As we have already said previously, this is not the philosophy of neural networks, and we generally prefer to use a statistical learning rule, which corresponds to an iterative algorithm:

$$\theta^{(t+1)} = \theta^{(t)} + \Delta\theta^{(t)}$$

where $\theta^{(t)}$ is the value of θ at the iteration (or epoch) t, and $\Delta\theta^{(t)}$ is the adjustment vector. The learning rule consists in defining how $\theta^{(t)}$ is updated, and is mostly based on the gradient of the loss function:

$$G(\theta) = \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$

Here are the main methods (Smith, 1993):

- The steepest descent method is defined by:

$$\Delta\theta^{(t)} = -\eta \cdot G\left(\theta^{(t)}\right)$$

  where $\eta > 0$ is the learning rate parameter. For minimizing the loss function, $\Delta\theta^{(t)}$ should go in the opposite direction of the gradient.

- For the momentum method, we have:

$$\begin{aligned} \Delta\theta^{(t)} &= -(1-\alpha_m)\,\eta \cdot G\left(\theta^{(t)}\right) + \alpha_m \cdot \Delta\theta^{(t-1)} \\ &= -\eta_m \cdot G\left(\theta^{(t)}\right) + \alpha_m \cdot \Delta\theta^{(t-1)} \end{aligned}$$

  where $\alpha_m \in [0,1]$ is the momentum weight and $\eta_m > 0$ is the momentum learning rate parameter. Therefore, the adjustment at iteration t is the weighted average of the adjustment at iteration $t-1$ and the steepest descent adjustment. The underlying idea of the term $\alpha_m \Delta\theta^{(t-1)}$ is to keep going in the previous direction. This method can speed up the algorithm because it may avoid oscillations[45].

[44] See Appendix A.1.3 on Page 1046.

[45] A better method is to consider the Nesterov approach:

$$\Delta\theta^{(t)} = -\eta_m \cdot G\left(\theta^{(t)} + \alpha_m \Delta\theta^{(t-1)}\right) + \alpha_m \cdot \Delta\theta^{(t-1)}$$

- The adaptive learning method is given by:

$$\Delta\theta^{(t)} = -\eta^{(t)} \cdot G\left(\theta^{(t)}\right)$$

where:

$$\eta^{(t)} = \begin{cases} \eta^{(t-1)} + \kappa & \text{if } G\left(\theta^{(t)}\right) \cdot K\left(\theta^{(t)}\right) \geq 0 \\ \phi \cdot \eta^{(t-1)} & \text{otherwise} \end{cases}$$

$\kappa > 0$, $0 < \phi < 1$ and $K\left(\theta^{(t)}\right) = G\left(\theta^{(t-1)}\right)$. Instead of using a fixed step η, we consider a variable step $\eta^{(t)}$ that depends on the previous value $\eta^{(t-1)}$. The variable step increases when the gradient does not change between iterations $t-1$ and t, and decreases otherwise. Another rule is to consider a moving average of the gradient:

$$\begin{aligned} K\left(\theta^{(t)}\right) &= (1-\varrho) \cdot G\left(\theta^{(t)}\right) + \varrho \cdot K\left(\theta^{(t-1)}\right) \\ &\approx (1-\varrho) \sum_{\tau=1} \varrho^{\tau-1} G\left(\theta^{(t-\tau)}\right) \end{aligned}$$

where $\varrho \in [0,1]$. In the case $\varrho = 0$, we retrieve the previous rule $K\left(\theta^{(t)}\right) = G\left(\theta^{(t-1)}\right)$.

- The adaptive learning with momentum method combines the two previous approaches:

$$\begin{aligned} \Delta\theta^{(t)} &= -(1-\alpha)\,\eta^{(t)} \cdot G\left(\theta^{(t)}\right) + \alpha \cdot \Delta\theta^{(t-1)} \\ &= -\eta_m^{(t)} \cdot G\left(\theta^{(t)}\right) + \alpha_m \cdot \Delta\theta^{(t-1)} \end{aligned}$$

There are numerous other algorithms[46] (adagrad, adam, nadam, rprop, rmsprop, etc.), and a lot of tricks for accelerating the convergence. First, we distinguish three approaches for evaluating the gradient of the objective function:

1. the batch gradient descent (BGD) computes the gradient with respect to the entire training dataset:

$$G\left(\theta^{(t)}\right) = \frac{\partial\,\mathcal{L}\left(\theta^{(t)}\right)}{\partial\,\theta}$$

2. the stochastic gradient descent (SGD) considers only one different training example at each iteration:

$$G\left(\theta^{(t)}\right) = \frac{\partial\,\mathcal{L}_i\left(\theta^{(t)}\right)}{\partial\,\theta}$$

where $\mathcal{L}_i(\theta)$ is the loss function for the i^{th} observation;

3. the mini-batch gradient descent (MGD) updates the parameters by using a subset of the training dataset:

$$G\left(\theta^{(t)}\right) = \sum_{i \in \mathcal{S}^{(t)}} \frac{\partial\,\mathcal{L}_i\left(\theta^{(t)}\right)}{\partial\,\theta}$$

where the subset $\mathcal{S}^{(t)}$ changes at each iteration.

The underlying idea is to evaluate the gradient not with respect to the current value $\theta^{(t)}$, but with respect to the prediction of the future value $\theta^{(t+1)}$. This prediction is equal to $\hat{\theta}^{(t+1)} = \theta^{(t)} + \alpha_m \Delta\theta^{(t-1)}$ in the momentum method.

[46] See Ruder (2016) for a review of recent approaches.

It is obvious that the choice of one approach depends on the size of the training data. Moreover, we better understand why the momentum approach is important when defining the learning rule. Indeed, in SGD and MGD approaches, the estimation of the gradient is more noisy than in the BGD approach. The momentum method helps to smooth the gradient and to obtain a more consistent direction.

We give here some default values that are used for the learning rules: $\eta = 1$, $\alpha_m = 0.75$, $\kappa = 0.1$, $\phi = 0.9$, $\alpha_m = 0.6$ and $\varrho = 0.5$. However, these parameters can change during the learning process. For instance, the learning rate parameter η can be greater at the beginning of the learning process, because of the large gradients. In a similar way, we can use a small momentum parameter α_m and then we can increase it progressively. We can also assume that the appropriate learning rules can vary between the parameters.

Error backpropagation In order to calculate the gradient $G(\theta)$, we consider a method called error backpropagation (or backward propagation). In the case of the loss function (15.22), we have $\mathcal{L}(\theta) = \sum_{i=1}^{n} \mathcal{L}_i(\theta)$ and:

$$\mathcal{L}_i(\theta) = \sum_{j=1}^{n_y} \frac{1}{2} (y_j(x_i) - y_{i,j})^2$$

It follows that $G(\theta) = \sum_{i=1}^{n} G_i(\theta)$ where $G_i(\theta)$ is the gradient of $\mathcal{L}_i(\theta)$. In the case of the previous loss, we can use the decomposition $\mathcal{L}_i(\theta) = \sum_{j=1}^{n_y} \mathcal{L}_{i,j}(\theta)$ where:

$$\mathcal{L}_{i,j}(\theta) = \frac{1}{2} (y_j(x_i) - y_{i,j})^2$$

Using chain rule, we obtain[47]:

$$\begin{aligned} \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial \gamma_{j,h}} &= \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial y_j(x_i)} \cdot \frac{\partial y_j(x_i)}{\partial v_{i,j}} \cdot \frac{\partial v_{i,j}}{\partial \gamma_{j,h}} \\ &= (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j}) z_{i,h} \end{aligned}$$

and $\partial_{\gamma_{j,h}} \mathcal{L}_{i,j'}(\theta) = 0$ when $j \neq j'$. We also deduce that[48]:

$$\begin{aligned} \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial \beta_{h,k}} &= \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial z_{i,h}} \cdot \frac{\partial z_{i,h}}{\partial u_{i,h}} \cdot \frac{\partial u_{i,h}}{\partial \beta_{h,k}} \\ &= \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial z_{i,h}} f'_{x,z}(u_{i,h}) x_{i,k} \\ &= (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j}) \gamma_{j,h} f'_{x,z}(u_{i,h}) x_{i,k} \end{aligned}$$

In the case of Model (15.20), there is a constant and we have:

$$\frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial \gamma_{j,0}} = (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j})$$

[47] The distinction between j and j' is important when we consider the softmax function (see Exercise 15.4.7 on page 1025).

[48] Because we have:

$$\begin{aligned} \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial z_{i,h}} &= \frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial y_j(x_i)} \cdot \frac{\partial y_j(x_i)}{\partial v_{i,j}} \cdot \frac{\partial v_{i,j}}{\partial z_{i,h}} \\ &= (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j}) \gamma_{j,h} \end{aligned}$$

and:

$$\frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial \beta_{h,0}} = (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j}) \gamma_{j,h} f'_{x,z}(u_{i,h})$$

In the case of Model (15.21), we have for the direct links:

$$\frac{\partial \mathcal{L}_{i,j}(\theta)}{\partial \gamma_{j,n_z+k}} = (y_j(x_i) - y_{i,j}) f'_{z,y}(v_{i,j}) x_{i,k}$$

It follows that the neural network consists in two steps. The forward propagation computes $u_{i,h}$, $z_{i,h}$, $v_{i,j}$ and $y_j(x_i)$, meaning that the information comes from left to right. The backward propagation computes all the derivatives using the chain rule, implying that the information goes from right to left.

Remark 191 *All the previous quantities can be calculated in a matrix form in order to avoid loop implementation (see Exercise 15.4.7 on page 1025).*

Since $f'_{z,y}$ and $f'_{x,z}$ are easy to calculate, all the derivatives are calculated in a closed-form expression. For instance, the derivative of the logistic activation function is equal to:

$$\begin{aligned} f'(z) &= \frac{e^{-z}}{(1+e^{-z})^2} \\ &= \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) \\ &= f(z)(1-f(z)) \end{aligned}$$

It follows that $f'_{z,y}(v_{i,j}) = f_{z,y}(v_{i,j})(1 - f_{z,y}(v_{i,j})) = y_j(x_i)(1 - y_j(x_i))$ and $f'_{x,z}(u_{i,h}) = z_{i,h}(1 - z_{i,h})$. In Exercise 15.4.7 on page 1025, we consider other activation functions and loss functions.

Examples Neural networks are sufficiently flexible that they can approximate any continuous function. Therefore, they are said to be '*universal approximators*' (Bishop, 2006). Figure 15.26 illustrates this property when the function is $f(x) = 2\cos(x)$ or $f(x) = |x| - 2$. For that, we use the network structure (15.20) with two constants and direct links. The activation function $f_{x,z}$ is the hyperbolic tangent function, while the output scaling function $f_{z,y}$ is the identity function. The training step is done with 201 uniform points between -4 and $+4$. We notice that the accuracy depends on the number n_z of hidden nodes. In particular, the approximation is very good when we consider three hidden nodes. The universal approximation property is certainly the main strength of neural networks. It suffices to increase the number of hidden nodes in order to achieve a given accuracy. However, this property is also the main weakness of neural networks. Indeed, the distinction between training and validation steps is not obvious, and overfitting risk is large.

The trade-off between n_z and $\mathcal{L}(\theta)$ is not the only issue with neural networks. Another problem is the scaling of data. By applying activation functions, the output domain is not necessarily the set $\mathbb{R}^{n_y}$. In Figure 15.27, we have reported the approximation of $f(x) = |x| - 2$ by considering two hidden nodes and different configurations. The first panel corresponds to the network structure (15.19) without constant and direct link ($\beta_0 = 0$, $\gamma_0 = 0$ and $\gamma_x = 0$). In the second panel, we include the two constants β_0 and γ_0, but not the direct links ($\gamma_x = 0$). We notice that this second structure is better to approximate the function than the structure of the first panel. The reason is the range of $\operatorname{dom} f(x)$, which is better managed by including a constant γ_0. This is confirmed by the third panel. Finally, the fourth panel assumes that the output scaling function $f_{z,y}$ is the logistic sigmoid function. In this

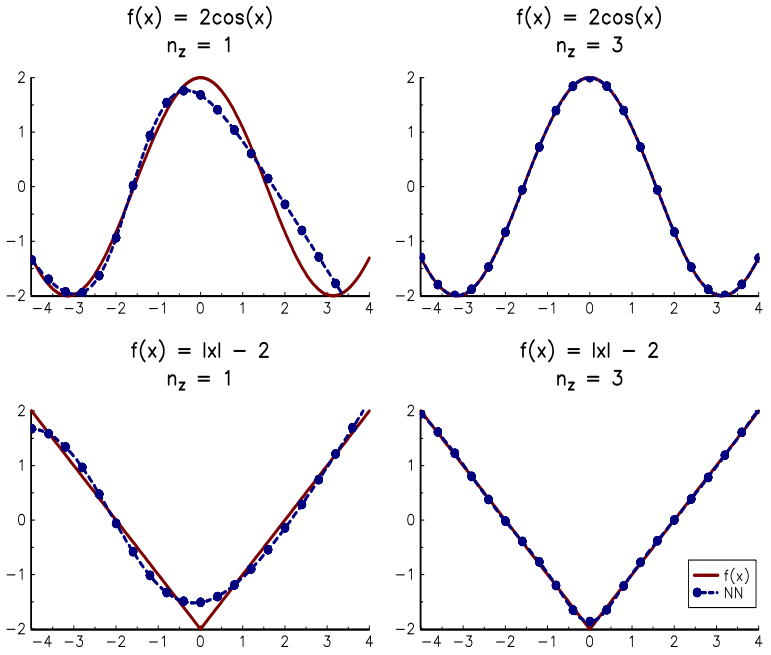


FIGURE 15.26: Neural networks as universal approximators

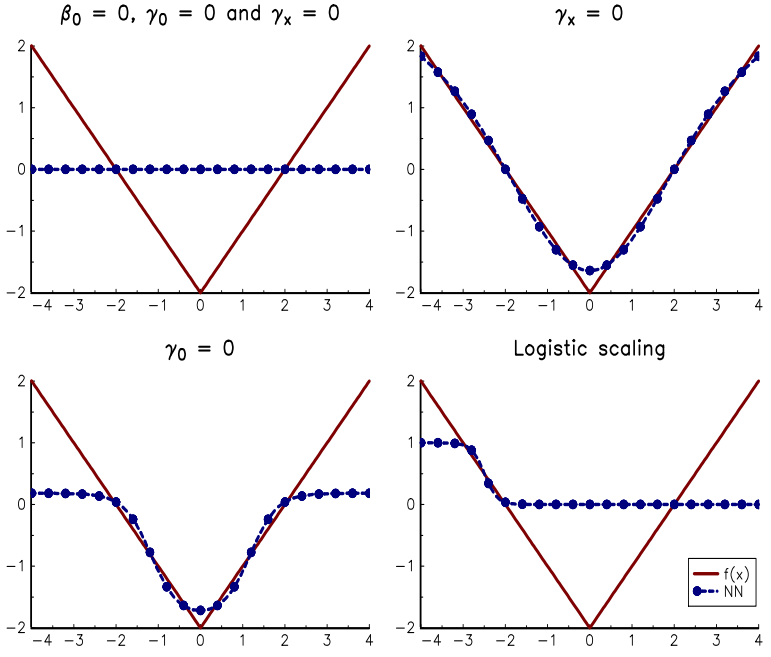


FIGURE 15.27: The scaling issue of neural networks ($f(x) = |x| - 2$)

case, it is obvious that the output $y(x_j) \in [0,1]$ cannot reach dom $f(x) = [-2,2]$. This case is trivial while the three previous cases are not. This means that the network structure is crucial, not only the number of hidden units, but also the choice of activation and scaling functions, adding or not constants and direct links, and the scaling of both input and output data.

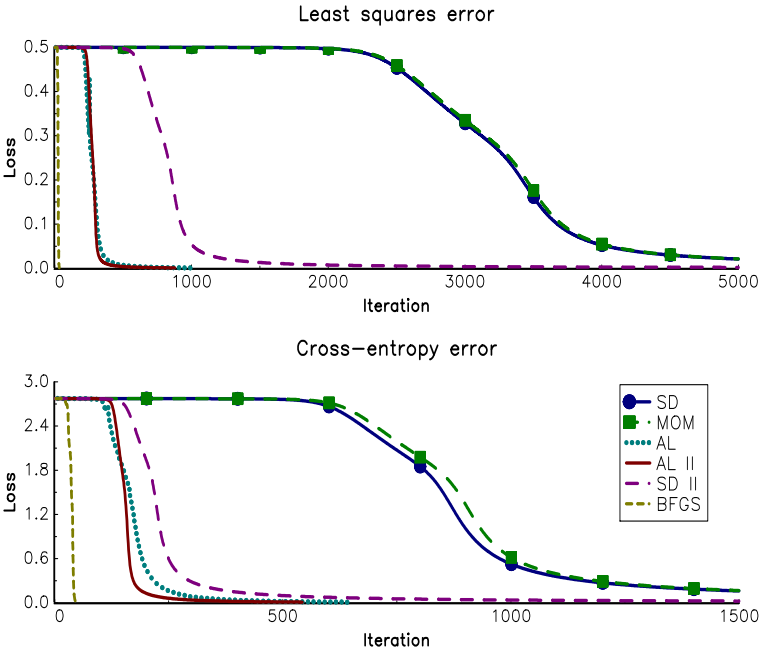


FIGURE 15.28: Convergence of the XOR problem

The XOR (or exclusive or) problem is a classic problem in neural networks. We note $y = x_1 \oplus x_2$ where x_1 and x_2 are two binary outputs:

$$\begin{cases} 0 \oplus 0 = 0 \\ 0 \oplus 1 = 1 \\ 1 \oplus 0 = 1 \\ 1 \oplus 1 = 0 \end{cases}$$

The XOR problem can be viewed as a supervised classification problem. In order to solve this problem, we use a neural network with three hidden nodes with no constant and no direct link. The activation and output scaling functions are set to the logistic function. In Figure 15.28, we have represented the evolution of the loss function $\mathcal{L}(\theta)$ with respect to the iterations of the learning rules, which are steepest descent (SD), momentum (MOM), adaptive learning (AL) and adaptive learning with momentum (AL II) methods. We also consider a steepest descent with optimal stepsize (SD II) and the BFGS algorithm. Moreover, we have used the two loss criteria: least squares and cross-entropy errors. The results show the following major lessons. First, we notice that the convergence highly depends on the learning rule, but also on the loss criterion. Second, a comparison of the optimal parameters $\hat{\theta}$ shows that they are all different. They differ from one learning rule to another, but they also differ from one loss criterion to another even if we use the same learning rule. This result is not surprising, because we observe that the solution $\hat{\theta}$ changes each time we consider new starting values. This means that neural networks produce models that

are overidentified. In this context, it is perfectly illusory to analyze and understand the estimated model. As we have already said, only the predictions $\hat{y}_{i,j}$ are relevant.

TABLE 15.15: Data of program effectiveness

OBS	GPA	TUCE	PSI	GRD	OBS	GPA	TUCE	PSI	GRD
1	2.66	20	0	0	17	2.75	25	0	0
2	2.89	22	0	0	18	2.83	19	0	0
3	3.28	24	0	0	19	3.12	23	1	0
4	2.92	12	0	0	20	3.16	25	1	1
5	4.00	21	0	1	21	2.06	22	1	0
6	2.86	17	0	0	22	3.62	28	1	1
7	2.76	17	0	0	23	2.89	14	1	0
8	2.87	21	0	0	24	3.51	26	1	0
9	3.03	25	0	0	25	3.54	24	1	1
10	3.92	29	0	1	26	2.83	27	1	1
11	2.63	20	0	0	27	3.39	17	1	1
12	3.32	23	0	0	28	2.67	24	1	0
13	3.57	23	0	0	29	3.65	21	1	1
14	3.26	25	0	1	30	4.00	23	1	1
15	3.53	26	0	0	31	3.10	21	1	0
16	2.74	19	0	0	32	2.39	19	1	1

Source: Greene (2017), Table F14.1 and Spector and Mazzeo (1980).

We consider the classification problem described in Greene (2017) based on the study of Spector and Mazzeo (1980), who examined whether a new method of teaching economics, the personalized system of instruction (PSI), significantly influenced performance in later economics courses. The corresponding data are reproduced in Table 15.15. OBS is the observation, that is the student. The output variable is GRD, which corresponds to the grade increase (1) or decrease (0) indicator for the student. The explanatory variables are the constant C, the grade point average GPA, the test score on economics test TUCE, and the binary variable PSI that indicates the participation to the new teaching method. Following Greene (2017), we estimate the following logit model:

$$\begin{aligned}\Pr\{\text{GRD}_i = 1\} &= \mathbf{F}\left(\beta_0 + \beta_1\,\text{GPA}_i + \beta_2\,\text{TUCE}_i + \beta_3\,\text{PSI}_i\right)\\ &= \mathbf{F}\left(x_i^\top \beta\right)\end{aligned}$$

where $\mathbf{F}$ is the cumulative distribution function of the logistic distribution. The results are reported in Table 15.16, and the value of the optimized log-likelihood function is $\boldsymbol{\ell}\left(\hat{\beta}\right) = -12.8896$. In order to challenge the logistic regression, we consider a neural network with three hidden nodes. The logistic function is used for both the activation and output scaling functions and we consider a direct link between the input variables (C, GPA, TUCE and PSI) and the output variable GRD. In Table 15.17, we have calculated the estimated probability $\hat{p}_i = \Pr\{\text{GRD}_i = 1\}$ in the cases of the logit model:

$$\hat{p}_i = \mathbf{F}\left(x_i^\top \hat{\beta}^{(\text{logit})}\right)$$

and the neural network:

$$\hat{p}_i = \mathbf{F}\left(\hat{\gamma}_z^{(\text{nn})}\mathbf{F}\left(\hat{\beta}_x^{(\text{nn})} x_i\right) + \hat{\gamma}_x^{(\text{nn})} x_i\right)$$

TABLE 15.16: Results of the logistic regression

Parameter	Estimate	Standard error	t-statistic	p-value
β_0	−13.0214	4.9313	−2.6405	0.0134
β_1	2.8261	1.2629	2.2377	0.0334
β_2	0.0952	0.1415	0.6722	0.5069
β_3	2.3787	1.0646	2.2344	0.0336

TABLE 15.17: Estimated probability $\hat{p}_i = \Pr\{\text{GRD}_i = 1\}$

OBS	Logit	NN	OBS	Logit	NN
1	2.658	2.658	17	5.363	5.363
2	5.950	5.950	18	3.859	3.859
3	18.726	18.726	19	58.987	58.987
4	2.590	2.590	20	66.079	66.079
5	56.989	56.989	21	6.138	6.138
6	3.486	3.486	22	90.485	90.485
7	2.650	2.650	23	24.177	24.177
8	5.156	5.156	24	85.209	85.209
9	11.113	11.113	25	83.829	83.829
10	69.351	69.351	26	48.113	48.113
11	2.447	2.447	27	63.542	63.542
12	19.000	19.000	28	30.722	30.722
13	32.224	32.224	29	84.170	84.170
14	19.321	19.321	30	94.534	94.534
15	36.099	36.099	31	52.912	52.912
16	3.018	3.018	32	11.103	11.103

The results are surprising. The estimated probability calculated with the neural network is exactly equal to the estimated probability calculated with the logit model. If we inspect the estimated coefficient, we obtain:

$$\hat{\beta}_x^{(\text{nn})} = \begin{pmatrix} 1.0343 & 0.8482 & 1.0678 & 0.5770 \\ 0.3856 & 0.1976 & 1.4420 & 0.8744 \\ 0.1925 & 0.8791 & 2.0427 & 0.5439 \end{pmatrix}$$

$\hat{\gamma}_z^{(\text{nn})} = (-2.9240, -2.9538, -3.6783)$ and $\hat{\gamma}_x^{(\text{nn})} = (-3.4652, 2.8261, 0.0952, 2.3787)$. Moreover, the loss error is equal to $\mathcal{L}\left(\hat{\theta}\right) = 12.8896$, which is exactly the opposite of the optimized log-likelihood function. This result is not surprising because the neural network encompasses the logit model:

$$\Pr\{\text{GRD}_i = 1\} = \mathbf{F}\left(\underbrace{\gamma_z^{(\text{nn})}\mathbf{F}\left(\beta_x^{(\text{nn})}x_i\right)}_{\text{specific nn effect}} + \underbrace{\gamma_x^{(\text{nn})}x_i}_{\text{logit effect}}\right)$$

We also notice that the logit coefficients are the same than the neural network coefficients for the direct link units ($\hat{\beta}^{(\text{logit})} = \hat{\gamma}_x^{(\text{nn})}$) with the exception of the constant[49]. Let us estimate

[49]The constant is equal to −13.0214 for the logit model and −3.4652 for the neural network.

the neural network by using other starting values for the optimization step. We obtain the same probability than previously, but the estimated coefficients are not the same. We have:

$$\hat{\beta} = \begin{pmatrix} 0.4230 & 0.9108 & 0.5875 & 0.0882 \\ 0.9586 & 0.2078 & 0.7862 & 0.4852 \\ 0.7835 & 2.9180 & 8.7259 & 0.4899 \end{pmatrix}$$

$\hat{\gamma}_z^{(\text{nn})} = (-4.5296, -4.3299, -4.2120)$ and $\hat{\gamma}_x^{(\text{nn})} = (0.0501, 2.8261, 0.0952, 2.3787)$. Again, the neural network coefficients for the direct link units are equal to the logit coefficients with the exception of the constant. We deduce that the neural network does not differ from the logit model, because we have:

$$\hat{\beta}_0^{(\text{logit})} = \gamma_z^{(\text{nn})} \mathbf{F}\left(\beta_x^{(\text{nn})} x_i\right) + \gamma_1^{(\text{nn})}$$

This result is interesting, because it shows that the neural network did not better than the logit model, although it presents more flexibility.

Remark 192 *The previous results are explained because we optimize the cross-entropy error loss for estimating the parameters of the neural network. This implies that the logit framework is perfectly compatible with the neural network framework.*

15.2.3.3 Support vector machines

The overidentification of neural networks is an important issue and the optimization step involves an objective function, which is generally not convex with respect to the parameters. This implies that there are many local minima. Moreover, the foundation of neural networks suffers from little theoretical basis of these learning models. Like neural networks, support vector machines (SVM) can be seen as an extension of the perceptron. However, it presents nice theoretical properties and a strong geometrical framework. Once SVMs have been first developed for linear classification, they have been extended for non-linear classification and regression.

TABLE 15.18: An example of linearly separable observations

i	1	2	3	4	5	6	7	
$x_{i,1}$	0.5	2.7	2.7	1.7	1.5	2.3	4.0	
$x_{i,2}$	2.5	4.2	2.0	4.2	0.7	5.3	6.9	
y_i	+1	+1	+1	+1	+1	+1	+1	
i	8	9	10	11	12	13	14	15
$x_{i,1}$	6.4	7.7	8.8	7.4	6.5	8.3	6.0	5.0
$x_{i,2}$	4.5	2.2	6.0	6.5	1.7	1.3	1.3	0.5
y_i	−1	−1	−1	−1	−1	−1	−1	−1

Separating hyperplanes We consider a training set $\{(x_i, y_i), i = 1, \ldots, n\}$, where the response variable y_i can take the values -1 and $+1$. This training set is said linearly separable if there is a hyperplane $\mathcal{H} = \left\{x \in \mathbb{R}^K : f(x) = \beta_0 + x^\top \beta = 0\right\}$ such that:

$$y_i = \text{sign}\, f(x_i)$$

This means that the hyperplane divides the affine space in two half-spaces[50] such that $\{i : y_i = +1\} \in \mathbf{H}^+$ and $\{i : y_i = -1\} \in \mathbf{H}^-$. Let us consider the example with two explanatory variables given in Table 15.18. We have represented the data $(x_{i,1}, x_{i,2})$ and the

[50] The upper half-space $\mathbf{H}^+$ is defined by $f(x) > 0$ while the lower half-space $\mathbf{H}^-$ corresponds to $f(x) < 0$.

corresponding label y_i in Figure 15.29. It is obvious that this training set is linearly separable. For example, we have reported three hyperplanes $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ that perform a perfect classification.

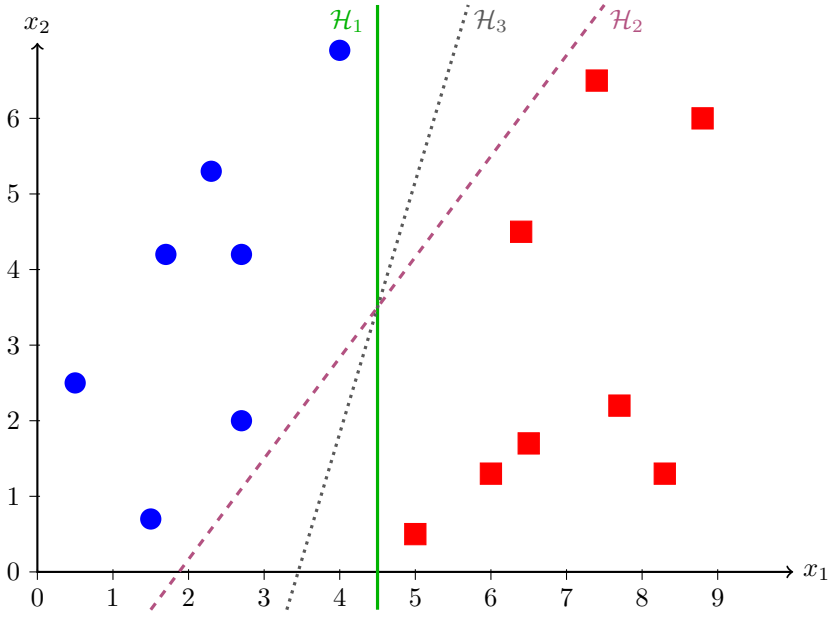


FIGURE 15.29: Separating hyperplane picking

Since there are many solutions, we may wonder if there exists one solution that dominates the others. The answer has been proposed by Vladimir Vapnik and Alexey Chervonenkis in the sixties, who have formulated the concept of support vector machines. Following Cortes and Vapnik (1995), the optimal hyperplane is the one that maximizes the margin. In Figure 15.30, we have represented an hyperplane and the two margins $M+$ and $M-$, which corresponds to the Euclidean distance between the hyperplane and the closest positive and negative points. The underlying idea of Vapnik and Chervonenkis is then to find the hyperplane $\mathcal{H}$ with the largest values of $M+$ and $M-$.

We notice that finding a hyperplane with two different margins $M_+ \neq M_-$ is equivalent to define a hyperplane with the same positive and negative margins: $M_+ = M_- = M$. This implies that the two separating hyperplanes $\mathcal{H}_+$ and $\mathcal{H}_-$ are equidistant to the hyperplane $\mathcal{H}$. The estimation of $\mathcal{H}_+$ and $\mathcal{H}_-$ requires identifying the training points that belongs to $\mathcal{H}_+$ and $\mathcal{H}_-$. These points are called the support vectors. In the case of Figure 15.30, two support vectors are necessary to define $\mathcal{H}_+$ and $\mathcal{H}_-$, or equivalently $\mathcal{H}$ and the margin M. By construction, the number of support points is at least equal to the number of explanatory variables. Except in degenerate cases, there are much less number of support points than the number of observations. This implies that not all observations are relevant for defining the decision boundary of an optimal linear classifier. Only the support vectors are important.

Hard margin classification The maximization problem is:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}\right\} &= \arg\max M \\ \text{s.t.} &\begin{cases} f(x_i) \geq M & \text{if} \quad y_i = +1 \\ f(x_i) \leq -M & \text{if} \quad y_i = -1 \end{cases} \end{aligned}$$

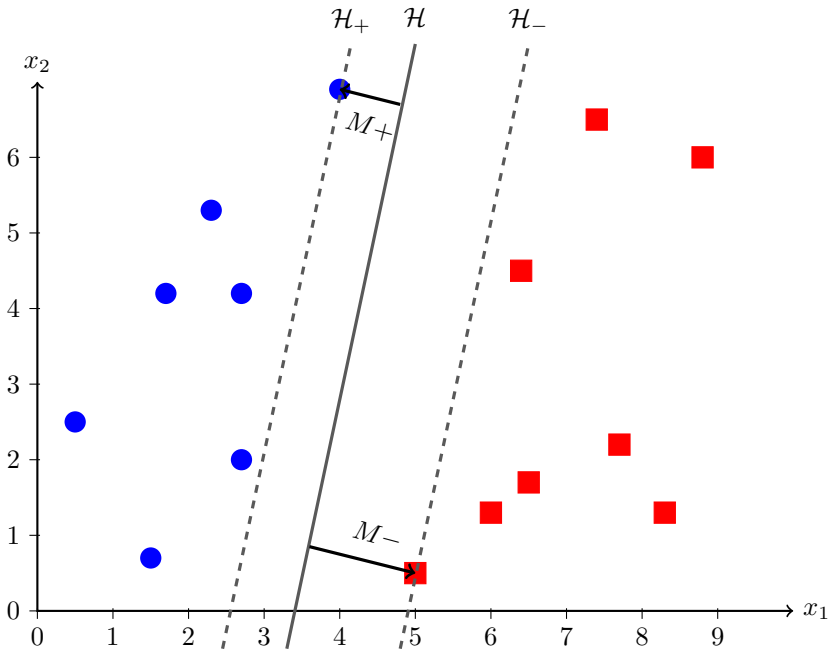


FIGURE 15.30: Margins of separation

However, this optimization problem is not well defined, since M depends on β. More precisely, it is inversely proportional to $\|\beta\|_2$. This is why we need to add another constraint, e.g. $\beta_1 = 1$ or $\|\beta\|_2 = 1$. Another approach is to standardize the problem by setting $M = 1$.

Let x_- and x_+ be two (negative and positive) support vectors, we deduce that the distance between x_- and x_+ is equal to[51]:

$$d(x_-, x_+) = \beta^\top (x_+ - x_-) = 2M$$

If we replace β by the corresponding unit vector $\hat{\beta} = \beta / \|\beta\|_2$, we obtain $\beta^\top (x_+ - x_-) = 2\hat{M} \|\beta\|_2$. By setting $M = 1$, we obtain $\hat{M} = 1/\|\beta\|_2$. Maximizing the margin is then equivalent to maximize $1/\|\beta\|_2$ or minimize $\|\beta\|_2$ (or $\|\beta\|_2^2$). Moreover, we notice that the inequality constraints[52] can be compacted as $y_i f(x_i) \geq 1$. Finally, we obtain the following optimization problem:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}\right\} &= \arg\min \frac{1}{2} \|\beta\|_2^2 \\ \text{s.t.} \quad & y_i \left(\beta_0 + x_i^\top \beta\right) \geq 1 \qquad \text{for } i = 1, \ldots, n \end{aligned} \tag{15.25}$$

We recognize a standard quadratic programming (QP) problem that can be easily solved from a numerical point of view.

Using the training set given in Table 15.18 on page 989, and solving the QP problem (15.25), we obtain $\hat{\beta}_0 = 2.416$, $\hat{\beta}_1 = -0.708$ and $\hat{\beta}_2 = 0.248$. It follows that the margin M is equal to 1.333. Since the equation $\beta_0 + \beta_1 x_1 + \beta_2 x_2 = c$ is equivalent to:

$$x_2 = \frac{c - \beta_0}{\beta_2} - \frac{\beta_1}{\beta_2} x_1$$

[51] We have $\beta_0 + x_-^\top \beta = -M$ and $\beta_0 + x_+^\top \beta = M$.
[52] Because we have set $M = 1$.

we deduce that the equations of the three hyperplanes $\mathcal{H}_-$, $\mathcal{H}$ and $\mathcal{H}_+$ are:

$$\begin{array}{lll} \mathcal{H}_- : & x_2 = -13.786 + 2.857 \cdot x_1 & (c=-1) \\ \mathcal{H} : & x_2 = -9.750 + 2.857 \cdot x_1 & (c=0) \\ \mathcal{H}_+ : & x_2 = -5.714 + 2.857 \cdot x_1 & (c=+1) \end{array}$$

We have reported the estimated hyperplanes in Figure 15.31, and have also indicated the support vectors, which are only three.

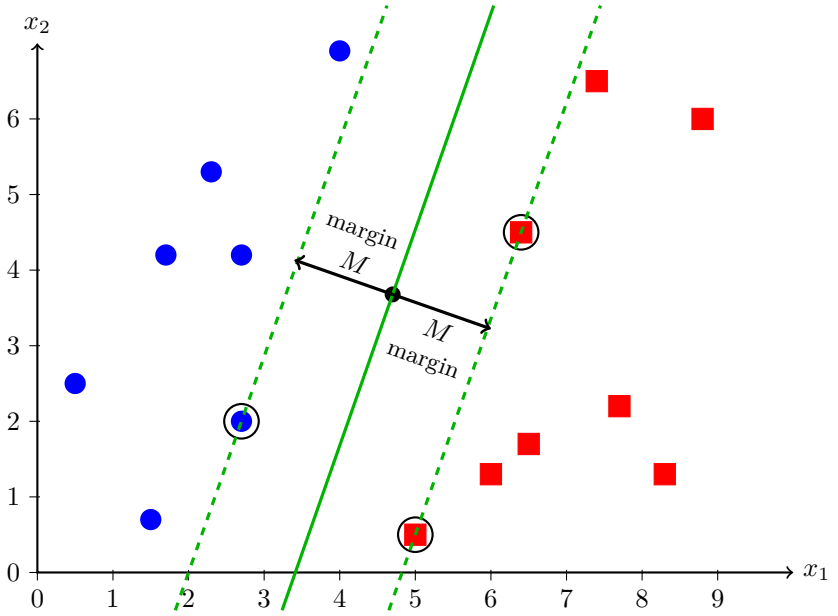


FIGURE 15.31: Optimal hyperplane

The historical approach to estimate a support vector machine is to map the primal QP problem to the dual QP problem. Using the results provided in Appendix A.1.3.1 on page 1046, we can show that[53]:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \Gamma \alpha - \alpha^\top \mathbf{1}_n \\ &\text{s.t.} \begin{cases} y^\top \alpha = 0 \\ \alpha \geq \mathbf{0}_n \end{cases} \end{aligned} \tag{15.26}$$

where α is the vector of Lagrange multipliers associated to the n inequality constraints and $\Gamma_{i,j} = y_i y_j x_i^\top x_j$. Moreover, we have:

$$\hat{\beta} = \sum_{i=1}^{n} \hat{\alpha}_i y_i x_i$$

The optimal value of $\hat{\beta}_0$ can be deduced from any support vectors. In the case of a positive support vector x_+, we have $\hat{\beta}_0 = 1 - x_+^\top \hat{\beta}$, while we have $\hat{\beta}_0 = -1 - x_-^\top \hat{\beta}$ for any negative support vector x_-. Moreover, we can classify new observations by considering the following rule:

$$\hat{y} = \text{sign}\left(\hat{\beta}_0 + x^\top \hat{\beta}\right)$$

[53] See Exercise 15.4.8 on page 1027.

If we consider our example, we observe that $\hat{\alpha}_i$ is different from zero for three observations: $i \in \{3, 8, 15\}$. They correspond to the three support vectors that we have found graphically. We obtain $\hat{\alpha}_3 = 0.2813$, $\hat{\alpha}_8 = 0.0435$ and $\hat{\alpha}_{15} = 0.2378$. With these values, we deduce that $\hat{\beta}_1 = -0.708$ and $\hat{\beta}_2 = 0.248$. In order to compute $\hat{\beta}_0$, we consider one of the support vectors and calculate $\hat{\beta}_0 = y_i - x_i^\top \hat{\beta}$. For example, in the case of the first support vector (or the third observation), we have: $\hat{\beta}_0 = 1 + 2.7 \times 0.708 - 2 \times 0.248 = 2.416$.

Remark 193 *We may wonder what the rational of using the dual problem is. The primal problem is a QP problem with $K + 1$ unknowns and n inequality constraints. The dual problem is a QP problem with n unknowns, one equality constraint and n box constraints. Since the last constraints are straightforward to manage, the second problem is easier to solve than the first problem. However, the dimension of the second problem is larger than this of the first problem, since we have to calculate the Γ matrix of dimension $n \times n$. Therefore, it is difficult to justify that the dual problem presents less computational issues than the primal problem. The reason is to be found elsewhere. In fact, the calculation of Γ involves the calculation of the inner product $\langle x_i, x_j \rangle = x_i^\top x_j$. We will see later that it corresponds to a covariance kernel, and the dual problem can be used in a more efficient way than the primal problem with other covariance kernels when we consider non-linear SVM problems.*

Soft margin classification The inequality constraints $y_i \left(\beta_0 + x_i^\top \beta\right) \geq 1$ ensure that all the training points are well-classified and belongs to the half-spaces $\mathbf{H}_+$ and $\mathbf{H}_-$. However, training data are generally not fully linearly separable. Therefore, we can relax these constraints by introducing slack variables $\xi_i > 0$:

$$y_i \left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i$$

We then face three situations:

1. if $\xi_i = 0$, the observation i is well-classified since we have $y_i \left(\beta_0 + x_i^\top \beta\right) \geq 1$;
2. if $0 < \xi_i \leq 1$, the observation i is located in the '*street*', that is in the area between the two separating planes $\mathcal{H}_-$ and $\mathcal{H}_+$; in this case, ξ_i can be interpreted as the margin error ($\xi_i \leq M$);
3. if $\xi_i > 1$, the observation i is fully misclassified.

The quality of the classification can be measured by the misclassification error sum, that we can bound:

$$\sum_{i=1}^{n} \xi_i \leq \xi^+$$

The parameter ξ^+ indicates the tolerance we have with respect to the hard margin classification. Instead of adding the inequality constraint $\sum_{i=1}^n \xi_i \leq \xi^+$ in Problem (15.25), we can penalize the objective function:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^{n} \xi_i \\ &\text{s.t.} \quad y_i \left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i \qquad \text{for } i = 1, \ldots, n \end{aligned} \tag{15.27}$$

where the parameter C controls the level of errors. If C is large, the norm $\|\beta\|_2$ can be large. On the contrary, if C is small, the sum $\sum_{i=1}^n \xi_i$ can be large, but not the norm $\|\beta\|_2$.

As the margin M is equal to $1/\|\beta\|_2$, C controls then the trade-off between the size of the margin and the misclassification error rate. The dual problem is[54]:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \Gamma \alpha - \alpha^\top \mathbf{1}_n \\ \text{s.t.} &\begin{cases} y^\top \alpha = 0 \\ \mathbf{0}_n \leq \alpha \leq C \cdot \mathbf{1}_n \end{cases} \end{aligned} \tag{15.28}$$

Again, we have $\hat{\beta} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$. Support vectors corresponds then to training points such that $0 < \alpha_i < C$. For computing $\hat{\beta}_0$, we average over all the support vectors:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n \mathbb{1}\{0 < \hat{\alpha}_i < C\} \cdot \left(y_i - x_i^\top \hat{\beta}\right)}{\sum_{i=1}^n \mathbb{1}\{0 < \hat{\alpha}_i < C\}}$$

Since we have $y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i$ and $\xi_i \geq 0$, the Kuhn-Tucker conditions implies that:

$$\hat{\xi}_i = \max\left(0, 1 - y_i\left(\hat{\beta}_0 + x_i^\top \hat{\beta}\right)\right) \tag{15.29}$$

The classification rule does not change, and we have $\hat{y} = \text{sign}\left(\hat{\beta}_0 + x^\top \hat{\beta}\right)$.

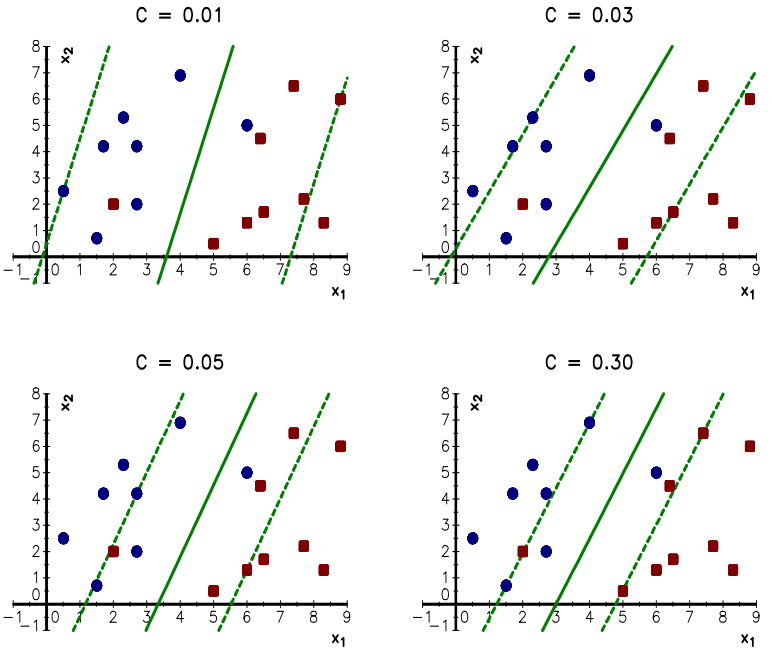


FIGURE 15.32: Soft margin SVM classifiers

We consider the previous training set given in Table 15.18 and we introduce two points $(6.0, 5.0, +1)$ $(i = 16)$ and $(2.0, 2.0, -1)$ $(i = 17)$. In this case, the training set is not linearly separable. Considering different values of C, we have represented the optimal hyperplanes in Figure 15.32. We verify that the margin decreases when C increases. In the case where C is equal to 0.05, we obtain $\hat{\beta}_0 = 1.533$, $\hat{\beta}_1 = -0.458$, $\hat{\beta}_2 = 0.168$, and the optimal value of α_i and ξ_i are reported in Table 15.19.

[54] See Exercise 15.4.8 on page 1027

TABLE 15.19: Soft margin classification with $C = 0.05$

i	y_i	$x_{i,1}$	$x_{i,2}$	$\hat{\alpha}_i$	$\hat{\xi}_i$
1	+1	0.5	2.5	0.000	0.000
2	+1	2.7	4.2	0.039	0.000
3	+1	2.7	2.0	0.050	0.369
4	+1	1.7	4.2	0.000	0.000
5	+1	1.5	0.7	0.050	0.038
6	+1	2.3	5.3	0.000	0.000
7	+1	4.0	6.9	0.050	0.143
8	−1	6.4	4.5	0.050	0.354
9	−1	7.7	2.2	0.000	0.000
10	−1	8.8	6.0	0.000	0.000
11	−1	7.4	6.5	0.050	0.231
12	−1	6.5	1.7	0.000	0.000
13	−1	8.3	1.3	0.000	0.000
14	−1	6.0	1.3	0.039	0.000
15	−1	5.0	0.5	0.050	0.324
16	+1	6.0	5.0	0.050	1.379
17	−1	2.0	2.0	0.050	1.952

If we combine Equations (15.27) and (15.29), we obtain:

$$\begin{aligned} f(\beta_0, \beta) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\max\left(0, 1 - y_i\left(\beta_0 + x_i^\top\beta\right)\right) \\ &= C\cdot\left(\sum_{i=1}^{n}\max\left(0, 1 - y_i\left(\beta_0 + x_i^\top\beta\right)\right) + \frac{1}{2C}\|\beta\|_2^2\right) \end{aligned}$$

We deduce that the optimization program is:

$$\arg\min \mathcal{R}(x, y) + \frac{1}{2C}\|\beta\|_2^2 \tag{15.30}$$

where $\mathcal{R}(x, y) = \sum_{i=1}^{n}\mathcal{L}(x_i, y_i)$ and $\mathcal{L}(x_i, y_i)$ is the binary hinge loss:

$$\mathcal{L}(x_i, y_i) = \max\left(0, 1 - y_i\left(\beta_0 + x_i^\top\beta\right)\right)$$

It follows that the soft margin classification corresponds to a risk minimization problem with a ridge penalization. The problem is convex but non-smooth because $\mathcal{L}(x_i, y_i)$ is non-differentiable. More generally, we can use other loss functions, for instance the $0-1$ loss:

$$\mathcal{L}^{0-1}(x_i, y_i) = \begin{cases} 0 & \text{if } y_i\left(\beta_0 + x_i^\top\beta\right) \geq 1 \\ 1 & \text{otherwise} \end{cases}$$

However, the associated risk measure is non-convex, and the minimization problem is computationally hard. A better approach is to consider the squared hinge loss:

$$\mathcal{L}^{\text{squared}}(x_i, y_i) = \mathcal{L}^{\text{hinge}}(x_i, y_i)^2$$

In this case, the problem is convex and smooth. Another popular loss function is the ramp loss:

$$\mathcal{L}^{\text{ramp}}(x_i, y_i) = \min\left(1, \mathcal{L}^{\text{hinge}}(x_i, y_i)\right)$$

The derivation of the dual problems and the comparison of these different loss functions are discussed in Exercise 15.4.8 on page 1028.

SVM regression Support vector machines can be extended to output variables that are continuous. In this case, we have to define an appropriate loss function. For instance, if we consider the least squares loss function, we have:

$$\mathcal{L}^{\text{ls}}(x_i, y_i) = (y_i - f(x_i))^2$$

where $f(x) = \beta_0 + x^\top \beta$. The corresponding SVM regression is then:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 \\ &\text{s.t.} \quad y_i = \beta_0 + x_i^\top \beta + \xi_i \qquad \text{for } i = 1, \ldots, n \end{aligned} \tag{15.31}$$

It is obvious that ξ_i plays the role of the residual. This regression problem looks very similar to the SVM problem for the soft margin classification and the squared hinge loss function. In particular, we can show that the dual problem is[55]:

$$\begin{aligned} \hat{\alpha} &= \arg\min \frac{1}{2}\alpha^\top \left(XX^\top + \frac{1}{2C}I_n\right)\alpha - \alpha^\top Y \\ &\text{s.t.} \quad \mathbf{1}_n^\top \alpha = 0 \end{aligned} \tag{15.32}$$

Once we have solved this QP problem, we can calculate the prediction for x: $\hat{y} = \hat{\beta}_0 + x^\top \hat{\beta}$.

Vapnik (1998) proposed another loss function in order to keep the formalism of the original soft margin problem:

$$\mathcal{L}^{\text{ls}}(x_i, y_i) = \mathbb{1}\{|y_i - f(x_i)| \geq \varepsilon\} \cdot (|y_i - f(x_i)| - \varepsilon)$$

where $\varepsilon > 0$. It follows that:

$$\mathcal{L}^{\text{ls}}(x_i, y_i) = \begin{cases} |y_i - f(x_i)| - \varepsilon & \text{if} \quad |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{if} \quad |y_i - f(x_i)| \leq \varepsilon \end{cases}$$

Therefore, we would like to find a hyperplane such that we don't care about the errors that are smaller than ε. We have:

$$\begin{aligned} \mathcal{L}^{\text{ls}}(x_i, y_i) &= \mathbb{1}\{y_i - f(x_i) \leq -\varepsilon\} \cdot (f(x_i) - y_i - \varepsilon) + \\ &\quad \mathbb{1}\{y_i - f(x_i) \geq \varepsilon\} \cdot (y_i - f(x_i) - \varepsilon) \\ &= \mathbb{1}\left\{\xi_i^- \geq 0\right\} \cdot \xi_i^- + \mathbb{1}\left\{\xi_i^+ \geq 0\right\} \cdot \xi_i^+ \end{aligned}$$

where $\xi_i^- = f(x_i) - y_i - \varepsilon$ and $\xi_i^+ = y_i - f(x_i) - \varepsilon$. We deduce that the ε-SVM regression problem is:

$$\begin{aligned} \left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}^-, \hat{\xi}^+\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \left(\xi_i^- + \xi_i^+\right) \\ &\text{s.t.} \quad \begin{cases} f(x_i) - y_i \leq \varepsilon + \xi_i^- \\ y_i - f(x_i) \leq \varepsilon + \xi_i^+ \\ \xi_i^- \geq 0 \\ \xi_i^+ \geq 0 \end{cases} \qquad \text{for } i = 1, \ldots, n \end{aligned} \tag{15.33}$$

[55] See Exercise 15.4.8 on page 1028.

We can show that the dual problem is[56]:

$$\begin{aligned}\left\{\hat{\alpha}^-, \hat{\alpha}^+\right\} &= \arg\min \frac{1}{2}\left(\alpha^- - \alpha^+\right)^\top XX^\top\left(\alpha^- - \alpha^+\right) + \\ &\quad \varepsilon\left(\alpha^- + \alpha^+\right)^\top \mathbf{1}_n + \left(\alpha^- - \alpha^+\right)^\top Y \\ &\text{s.t.} \begin{cases} \mathbf{1}_n^\top\left(\alpha^- - \alpha^+\right) = 0 \\ \mathbf{0}_n \leq \alpha^- \leq C \cdot \mathbf{1}_n \\ \mathbf{0}_n \leq \alpha^+ \leq C \cdot \mathbf{1}_n \end{cases}\end{aligned} \tag{15.34}$$

where α^- and α^+ are the Lagrange multipliers of the inequality constraints. We have $\hat{\beta} = \sum_{i=1}^n \left(\hat{\alpha}_i^+ - \hat{\alpha}_i^-\right) x_i$ and:

$$\hat{\beta}_0 = \frac{1}{n_{\mathcal{SV}}}\left(\sum_{i\in\mathcal{SV}^-}\left(y_i + \varepsilon - x_i^\top\hat{\beta}\right) + \sum_{i\in\mathcal{SV}^+}\left(y_i - \varepsilon - x_i^\top\hat{\beta}\right)\right)$$

where $\mathcal{SV}^- = \left\{i : 0 < \hat{\alpha}_i^- < C\right\}$ and $\mathcal{SV}^+ = \left\{i : 0 < \hat{\alpha}_i^+ < C\right\}$ are the set of negative and positive support vectors, and $n_{\mathcal{SV}}$ is the number of support vectors.

TABLE 15.20: Comparison of OLS, LAD and SVM estimates

$\hat{\beta}_k$	OLS	LAD	LS-SVM ($C=1, \varepsilon=1$)	ε-SVM ($C=1, \varepsilon=1$)	LS-SVM ($C=\infty, \varepsilon=0$)	ε-SVM ($C=\infty, \varepsilon=0$)
$\hat{\beta}_0$	3.446	2.331	3.389	3.262	3.446	2.331
$\hat{\beta}_1$	1.544	1.893	1.542	1.631	1.544	1.893
$\hat{\beta}_2$	−1.645	−1.735	−1.616	−1.526	−1.645	−1.735
$\hat{\beta}_3$	2.895	2.908	2.885	2.726	2.895	2.908

We consider Example 100 on page 606, which has been used to illustrate the linear regression. In Table 15.20, we report OLS, LAD and SVM estimates for $C = 1$ and $\varepsilon = 1$. In the last two columns, we consider the limit cases, when the constant C tends to $+\infty$ and ε is equal to zero. We notice that the LS-SVM estimator converges to the OLS estimator. This is quite intuitive since we use a least squares loss function. In some sense, the LS-SVM regression can be seen as a ridge regression. When C tends to $+\infty$, the ridge penalization disappears. More curiously, the ε-SVM estimator converges to the LAD estimator. In fact, the ε-SVM regression is close to a ridge quantile regression. When ε is equal to zero, we obtain a median regression with a L_2 penalization. This is why the ε-SVM estimator converges to the LAD (or median regression) estimator.

Non-linear support vector machines As we have previously seen, we can introduce non-linearity by replacing the input data x by $\phi(x)$, where ϕ is a map from K-dimension to m-dimension non-linear feature space. In the case of SVM, we notice that the dual formulation generally requires the computation of the inner product $\langle x, x'\rangle$. This implies that we can use the same framework by replacing $\langle x, x'\rangle$ by $\langle\phi(x), \phi(x')\rangle$. Manipulating $\phi(x)$ can be tricky and not always obvious[57], because of the high dimension of the non-linear space. Sometimes, it is better to manipulate the inner product, which is called a kernel function $\mathcal{K}(x, x')$. For example, let us consider $x = (x_1, x_2)$ and $\phi(x) = \left(x_1^2, x_1x_2, x_2x_1, x_2^2\right)$. The corresponding kernel function is $\mathcal{K}(x, x') = \langle x, x'\rangle^2$. We also notice that two mapping

[56] See Exercise 15.4.8 on page 1028.

[57] The dimension m is generally much larger than the original dimension.

functions can give the same kernel. For instance, $\mathcal{K}(x,x') = \langle x,x'\rangle^2$ can be generated by $\phi(x) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right)$.

Since we have $\mathcal{K}(x,x') = \phi(x)^\top \phi(x')$, we see that the kernel function is symmetric (Bishop, 2006). This is the main property to define kernel functions. Another way to characterize a kernel is to verify that the Kernel (or Gram) matrix $K = (K_{i,j})$, whose elements are $K_{i,j} = \mathcal{K}(x_i,x_j)$, is positive definite. Therefore, we can directly construct kernels without specifying ϕ. For instance, $e^{c\mathcal{K}}$, $\mathcal{K}+c$, $c\mathcal{K}$ and $\mathcal{K}^d$ are also kernel functions when $c>0$ and $d \in \mathbb{N}$. If $\mathcal{K}_1$ and $\mathcal{K}_2$ are two kernels, the sum $\mathcal{K}_1+\mathcal{K}_2$ and the product $\mathcal{K}_1 \cdot \mathcal{K}_2$ are also kernel functions. The simplest kernel function is obtained by considering the identify function $\phi(x) = x$. It follows that $\langle x,x'\rangle + c$ and $(\langle x,x'\rangle + c)^d$ are also kernel functions. This last one is called the polynomial kernel and is very popular in SVM non-linear classification. Another popular kernel functions are the Gaussian (or radial basis function) kernel[58]:

$$\mathcal{K}(x,x') = \exp\left(-\frac{1}{2\sigma^2}\|x-x'\|_2^2\right)$$

and the neural network (or sigmoid) kernel:

$$\mathcal{K}(x,x') = \tanh\left(c_1\langle x,x'\rangle + c_2\right)$$

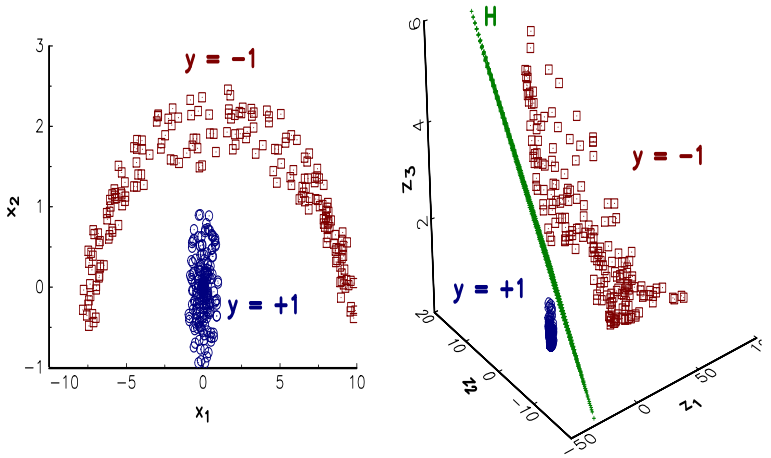


FIGURE 15.33: Transforming a non-linearly separable training set into a linearly separable training set

In order to understand the interest of kernel, we consider a training set[59], which is not linearly separable. In the left panel in Figure 15.33, we have represented the two input variables x_1 and x_2, and the response variable[60] y. Let us apply the polynomial mapping

[58] We can show that dimension of the feature space is infinite: $\phi(x) = (\phi_0(x), \ldots, \phi_s(x), \ldots, \phi_\infty(x))$ where:

$$\phi_s(x) = \left(\frac{1}{\sqrt{s!\sigma^{2s}}} e^{-\frac{x^2}{2\sigma^2}} x^s\right)$$

[59] The data are generated as follows: $x_{1,i} = c_{1,i} + r_{1,i}\cos\theta_i$ and $x_{2,i} = f_i(c_{2,i} + r_{2,i}\sin\theta_i)$ where $\theta_i \sim \mathcal{U}_{[0,2\pi]}$. In the case $y=-1$, we have $c_{1,i} = c_{2,i} = 0$, $r_{1,i} = r_{2,i} \sim \mathcal{U}_{[0,1]}$ and $f_i(x) = x$, otherwise we have $c_{1,i} = 1$, $c_{2,i} = 0$, $r_{1,i} \sim \mathcal{U}_{[8,9]}$, $r_{2,i} \sim \mathcal{U}_{[0,1]}$ and $f_i(x) = |x| - 0.5$.

[60] $y=+1$ corresponds to a circle while $y=-1$ corresponds to a square.

$z = \phi(x) = \left(x_1^2 - 10, \sqrt{2}x_1x_2, x_2^2\right)$. We have reported this transformation in the right panel in Figure 15.33. We observe that the training sets (x, y) and (z, y) are very different, since (z, y) is linearly separable.

All the previous SVM algorithms are valid in the non-linear case and we obtain the following generic framework:

1. the first step consists of defining the mapping function ϕ. Let $z_i = \phi(x_i)$ be the transformed data;
2. in the second step, we calculate the estimated parameters $\hat{\beta}_0$ and $\hat{\beta}$ in the feature space $\mathcal{Z}$;
3. finally, a new observation x is classified by computing $\hat{y} = \text{sign}\left(\hat{\beta}_0 + \phi(x)^\top \hat{\beta}\right)$; in the case of the SVM regression, we have $\hat{y} = \hat{\beta}_0 + \phi(x)^\top \hat{\beta}$.

The previous framework can be simplified by considering the kernel function $\mathcal{K}$ instead of the mapping function ϕ. Indeed, in the dual problems, the input variables are evaluated through the inner product $\langle \phi(x_i), \phi(x_j) \rangle$, that can be replaced[61] by the kernel value $K_{i,j} = \mathcal{K}(x_i, x_j)$. The elements of the Γ matrix used in hard and soft margin QP problem becomes:

$$\begin{aligned} \Gamma_{i,j} &= y_i y_j \phi(x_i)^\top \phi(x_j) \\ &= y_i y_j K_{i,j} \end{aligned}$$

and we have $\Gamma = y \odot y^\top \odot K$ where $K = (k_{i,j})$ is the Gram matrix. Since we have $\hat{\beta} = \sum_{i=1}^n \hat{\alpha}_i y_i \phi(x_i)$ and $\hat{\beta}_0 = \sum_{j \in \mathcal{SV}} \left(y_j - \phi(x_j)^\top \hat{\beta}\right)$, we deduce that $\hat{y} = \text{sign}\, \hat{f}(x)$ where:

$$\begin{aligned} \hat{f}(x) &= \hat{\beta}_0 + \phi(x)^\top \hat{\beta} \\ &= \sum_{j \in \mathcal{SV}} \left(y_j - \phi(x_j)^\top \sum_{i=1}^n \hat{\alpha}_i y_i \phi(x_i) \right) + \sum_{i=1}^n \hat{\alpha}_i y_i \phi(x)^\top \phi(x_i) \\ &= \sum_{j \in \mathcal{SV}} \left(y_j - \sum_{i=1}^n \hat{\alpha}_i y_i \mathcal{K}(x_j, x_i) \right) + \sum_{i=1}^n \hat{\alpha}_i y_i \mathcal{K}(x, x_i) \end{aligned}$$

for a new feature x. The estimation of $\hat{y}$ involves the computation of $\mathcal{K}(x_j, x_i)$ and $\mathcal{K}(x, x_i)$. However, this expression can be reduced because most of the estimates $\hat{\alpha}_i$ are equal to zero.

Remark 194 *The derivation of the SVM non-linear regression is similar to the framework above, because the dual problem involves the computation of* $\phi(X)\phi(X)^\top$*, which is exactly equal to the Gram matrix* K*.*

In Figure 15.33, we have shown that it was possible to transform the data in order to obtain separable training sets. For instance, the hyperplane $\mathcal{H}$, which is estimated using the hard margin classifier, is defined by:

$$0.884 - 0.142 \cdot z_1 + 0.268 \cdot z_2 - 1.422 \cdot z_3 = 0$$

or equivalently:

$$0.884 - 0.142 \cdot \left(x_1^2 - 10\right) + 0.268\sqrt{2} \cdot x_1 x_2 - 1.422 \cdot x_2^2 = 0$$

[61] The fact that we can easily substitute inner products by the Gram matrix in SVM classification and regression is called the kernel trick.

Let us now consider a Monte Carlo simulation. We assume that $X \sim \mathcal{N}(\mathbf{0}_4, I_4)$ and $Y = \text{sign}(\mathcal{N}(0,1))$, meaning that there is no relationship between X and Y. We simulate 300 observations for the training set, and we compute the hard margin classifier for several kernels: linear, quadratic and cubic polynomial with $c = 0$, and RBF ($\sigma = 50$ and $\sigma = 20$). Then, we estimate the predicted value $\hat{y}_i$ for all the observations and calculate the error rate. Since Y is independent from X, the true error rate is equal to 50%, because the score is purely random. Using 500 replications, we have estimated the density function of the error rate in Figure 15.34. We notice that the linear kernel classifier is the worst method, while the RBF kernel with $\sigma = 20$ is the best method. On average, the error rate is respectively equal to 45.0%, 41.5%, 37.7%, 31.8% and 22.3%. Therefore, we have overfitted the model, and this is particularly true with the kernel approach. Indeed, if we consider a validation set, we obtain an average error rate of 50% whatever the kernel function we have used. We conclude that kernel functions are very powerful, but they can lead to large overfitting problems.

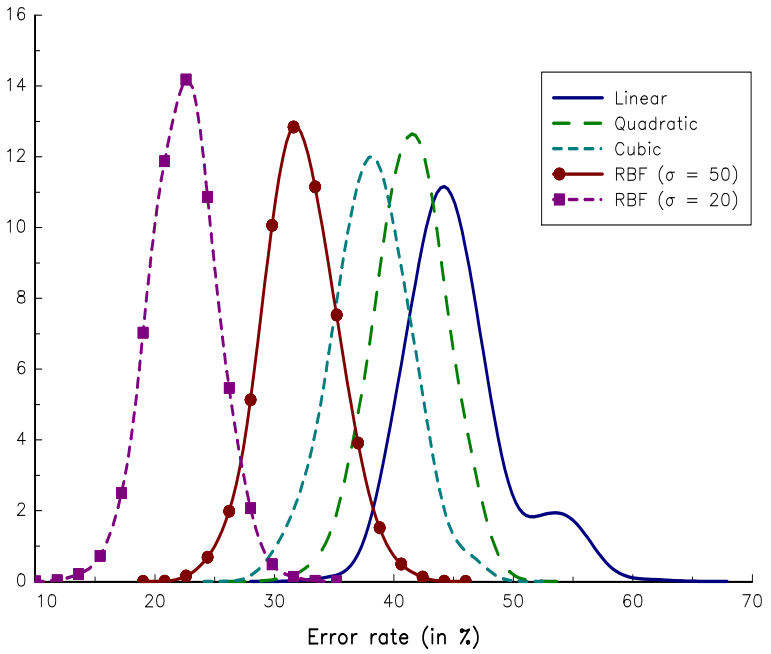


FIGURE 15.34: Probability density function of in-sample error rates

Extension to the multi-class problem We assume that we have n_C disjoint classes $\mathcal{C}_j$ where $j = 1, \ldots, J$. SVMs are inherently two-class classifiers, and the extension to the multi-class problem is not straightforward. However, we distinguish two main approaches. The first approach uses binary classification. In the case of the '*one-against-all*' strategy, we construct J single SVM classifiers in order to separate the training data from every class to the other classes. For the j^{th} classifier, the response variable is then $z_i^{(j)} = +1$ if $y_i \in \mathcal{C}_j$ and $z_i^{(j)} = -1$ if $y_i \notin \mathcal{C}_j$. Using this modified training set, we can estimate the discriminant function $\hat{f}^{(j)}(x) = \hat{\beta}_0^{(j)} + x^\top \hat{\beta}^{(j)}$. In the two-class case, we have $\hat{y} = \text{sign}\, \hat{f}(x)$. In the multi-class problem, the prediction corresponds to the binary classifier that gives the

largest value $\hat{f}^{(j)}(x)$:

$$\hat{y} \in \mathcal{C}_{j^\star} \quad \text{where} \quad j^\star = \arg\max_j \hat{f}^{(j)}(x)$$

Another approach based on the binary classification is called the '*one-against-one*' strategy. In this case, we construct $J(J-1)/2$ single SVM classifiers in order to separate the training data from the class $\mathcal{C}_j$ to the class $\mathcal{C}_{j'}$. Using the estimated discriminant function $\hat{f}^{(j|j')}(x) = \hat{\beta}_0^{(j|j')} + x^\top \hat{\beta}^{(j|j')}$, we can calculate the prediction $\hat{y}^{(j|j')} = \operatorname{sign} \hat{f}^{(j|j')}(x)$. The empirical probability that the observation belongs to the class $\mathcal{C}_j$ is then equal to[62]:

$$\hat{p}_j(x) = \frac{2\sum_{j'=1}^{J} \mathbb{1}\left\{\hat{y}^{(j|j')} = +1\right\}}{J(J-1)}$$

We deduce that the classification rule is defined as follows:

$$\hat{y} \in \mathcal{C}_{j^\star} \quad \text{where} \quad j^\star = \arg\max_j \hat{p}_j(x)$$

The second approach of multi-classification extends the mathematical framework of SVM that has been developed for the binary classification. The idea is then to consider a function $y = f(x) : \mathbb{R}^K \to \{1, \ldots, J\}$ where:

$$f(x) = \arg\max_j \beta_0^{(j)} + x^\top \beta^{(j)}$$

We have now to estimate the $J \times 1$ vector β_0 and the $K \times J$ matrix β. Crammer and Singer (2001) developed both hard and soft margin primal and dual problems in an elegant way. For a review and a comparison of these different methods, the reader can refer to Hsu and Lin (2002).

15.2.3.4 Model averaging

Model averaging (or ensemble averaging) combines multiple learning algorithms to obtain better predictive performance than could be obtained from the individual models. Two types of approaches are generally used. The first one constructs a family of '*random*' models (bagging/random forests), whereas the second one generates a family of '*adaptive*' models (boosting).

The motivation of model averaging is to replace a single expert by a committee of experts. Sometimes, it is difficult to find a skilled expert, or his search has a large cost. In this case, we can imagine that the work produced by this high skilled expert can be done by a committee of less skilled experts. For establishing the committee, we can choose (randomly) experts with similar skills or we can choose experts that are complementary. The parallel with model averaging is obvious when we distinguish random and adaptive models.

Bagging (bootstrap aggregation) Breiman (1996) proposed to use the bootstrap method to improve the performance of weak learners, in particular to reduce their variance and the overfitting bias. Given a training set $\mathcal{Z} = \{(x_i, y_i), i = 1, \ldots, n\}$, the bagging method generates n_S bootstrapped training sets $\mathcal{Z}_{(s)}$ and estimates the output function

[62] We have $\hat{y}^{(j|j')} = +1 \Leftrightarrow \hat{y}^{(j'|j)} = -1$.

$\hat{f}_{(s)}(x)$ for each training set. The s^{th} model is then defined by the pair $\left(\mathcal{Z}_{(s)}, \hat{f}_{(s)}\right)$. In the case of regression, the predicted value is the mean of the predicted values of the different models:

$$\hat{y} = \frac{1}{n_S} \sum_{s=1}^{n_S} \hat{f}_{(s)}(x)$$

In the case of classification, we generally implement the majority vote rule:

$$\hat{y} = \text{MaxVote}\left(\hat{f}_{(1)}(x), \ldots, \hat{f}_{(n_S)}(x)\right)$$

In this approach, the predictions of each model are considered as a '*vote*'. The final prediction corresponds to the class that has the maximum number of votes for multi-classification, or the majority vote for binary classification. As shown by Breiman (1996), the bagging method makes only sense when we consider non-linear models.

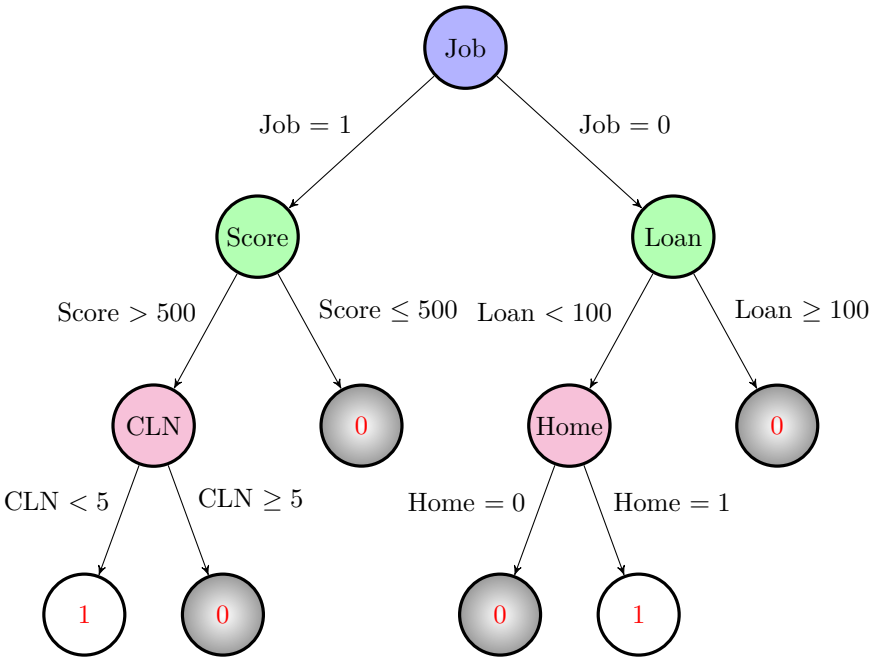


FIGURE 15.35: An example of decision tree

The bagging method is extensively used when considering decision trees. A tree is represented by a series of binary splits. Each node represents a query, except the terminal nodes that correspond to the decision nodes. In the case of a classification tree, the output variable takes a discrete set of class labels, whereas the output variable takes continuous values when considering regression trees. In Figure 15.35, we report an example of a classification tree. We consider an applicant that would like a new credit. If the applicant has not a job, the credit will be automatically refused if the amount of the loan is too high. If the amount of the loan is less than 100, the final decision will depend upon whether the applicant owns his house. In this case, the client can obtain the credit if he applies for a mortgage or a home equity line of credit. If the applicant has a job, the bank computes his credit score. If the score is less than 500, the credit is rejected. Otherwise, the final decision will depend on the number of credits. If the applicant has less than 5 credits, the new credit is accepted,

otherwise it is refused. Decision trees are very popular in credit scoring for three main reasons. First, they can handle different types of variables (numeric, continuous, discrete, qualitative, etc.). Second, the rules and the decision process are very easy to understand. Third, they can be estimated with statistical models, and adjusted by experts. In practice, we use greedy approaches based on recursive binary splitting algorithms. One drawback of classification trees is that their prediction power is generally lower than the ones observed with logistic models, neural networks or support vector machines. We generally say that they produce weak classifiers (or learners). However, by combining classification trees and bagging, we can obtain the same performance than strong classifiers (Hastie *et al.*, 2009).

Remark 195 *Bagging is also extensively used when we have a large set of predictors. Instead of running one logistic regression with all the input variables, we can estimate many logit models with a limited number of explanatory variables (e.g. less than 10). In this approach, the bootstrap procedure concerns the variables, not the observations. By construction, the bagging model will produce better and more stable predictions than the single logit model*[63].

Random forests Let $\hat{Y}_{(s)} = \hat{f}_{(s)}(X)$ be the output random variable produced by the s^{th} bootstrapped model. If we assume that $\hat{Y}_{(1)}, \ldots, \hat{Y}_{(n_S)}$ are *iid* random variables with mean μ and variance σ^2, we have:

$$\operatorname{var}\left(\hat{Y}\right) = \operatorname{var}\left(\frac{1}{n_S}\sum_{s=1}^{n_S}\hat{Y}_{(s)}\right) = \frac{\sigma^2}{n_s}$$

where $\hat{Y}$ is the bagging estimator. We deduce that $\operatorname{var}\left(\hat{Y}\right) \to 0$ when $n_S \to \infty$. Theoretically, the bagging method can highly reduce the variance of the prediction. However, the hypothesis that $\hat{Y}_{(1)}, \ldots, \hat{Y}_{(n_S)}$ are not correlated is too strong. If we assume that the average correlation between bootstrapped models is equal to ρ, we obtain:

$$\begin{aligned}
\operatorname{var}\left(\hat{Y}\right) &= \mathbb{E}\left[\left(\frac{1}{n_S}\sum_{s=1}^{n_S}\left(\hat{Y}_{(s)} - \mu\right)\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n_S^2}\sum_{s=1}^{n_S}\left(\hat{Y}_{(s)} - \mu\right)^2 + \frac{1}{n_S^2}\sum_{r\neq s}\left(\hat{Y}_{(r)} - \mu\right)\left(\hat{Y}_{(s)} - \mu\right)\right] \\
&= \frac{n_S\sigma^2}{n_S^2} + \frac{n_S\left(n_S - 1\right)\rho\sigma^2}{n_S^2} \\
&= \rho\sigma^2 + \frac{1-\rho}{n_s}\sigma^2
\end{aligned}$$

It follows that $\operatorname{var}\left(\hat{Y}\right) > \rho\sigma^2$. For example, if $\rho = 90\%$, the maximum reduction of the variance is only 10%. It follows that the improvement due to the bagging method can be highly limited when the correlation is high. Breiman (2001) proposed a modification of bagging by building de-correlated trees. At each iteration s, we select randomly a subset of predictors $\mathcal{X}_{(s)}$, implying that the model is then defined by the 3-tuple $\left(\mathcal{Z}_{(s)}, \mathcal{X}_{(s)}, \hat{f}_{(s)}\right)$. Generally, the randomization step is done with a fixed number $K^\star$ of bootstrapped predictors[64].

[63] For instance, if we consider the degenerate case when the number of observations is lower than the number of predictors ($n < K$), the single logit model is highly noisy, which is not the case of the bagging model.

[64] The recommended default value is $K^\star = \sqrt{K}$ for classification and $K^\star = K/3$ for regression (Hastie *et al.*, 2009).

Remark 196 *The method of random forests can be viewed as a double bagging method. Indeed, it mixes observation-based and feature-based bagging methods.*

Boosting In this approach, the training set $(\mathcal{Z}, \mathcal{W})$ is defined by including the weight of each observation:

$$(\mathcal{Z}, \mathcal{W}) = \{(x_i, y_i, w_i), i = 1, \ldots, n\}$$

At each iteration s, boosting computes adaptive weights $\mathcal{W}_{(s)}$ and fits the learning algorithm $\hat{f}_{(s)}$ with the training set $\left(\mathcal{Z}, \mathcal{W}_{(s)}\right)$. Then, it combines the different learning models through a weighting rule:

$$\hat{y} = \hat{f}(x) = \text{Avg}\left(\omega_s \cdot \hat{f}_{(s)}(x)\right)_{s=1}^{n_S}$$

where ω_s is the weight of the s^{th} learning model and Avg is the averaging function. In the case of a binary classification, we have:

$$\hat{y} = \hat{f}(x) = \text{sign}\left(\sum_{s=1}^{n_S} \omega_s \hat{f}_{(s)}(x)\right)$$

The concept of boosting has been introduced by Schapire (1990), who proved that a '*weak*' learning algorithm can be '*boosted*' into a '*strong*' learning algorithm[65]. In the 1990s, many boosting algorithms have been developed, but the high recognition comes with the adaptive boosting method proposed by Freund and Schapire (1997), and described in Algorithm 2.

The algorithm concerns the classification problem $y \in \{-1, +1\}$. We begin by initializing the observation weights w_i to $1/n$. Then, we fit the classifier $\hat{f}_{(1)}$ using the training set $\mathcal{Z}$, because the initial weights have no impact. The first step is the usual manner to fit a classification model. To improve the accuracy, boosting constructs at iteration s another training set by calculating new observation weights:

$$w_{i,s+1} = \begin{cases} w_{i,s} & \text{if } i \text{ is well-classified} \\ w_{i,s} e^{\omega_s} & \text{otherwise} \end{cases}$$

If the observation i is well-classified, the weight remain the same, otherwise it increases: $w_{i,s+1} > w_{i,s}$. Indeed, the update makes only sense if the error rate $\mathcal{L}_{(s)}$ is smaller than 50%, implying that ω_s is strictly positive[66]. At iteration $s+1$, the classifier will be fitted with the training set $\left(\mathcal{Z}, \mathcal{W}_{(s+1)}\right)$, where the misclassified observations at iteration s are more weighted than the well-classified observations. Therefore, the weighting scheme $\mathcal{W}_{(s+1)}$ forces the new classifier $\hat{f}_{(s+1)}$ to be more focus on the training observations that are difficult to classify. Finally, we use the majority vote to predict y:

$$\hat{y} = \text{sign}\left(\sum_{s=1}^{n_S} \omega_s \cdot \hat{f}_{(s)}(x)\right)$$

We represent the classifier weight ω_s with respect to the loss function (or the error rate) in Figure 15.36. If the error rate is equal to 50%, the weight ω_s of the s^{th} classifier is equal to zero. This classifier does not participate to the final model, because it corresponds to a random guessing model. On the contrary, if the error rate of one classifier is equal to zero, its allocation is infinite in the final model. In Figure 15.36, we also show the impact of the

[65] A training set is said to be strongly learnable if "*there exists a polynomial-time algorithm that achieves low error with high confidence*" for all the observations (Schapire, 1990). A weak learning algorithm performs just slightly better than a random learning algorithm.

[66] If the classifier has an error rate greater than 50%, it performs worse than random guessing.

Algorithm 2 AdaBoost.M1 binary classifier

Estimate the AdaBoost.M1 classifier $\hat{y} = \hat{f}(x)$
Initialize the observation weights $w_{i,1} = 1/n$ for $i = 1, \ldots, n$
for $s = 1 : n_S$ **do**
$\mathcal{W}_{(s)} \leftarrow (w_{1,s}, \ldots, w_{n,s})$
Fit the classifier $\hat{f}_{(s)}$ using the training set $\left(\mathcal{Z}, \mathcal{W}_{(s)}\right)$
Compute the loss function:

$$\mathcal{L}_{(s)} = \frac{\sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq \hat{f}_{(s)}(x_i)\right\}}{\sum_{i=1}^{n} w_{i,s}}$$

Calculate the classifier weight ω_s:

$$\omega_s \leftarrow \ln\left(\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}\right)$$

Update the observation weights:

$$w_{i,s+1} \leftarrow w_{i,s} e^{\omega_s \cdot \mathbb{1}\left\{y_i \neq \hat{f}_{(s)}(x_i)\right\}}$$

Normalize the observation weights:

$$w_{i,s+1} \leftarrow \frac{w_{i,s+1}}{\sum_{i'=1}^{n} w_{i',s+1}}$$

end for
return $\hat{f}(x) = \text{sign}\left(\sum_{s=1}^{n_S} \omega_s \hat{f}_{(s)}(x)\right)$

error rate on the weights $w_{i,s+1}$ when we consider a sample of two observations. We assume that the first observation is misclassified while the second observation is well-classified at the step s. This implies that the first observation will have more weight at the step $s+1$. The re-weighting of observations also depends on the error rate. If the error rate of the s^{th} model is low, the re-weighting is strong, in order to separate well-classified and misclassified observations. It is not obvious that the error rate is a monotonous function of the iteration s. At the beginning, the error rate can increase or decrease depending whether the initial classifier is good or bad. But, at the end, the error rate must reach the upper bound 50%.

In order to illustrate the boosting method, we consider the data given in Table 15.21 and the logit model:

$$\Pr\{y_i = 1\} = \mathbf{F}(\beta_0 + \beta_1 x_i)$$

where $\mathbf{F}(x)$ is the logit function. We have $\Pr\{y_i = -1\} = 1 - \mathbf{F}(\beta_0 + \beta_1 x_i)$. Given the pattern x, the classification rule is then:

$$\hat{y} = 2 \cdot \mathbb{1}\left\{\mathbf{F}\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) > \frac{1}{2}\right\} - 1$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the parameters, which have been estimated by the method of maximum likelihood. Using our data, we obtain $\hat{\beta}_0 = 0.4133$ and $\hat{\beta}_1 = 0.2976$, and the error rate is equal to 45%. In the case of the boosting algorithm, the first iteration is exactly the same as the previous logit estimation. For the second iteration, we have to calculate the weights $w_{i,2}$ of each observation. We have $\omega_1 = 0.2007$ because $\mathcal{L}_{(1)} = 45\%$. Therefore, we update

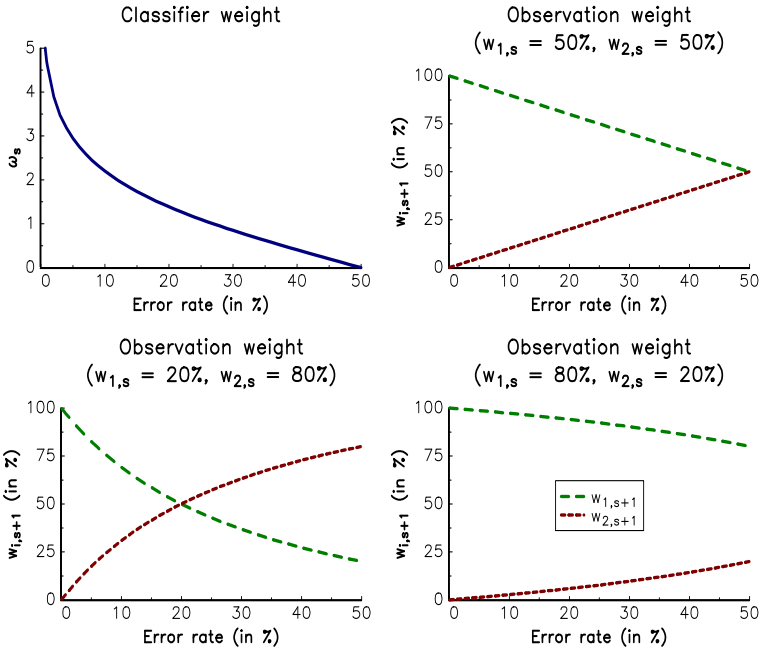


FIGURE 15.36: Weighting schemes of the boosting approach

the weights. In Table 15.21, we have reported the predicted value $\hat{y}_{i,s}$ at the iteration step s, and also the variable $\vartheta_{i,s}$ which indicates the misclassified observations. For instance, observations 3, 4, 5, 7, 9, 10, 12, 16 and 17 are not well-classified at the first iteration by the logit model. This is why the weight of these observations increase by e^{ω_1}. While the weights $w_{i,1}$ take the uniform value of 5%, the weights $w_{i,2}$ are different with respect to observations. After normalizing, $w_{i,2}$ is equal to 5.56% for observations that are misclassified at the first iteration, otherwise it is equal to 4.55%. Using these weights, we estimate the logit model, and found $\hat{\beta}_{0,2} = 0.2334$ and $\hat{\beta}_{1,2} = 0.2558$ (Table 15.22). The loss function is then equal to $\mathcal{L}_{(2)} = 38.89\%$. We see that the second logit model has improved the classification for two observations ($i = 3$ and $i = 10$). We can continue the algorithm. In our example, the boosting method stops after 5 iterations, because $\mathcal{L}_{(5)} = 50.00\%$. The fifth estimated classifier is then a pure random guessing model. While the number of well-classified observations is equal to 11 for the logit model, it is equal to 13 for the boosting model[67]. From a general point of view, the boosting is interesting only if we use a large dataset of observations and variables. When considering small datasets, we face an obvious overfitting issue.

Remark 197 *The boosting method is based on weighted estimation methods. In Chapter 10, we have already defined the weighted least squares estimator*[68]*. In Exercise 15.4.10 on page 1029, we extend the method of maximum likelihood, neural networks and support vector machines when observations are weighted.*

Hastie *et al.* (2009) showed that boosting is related to additive models:

$$g(x) = \sum_{s=1}^{n_S} \beta_{(s)} \mathcal{B}\left(x; \gamma_{(s)}\right)$$

[67]The boosting classifier corresponds to the column $\hat{y}_i$ in Table 15.21.
[68]See Section 10.1.1.5 on page 612.

TABLE 15.21: Illustration of the boosting algorithm ($n_S = 2$)

i	y_i	x_i	$w_{i,1}$ (in %)	$\hat{y}_{i,1}$	$\vartheta_{i,1}$	$w_{i,2}$ (in %)	$\hat{y}_{i,2}$	$\vartheta_{i,2}$	$\hat{y}_i$	ϑ_i
1	1	0.597	5.00	1		4.55	1		1	
2	1	1.496	5.00	1		4.55	1		1	
3	−1	−0.914	5.00	1	✓	5.56	−1		−1	
4	−1	−0.497	5.00	1	✓	5.56	1	✓	1	✓
5	−1	0.493	5.00	1	✓	5.56	1	✓	1	✓
6	1	0.841	5.00	1		4.55	1		1	
7	−1	−0.885	5.00	1	✓	5.56	1	✓	1	✓
8	1	1.418	5.00	1		4.55	1		1	
9	−1	−0.183	5.00	1	✓	5.56	1	✓	1	✓
10	−1	−1.298	5.00	1	✓	5.56	−1		−1	
11	1	−0.324	5.00	1		4.55	1		1	
12	1	−1.454	5.00	−1	✓	5.56	−1	✓	−1	✓
13	1	−0.270	5.00	1		4.55	1		1	
14	1	−0.770	5.00	1		4.55	1		1	
15	1	0.232	5.00	1		4.55	1		1	
16	−1	0.970	5.00	1	✓	5.56	1	✓	1	✓
17	−1	1.196	5.00	1	✓	5.56	1	✓	1	✓
18	1	0.578	5.00	1		4.55	1		1	
19	1	−0.686	5.00	1		4.55	1		1	
20	1	−0.590	5.00	1		4.55	1		1	

TABLE 15.22: Estimated model at each boosting iteration ($n_S = 5$)

s	1	2	3	4	5
$\hat{\beta}_{0,s}$	0.4133	0.2334	−0.0771	0.0009	0.0103
$\hat{\beta}_{1,s}$	0.2976	0.2558	0.0278	0.0277	−0.0751
$\mathcal{L}_{(s)}$	0.4500	0.3889	0.4805	0.4741	0.5000
ω_s	0.2007	0.4520	0.0780	0.1038	0.0000

where $\beta_{(s)}$ is the expansion coefficient and $\mathcal{B}\left(x;\gamma_{(s)}\right)$ is the basis function at iteration s. Forward stagewise regression consists in finding the optimal values of $\hat{\beta}_{(s)}$ and $\hat{\gamma}_{(s)}$:

$$\left(\hat{\beta}_{(s)},\hat{\gamma}_{(s)}\right)=\arg\min\sum_{i=1}^{n}\mathcal{L}\left(y_i,\hat{g}_{(s-1)}\left(x_i\right)+\beta_{(s)}\mathcal{B}\left(x;\gamma_{(s)}\right)\right)$$

where $\hat{g}_{(s)}\left(x_i\right)=\sum_{s'=1}^{s}\hat{\beta}_{(s')}\mathcal{B}\left(x;\hat{\gamma}_{(s')}\right)$ and $\mathcal{L}$ is the loss function. In the case of boosting, we can show that $\mathcal{B}\left(x;\gamma_{(s)}\right)=\hat{f}_{(s)}\left(x\right)$, $\hat{\beta}_{(s)}=\omega_s$ and $\mathcal{L}\left(y,f\left(x\right)\right)=e^{-yf(x)}$. We recognize an additive logit model with the softmax loss function. Using this framework, Friedman (2002) proposed gradient boosting models. The idea is to minimize the loss function $\sum_{i=1}^{n}\mathcal{L}\left(y_i,f\left(x_i\right)\right)$ with respect to the learning algorithm $f\left(x\right)$. The steepest descend algorithm consists in the following iterations:

$$\hat{f}_{(s)}\left(x_i\right)=\hat{f}_{(s-1)}\left(x_i\right)-\eta_{(s)}\frac{\partial\mathcal{L}\left(y_i,\hat{f}_{(s-1)}\left(x_i\right)\right)}{\partial f\left(x_i\right)}$$

Instead of finding the optimal classifier $\hat{f}_{(s)}$, gradient boosting estimates the optimal step $\eta_{(s)}$ and iterates the previous formula. Finally, the optimal model $\hat{f}\left(x\right)$ is given by the estimate $\hat{f}_{(n_S)}\left(x\right)$ at the last iteration.

Remark 198 *The table below summarizes the differences between bagging, random forests and boosting:*

Method	*Model definition*	$\mathcal{Z}_{(s)}$	$\mathcal{X}_{(s)}$	$\mathcal{W}_{(s)}$	*Weighted average*
Bagging	$\left(\mathcal{Z}_{(s)},\hat{f}_{(s)}\right)$	✓			
Random forests	$\left(\mathcal{Z}_{(s)},\mathcal{X}_{(s)},\hat{f}_{(s)}\right)$	✓	✓		
Boosting	$\left(\mathcal{Z},\mathcal{W}_{(s)},\hat{f}_{(s)}\right)$			✓	✓

In the bagging method, the randomization step concerns observations. In the case of random forests, the models are generated by randomizing both observations and variables. Boosting is a very different approach, since all the observations and variables are used to construct the weak learning models. In this method, the perturbations are introduced by using a weighting scheme for the observations that changes at each iteration. The randomization step is then replaced by an adaptive step, where the $(s+1)^{\text{th}}$ *model depends on the accuracy of the* s^{th} *model. Finally, boosting uses a weighted average of the different weak learning algorithms.*

15.3 Performance evaluation criteria and score consistency

This section is dedicated to the performance assessment of a score. Using information theory, we would like to know if the scoring system is informative or not. The second paragraph presents the graphical tools in order to measure the classification accuracy of the score. Finally, we define the different statistical measures to estimate the performance of the score. We also notice that the tools presented here can be used with both the training set or the validation set.

15.3.1 Shannon entropy

15.3.1.1 Definition and properties

The entropy is a measure of unpredictability or uncertainty of a random variable. Let (X, Y) be a random vector where $p_{i,j} = \Pr\{X = x_i, Y = y_j\}$, $p_i = \Pr\{X = x_i\}$ and $p_j = \Pr\{Y = y_j\}$. The Shannon entropy of the discrete random variable X is given by[69]:

$$H(X) = -\sum_{i=1}^{n} p_i \ln p_i$$

We have the property $0 \leq H(X) \leq \ln n$. H is equal to zero if there is a state i such that $p_i = 1$ and is equal to $\ln n$ in the case of the uniform distribution ($p_i = 1/n$). The Shannon entropy is a measure of the average information of the system. The lower the Shannon entropy, the more informative the system. For a random vector (X, Y), we have:

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{i,j} \ln p_{i,j}$$

We deduce that the conditional information of Y given X is equal to:

$$\begin{aligned} H(Y \mid X) &= \mathbb{E}_X[H(Y \mid X = x)] \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{i,j} \ln \frac{p_{i,j}}{p_i} \\ &= H(X, Y) - H(X) \end{aligned}$$

We have the following properties:

- if X and Y are independent, we have $H(Y \mid X) = H(Y)$ and $H(X, Y) = H(Y) + H(X)$;
- if X and Y are perfectly dependent, we have $H(Y \mid X) = 0$ and $H(X, Y) = H(X)$.

The amount of information obtained about one random variable, through the other random variable is measured by the mutual information:

$$\begin{aligned} I(X, Y) &= H(Y) + H(X) - H(X, Y) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} p_{i,j} \ln \frac{p_{i,j}}{p_i p_j} \end{aligned}$$

Figure 15.37 shows some examples of Shannon entropy calculation. For each example, we indicate the probabilities $p_{i,j}$ and the values taken by $H(X)$, $H(Y)$, $H(X, Y)$ and $I(X, Y)$. The top/left panel corresponds to a diffuse system. The value of $H(X, Y)$ is maximum, meaning that the system is extremely disordered. The top/right panel represents a highly ordered system in the bivariate case and a diffuse system in the univariate case. We have $H(X \mid Y) = H(Y \mid X) = 0$, implying that the knowledge of X is sufficient to find the state of Y. Generally, the system is not perfectly ordered or perfectly disordered. For instance, in the case of the system described in the bottom/left panel, the knowledge of X informs us about the state of Y. Indeed, if X is in the third state, then we know that Y cannot be in the first or sixth state. Another example is provided in the bottom/right panel.

Remark 199 *If we apply the Shannon entropy to the transition matrix of a Markov chain, we set $X = \mathfrak{R}(s)$ and $Y = \mathfrak{R}(t)$ where $\mathfrak{R}(t)$ is the state variable at the date t. We obtain:*

$$H(\mathfrak{R}(t) \mid \mathfrak{R}(s)) = -\sum_{i=1}^{K} \pi_i^\star \sum_{j=1}^{K} p_{i,j}^{(t-s)} \ln p_{i,j}^{(t-s)}$$

where $p_{i,j} = \Pr\{\mathfrak{R}(t+1) = j \mid \mathfrak{R}(t) = i\}$, $\mathcal{S} = \{1, 2, \ldots, K\}$ is the state space of the Markov chain and $\pi^\star$ is the associated stationary distribution.

[69] We use the convention $p_i \ln p_i = 0$ when p_i is equal to zero.

1/36	1/36	1/36	1/36	1/36	1/36
1/36	1/36	1/36	1/36	1/36	1/36
1/36	1/36	1/36	1/36	1/36	1/36
1/36	1/36	1/36	1/36	1/36	1/36
1/36	1/36	1/36	1/36	1/36	1/36
1/36	1/36	1/36	1/36	1/36	1/36

$$H(X) = H(Y) = 1.792$$
$$H(X,Y) = 3.584$$
$$I(X,Y) = 0$$

1/6					
	1/6				
		1/6			
			1/6		
				1/6	
					1/6

$$H(X) = H(Y) = 1.792$$
$$H(X,Y) = 1.792$$
$$I(X,Y) = 1.792$$

1/24	1/24				
1/24	1/24	1/24	1/48		
	1/24	1/6	1/24	1/48	
	1/48	1/24	1/6	1/24	
		1/48	1/24	1/24	1/24
				1/24	1/24

$$H(X) = H(Y) = 1.683$$
$$H(X,Y) = 2.774$$
$$I(X,Y) = 0.593$$

					1/12
1/8			1/8		
	1/24				
5/24		1/24			
3/24				1/24	
3/24	1/24	1/24			

$$H(X) = 1.658$$
$$H(Y) = 1.328$$
$$I(X,Y) = 0.750$$

FIGURE 15.37: Examples of Shannon entropy calculation

15.3.1.2 Application to scoring

Let S and Y be the score and the control variable. For instance, Y is a binary random variable that may indicate a bad credit ($Y = 0$) or a good credit ($Y = 1$). Y may also correspond to classes defined by some quantiles. With Shannon entropy, we can measure the information of the system (S, Y). We can also compare two scores S_1 and S_2 by using the statistical measures $I(S_1, Y)$ and $I(S_2, Y)$. Let S_3 be the aggregated score obtained from the two individual scores S_1 and S_2. We can calculate the information contribution of each score with respect to the global score. Therefore, we can verify that a score really adds an information.

We consider the following decision rule:

$$\begin{cases} S \leq 0 \Rightarrow S^\star = 0 \\ S > 0 \Rightarrow S^\star = 1 \end{cases}$$

We note $n_{i,j}$ the number of observations such that $S^\star = i$ and $Y = j$. We obtain the following system $(S^\star, Y)$:

	$Y=0$	$Y=1$
$S^\star=0$	$n_{0,0}$	$n_{0,1}$
$S^\star=1$	$n_{1,0}$	$n_{1,1}$

where $n = n_{0,0} + n_{0,1} + n_{1,0} + n_{1,1}$ is the total number of observations. The hit rate is the ratio of good bets:

$$H = \frac{n_{0,0} + n_{1,1}}{n}$$

This statistic can be viewed as an information measure of the system (S, Y). When there are more states, we can consider the Shannon entropy. In Figure 15.38, we report the contingency table of two scores S_1 and S_2 for 100 observations[70]. We have $I(S_1, Y) = 0.763$ and $I(S_2, Y) = 0.636$. We deduce that S_1 is more informative than S_2.

	y_1	y_2	y_3	y_4	y_5
s_1	10	9			
s_2	7	9			
s_3	3		7	2	
s_4		2	10	4	5
s_5				10	2
s_6			3	4	13

$H(S_1) = 1.767$
$H(Y) = 1.609$
$H(S_1, Y) = 2.614$
$I(S_1, Y) = 0.763$

	y_1	y_2	y_3	y_4	y_5
s_1	7	10			
s_2	10	8			
s_3			5	4	3
s_4	3		10	6	4
s_5	2			5	8
s_6			5	5	5

$H(S_1) = 1.771$
$H(Y) = 1.609$
$H(S_1, Y) = 2.745$
$I(S_1, Y) = 0.636$

FIGURE 15.38: Scorecards S_1 and S_2

15.3.2 Graphical methods

We assume that the control variable Y can takes two values: $Y = 0$ corresponds to a bad risk (or bad signal) while $Y = 1$ corresponds to a good risk (or good signal). Gouriéroux (1992) introduced 3 graphical tools for assessing the quality of a score: the performance curve, the selection curve and the discrimination curve[71]. In the following, we assume that the probability $\Pr\{Y = 1 \mid S \geq s\}$ is increasing with respect to the level $s \in [0, 1]$, which corresponds to the rate of acceptance. We deduce that the decision rule is the following:

- if the score of the observation is above the threshold s, the observation is selected;
- if the score of the observation is below the threshold s, the observation is not selected.

[70] Each score is divided into 6 intervals $(s_1, \ldots, s_6)$ while the dependent variable is divided into 5 intervals $(y_1, \ldots, y_5)$.
[71] See also Gouriéroux and Jasiak (2007).

If s is equal to one, we select no observation. If s is equal to zero, we select all the observations. In a scoring system, the threshold s is given. Below, we assume that s is varying and we analyze the relevance of the score with respect to this parameter.

15.3.2.1 Performance curve, selection curve and discriminant curve

The performance curve is the parametric function $y = \mathcal{P}(x)$ defined by:

$$\begin{cases} x(s) = \Pr\{S \geq s\} \\ y(s) = \dfrac{\Pr\{Y = 0 \mid S \geq s\}}{\Pr\{Y = 0\}} \end{cases}$$

where $x(s)$ corresponds to the proportion of selected observations and $y(s)$ corresponds to the ratio between the proportion of selected bad risks and the proportion of bad risks in the population. The score is efficient if the ratio is below one. If $y(s) > 1$, the score selects more bad risks than those we can find in the population[72]. If $y(s) = 1$, the score is random and the performance is equal to zero. In this case, the selected population is representative of the total population.

The selection curve is the parametric curve $y = \mathcal{S}(x)$ defined by:

$$\begin{cases} x(s) = \Pr\{S \geq s\} \\ y(s) = \Pr\{S \geq s \mid Y = 0\} \end{cases}$$

where $y(s)$ corresponds to the ratio of observations that are wrongly selected. By construction, we would like that the curve $y = \mathcal{S}(x)$ is located below the bisecting line $y = x$ in order to verify that $\Pr\{S \geq s \mid Y = 0\} < \Pr\{S \geq s\}$.

Remark 200 *The performance and selection curves are related as follows*[73]*:*

$$\mathcal{S}(x) = x\mathcal{P}(x)$$

The discriminant curve is the parametric curve $y = \mathcal{D}(x)$ defined by:

$$\mathcal{D}(x) = g_1\left(g_0^{-1}(x)\right)$$

where:

$$g_y(s) = \Pr\{S \geq s \mid Y = y\}$$

It represents the proportion of good risks in the selected population with respect to the proportion of bad risks in the selected population. The score is said to be discriminant if the curve $y = \mathcal{D}(x)$ is located above the bisecting line $y = x$.

[72] In this case, we have $\Pr\{Y = 0 \mid S \geq s\} > \Pr\{Y = 0\}$.

[73] We have:

$$\begin{aligned} \Pr\{S \geq s \mid Y = 0\} &= \frac{\Pr\{S \geq s, Y = 0\}}{\Pr\{Y = 0\}} \\ &= \Pr\{S \geq s\} \cdot \frac{\Pr\{S \geq s, Y = 0\}}{\Pr\{S \geq s\}\Pr\{Y = 0\}} \\ &= \Pr\{S \geq s\} \cdot \frac{\Pr\{Y = 0 \mid S \geq s\}}{\Pr\{Y = 0\}} \end{aligned}$$

15.3.2.2 Some properties

We first notice that the previous parametric curves do not depend on the probability distribution of the score S, but only on the ranking of the observations. They are then invariant if we apply an increasing function to the score. Gouriéroux (1992) also established the following properties:

1. the performance curve (respectively, the selection curve) is located below the line $y = 1$ (respectively, the bisecting line $y = x$) if and only if $\operatorname{cov}(f(Y), g(S)) \geq 0$ for any increasing functions f and g;
2. the performance curve is increasing if and only if:
$$\operatorname{cov}(f(Y), g(S) \mid S \geq s) \geq 0$$
for any increasing functions f and g, and any threshold level s;
3. the selection curve is convex if and only if $\mathbb{E}[f(Y) \mid S = s]$ is increasing with respect to the threshold level s for any increasing function f.

Remark 201 *The first property is the least restrictive. It allows us to verify that the score S is better than a random score. We can show that* (3) $\Rightarrow$ (2) $\Rightarrow$ (1). *The last property is then the most restrictive.*

A score is perfect or optimal if there is a threshold level $s^\star$ such that $\Pr\{Y = 1 \mid S \geq s^\star\} = 1$ and $\Pr\{Y = 0 \mid S < s^\star\} = 1$. It separates the population between good and bad risks. Graphically, the selection curve of a perfect score is equal to:

$$y = \mathbb{1}\{x > \Pr\{Y = 1\}\} \cdot \left(1 + \frac{x - 1}{\Pr\{Y = 0\}}\right)$$

Using the relationship $\mathcal{S}(x) = x\mathcal{P}(x)$, we deduce that the performance curve of a perfect score is given by:

$$y = \mathbb{1}\{x > \Pr\{Y = 1\}\} \cdot \left(\frac{x - \Pr\{Y = 1\}}{x \cdot \Pr\{Y = 0\}}\right)$$

For the discriminant curve, a perfect score satisfies $\mathcal{D}(x) = 1$. When the score is random, we have $\mathcal{S}(x) = \mathcal{D}(x) = x$ and $\mathcal{P}(x) = 1$. In Figure 15.39, we have reported the performance, selection and discriminant curves of a given score S. We also show the curves obtained with an optimal (or perfect) score and a random score. A score must be located in the area between the curve computed with a random score and the curve computed with a perfect score, except if the score ranks the observations in a worst way than a random score.

Gouriéroux (1992) also established two properties for comparing two scores S_1 and S_2:

- the score S_1 is more performing on the population P_1 than the score S_2 on the population P_2 if and only if the performance (or selection) curve of (S_1, P_1) is below the performance (or selection) curve of (S_2, P_2);
- the score S_1 is more discriminatory on the population P_1 than the score S_2 on the population P_2 if and only if the discriminant curve of (S_1, P_1) is above the discriminant curve of (S_2, P_2).

Figure 15.40 illustrates the case where the score S_1 is better than the score S_2. However, the order is only partial. Most of the time, the two scores cannot be globally compared. An example is provided in Figure 15.41. The second score is not very good to distinguish good and bad risks when it takes small values, but it is close to a perfect score when it takes high values.

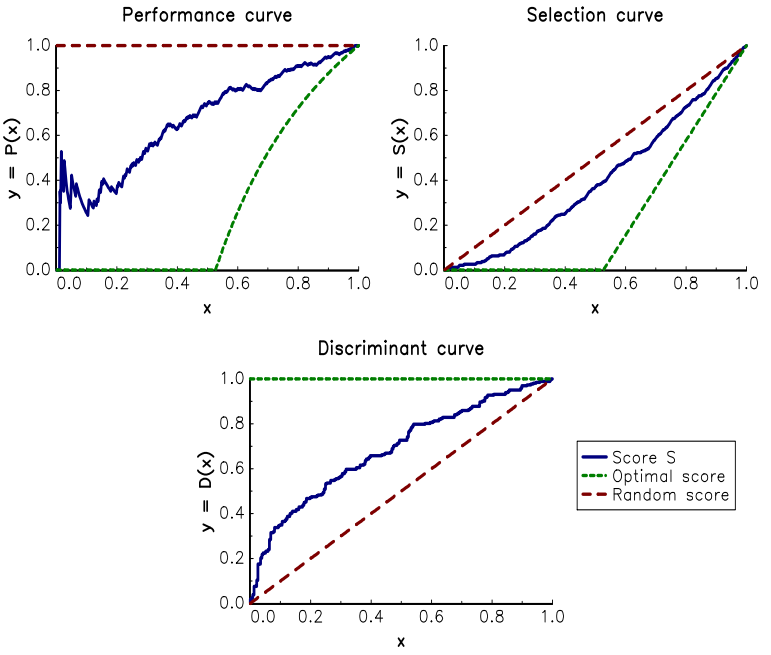


FIGURE 15.39: Performance, selection and discriminant curves

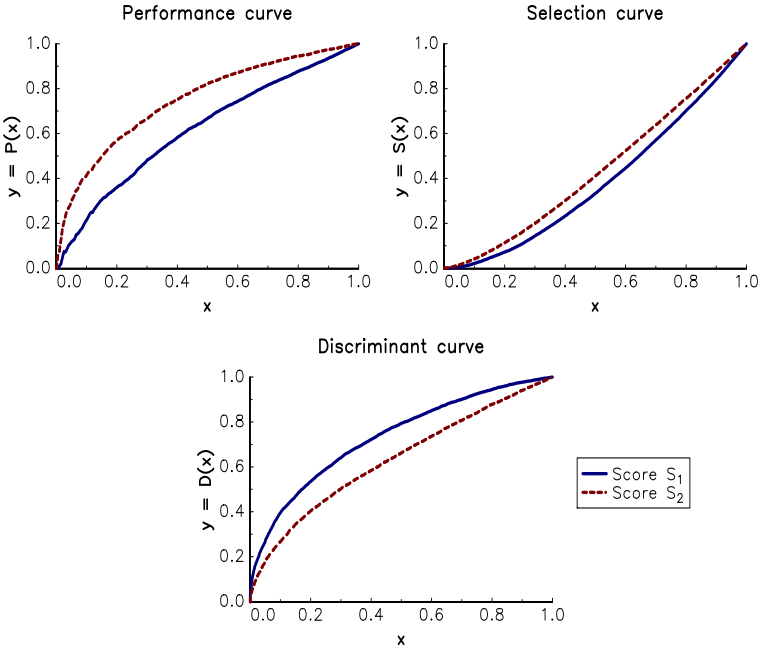


FIGURE 15.40: The score S_1 is better than the score S_2

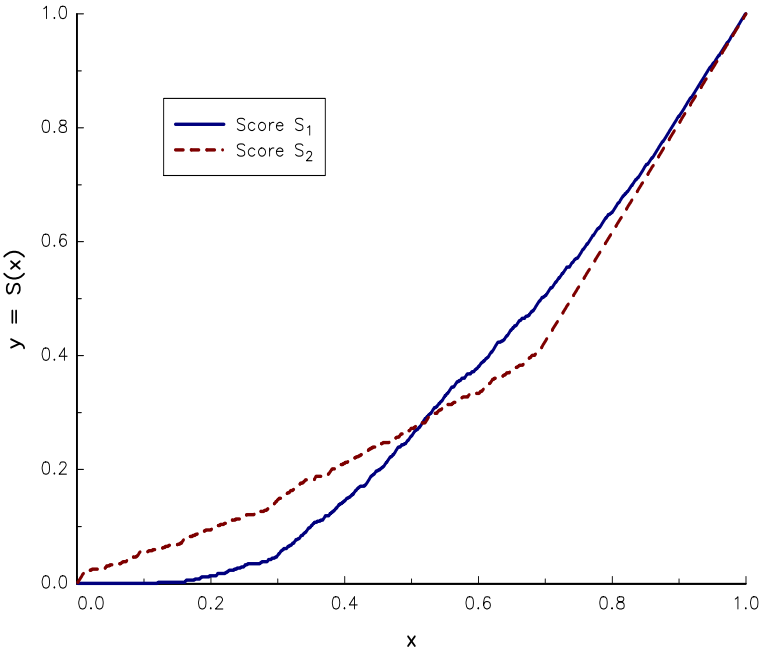


FIGURE 15.41: Illustration of the partial ordering between two scores

15.3.3 Statistical methods

Since the quantitative tools for comparing two scores are numerous, we focus on two non-parametric measures: the Kolmogorov-Smirnov test and the Gini coefficient.

15.3.3.1 Kolmogorov-Smirnov test

We consider the cumulative distribution functions:

$$\mathbf{F}_0(s) = \Pr\{S \leq s \mid Y = 0\}$$

and:

$$\mathbf{F}_1(s) = \Pr\{S \leq s \mid Y = 1\}$$

The score S is relevant if we have the stochastic dominance order $\mathbf{F}_0 \succ \mathbf{F}_1$. In this case, the score quality is measured by the Kolmogorov-Smirnov statistic:

$$\text{KS} = \max_s |\mathbf{F}_0(s) - \mathbf{F}_1(s)|$$

It takes the value 1 if the score is perfect. The KS statistic may be used to verify that the score is not random. We then test the assumption $\mathcal{H}_0 : \text{KS} = 0$ by using the tabulated critical values[74] In Figure 15.42, we give an example with 5 000 observations. The KS statistic is equal to 36%, which implies that $\mathcal{H}_0$ is rejected at the confidence level 1%.

[74]The critical values at the 5% confidence level are equal to:

n	10	50	100	500	5000
CV	40.9%	18.8%	13.4%	6.0%	1.9%

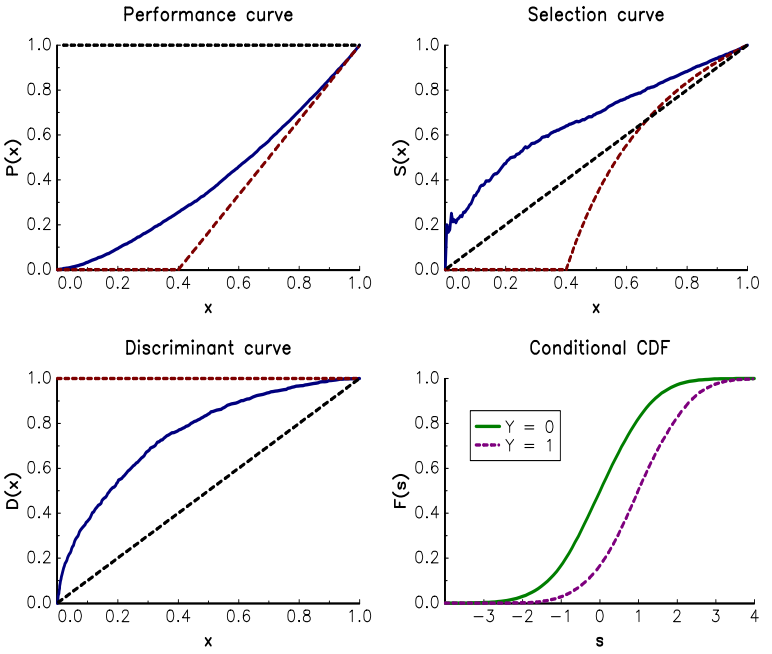


FIGURE 15.42: Comparison of the distributions $\mathbf{F}_0(s)$ and $\mathbf{F}_1(s)$

15.3.3.2 Gini coefficient

The Lorenz curve The Gini coefficient is the statistic, which is the most used for measuring the performance of a score. It is related to the concept of Lorenz curve, which is a graphical representation of the concentration. Let X and Y be two random variables. The Lorenz curve $y = \mathcal{L}(x)$ is the parametric curve defined by:

$$\begin{cases} x = \Pr\{X \leq x\} \\ y = \Pr\{Y \leq y \mid X \leq x\} \end{cases}$$

In economics, x represents the proportion of individuals that are ranked by income while y represents the proportion of income. In this case, the Lorenz curve is a graphical representation of the distribution of income and is used for illustrating inequality of the wealth distribution between individuals. For example, we observe that 70% of individuals have only 34% of total income in Figure 15.43.

Definition of the Gini coefficient The Lorenz curve has two limit cases. If the wealth is perfectly concentrated, one individual holds 100% of the total wealth. If the wealth is perfectly allocated between all the individuals, the corresponding Lorenz curve is the bisecting line. We define the Gini coefficient by:

$$\mathcal{G}ini(\mathcal{L}) = \frac{A}{A+B}$$

where A is the area between the Lorenz curve and the curve of perfect equality, and B is the area between the curve of perfect concentration and the Lorenz curve. By construction, we have $0 \leq \mathcal{G}ini(\mathcal{L}) \leq 1$. The Gini coefficient is equal to zero in the case of perfect equality and one in the case of perfect concentration. We have:

$$\mathcal{G}ini(\mathcal{L}) = 1 - 2\int_0^1 \mathcal{L}(x)\,\mathrm{d}x$$

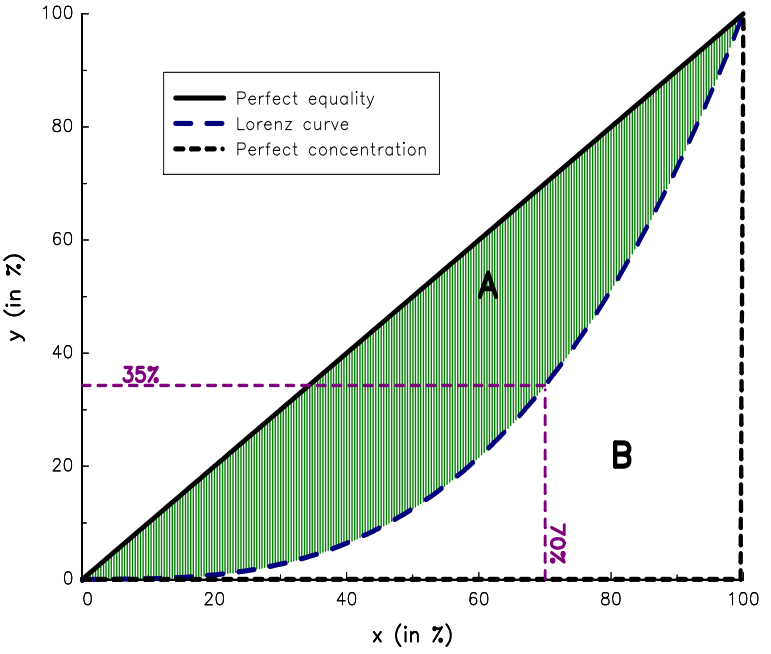


FIGURE 15.43: An example of Lorenz curve

Application to credit scoring We can interpret the selection curve as a Lorenz curve. We recall that $\mathbf{F}(s) = \Pr\{S \leq s\}$, $\mathbf{F}_0(s) = \Pr\{S \leq s \mid Y = 0\}$ and $\mathbf{F}_1(s) = \Pr\{S \leq s \mid Y = 1\}$. The selection curve is defined by the following parametric coordinates:

$$\begin{cases} x(s) = 1 - \mathbf{F}(s) \\ y(s) = 1 - \mathbf{F}_0(s) \end{cases}$$

The selection curve measures the capacity of the score for not selecting bad risks. We could also build the Lorenz curve that measures the capacity of the score for selecting good risks:

$$\begin{cases} x(s) = \Pr\{S \geq s\} = 1 - \mathbf{F}(s) \\ y(s) = \Pr\{S \geq s \mid Y = 1\} = 1 - \mathbf{F}_1(s) \end{cases}$$

It is called the precision curve. Another popular graphical tool is the receiver operating characteristic (or ROC) curve (Powers, 2011), which is defined by:

$$\begin{cases} x(s) = \Pr\{S \geq s \mid Y = 0\} = 1 - \mathbf{F}_0(s) \\ y(s) = \Pr\{S \geq s \mid Y = 1\} = 1 - \mathbf{F}_1(s) \end{cases}$$

An example for a given score S is provided in Figure 15.44. For all the three curves, we can calculate the Gini coefficient. Since the precision and ROC curves are located above the bisecting line, the Gini coefficient associated to the Lorenz curve $\mathcal{L}$ becomes[75]:

$$\mathcal{G}ini(\mathcal{L}) = 2\int_0^1 \mathcal{L}(x)\,\mathrm{d}x - 1$$

[75] An alternative to the Gini coefficient is the AUC measure, which corresponds to the area under the ROC curve. However, they give the same information since they are related by the equation:

$$\mathcal{G}ini(\mathrm{ROC}) = 2 \times \mathrm{AUC}(\mathrm{ROC}) - 1$$

The Gini coefficient of the score S is then computed as follows:

$$\mathcal{G}ini^{\star}(S) = \frac{\mathcal{G}ini(\mathcal{L})}{\mathcal{G}ini(\mathcal{L}^{\star})}$$

where $\mathcal{L}^{\star}$ is the Lorenz curve associated to the perfect score.

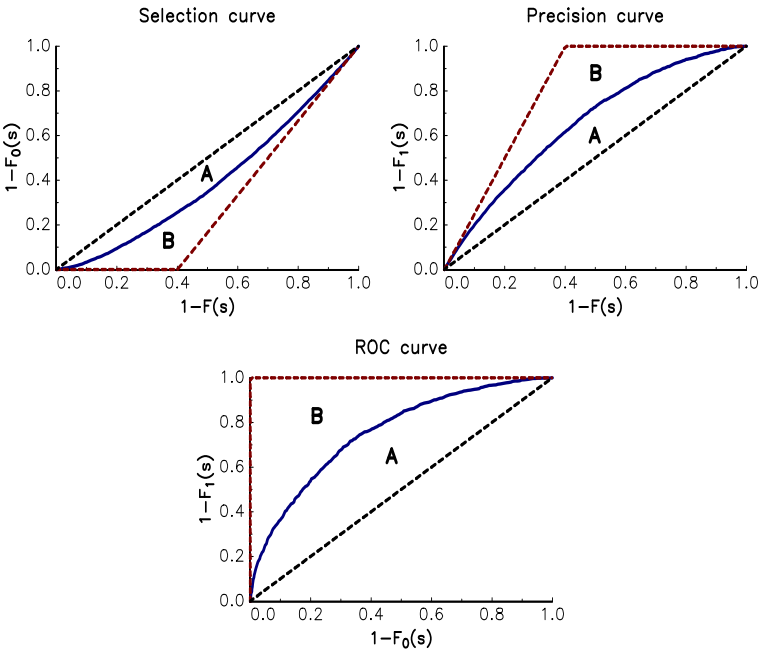


FIGURE 15.44: Selection, precision and ROC curves

Remark 202 *The Gini coefficient is not necessarily the same for the three curves. However, if the population is homogeneous, we generally obtain very similar figures*[76].

15.3.3.3 Choice of the optimal cut-off

The choice of the optimal cut-off $s^{\star}$ depends on the objective function. For instance, we can calibrate $s^{\star}$ in order to achieve a minimum market share. We can also fix a given selection rate. More generally, the objective function can be the profitability of the activity. From a statistical point of view, we must distinguish the construction of the scoring model and the decision rule. In statistical learning, we generally consider three datasets: the training set, the validation set and the test set. The training set is used for calibrating the model and its parameters whereas the validation set helps to avoid overfitting. But the decision rule is based on the test set.

[76] For instance, we obtain the following results with the score S that has been used in Figure 15.44:

Curve	$\mathcal{G}ini(\mathcal{L})$	$\mathcal{G}ini(\mathcal{L}^{\star})$	$\mathcal{G}ini^{\star}(S)$
Selection	20.41%	40.02%	51.01%
Precision	30.62%	59.98%	51.05%
ROC	51.03%	100.00%	51.03%

Confusion matrix A confusion matrix is a special case of contingency matrix. Each row of the matrix represents the frequency in a predicted class while each column represents the frequency in an actual class. Using the test set, it takes the following form:

	$Y=0$	$Y=1$
$S<s$	$n_{0,0}$	$n_{0,1}$
$S\geq s$	$n_{1,0}$	$n_{1,1}$
	$n_0=n_{0,0}+n_{1,0}$	$n_1=n_{0,1}+n_{1,1}$

where $n_{i,j}$ represents the number of observations of the cell (i,j). We notice that each cell of this table can be interpreted as follows:

	$Y=0$	$Y=1$
$S<s$	It is rejected and it is a bad risk (true negative)	It is rejected, but it is a good risk (false negative)
$S\geq s$	It is accepted, but it is a bad risk (false positive)	It is accepted and it is a good risk (true positive)
	(negative)	(positive)

The cells $(S<s,Y=0)$ and $(S\geq s,Y=1)$ correspond to observations that are well-classified: true negative (TN) and true positive (TP). The cells $(S\geq s,Y=0)$ and $(S<s,Y=1)$ correspond to two types of errors:

1. a false positive (FP) can induce a future loss, because it may default: this is a type I error;
2. a false negative (FN) potentially corresponds to a loss of a future P&L[77]: this is a type II error.

Classification ratios Binary classification defines many metrics for measuring the performance of the classifier[78] (Fawcett, 2006):

$$\begin{aligned}
&\text{True Positive Rate} && \text{TPR}=\frac{\text{TP}}{\text{TP}+\text{FN}}\\
&\text{False Negative Rate} && \text{FNR}=\frac{\text{FN}}{\text{FN}+\text{TP}}=1-\text{TPR}\\
&\text{True Negative Rate} && \text{TNR}=\frac{\text{TN}}{\text{TN}+\text{FP}}\\
&\text{False Positive Rate} && \text{FPR}=\frac{\text{FP}}{\text{FP}+\text{TN}}=1-\text{TNR}
\end{aligned}$$

The true positive rate (TPR) is also known as the sensitivity or the recall. It measures the proportion of real good risks that are correctly predicted good risk. Fawcett (2006) also defines the precision or the positive predictive value (PPV):

$$\text{PPV}=\frac{\text{TP}}{\text{TP}+\text{FP}}$$

[77] This is an opportunity cost.

[78] We rewrite the confusion matrix as follows:

	$Y=0$	$Y=1$
$S<s$	TN	FN
$S\geq s$	FP	TP
	$\text{N}=\text{TN}+\text{FP}$	$\text{P}=\text{FN}+\text{TP}$

It measures the proportion of predicted good risks that are correctly real good risk. Besides these metrics, statisticians also use two generic metrics:

1. the accuracy considers the classification of both negatives and positives:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}$$

2. the F_1 score is the harmonic mean of precision and sensitivity:

$$\begin{aligned} F_1 &= \frac{2}{1/\text{precision} + 1/\text{sensitivity}} \\ &= \frac{2 \cdot \text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} \end{aligned}$$

Example 171 *We consider three scoring systems that have been calibrated on a training set. These systems produce a score between* 0 *and* 1 000. *A low value predicts a bad risk while a high value predicts a good risk. In order to calibrate the cut-off, we consider a test set, which is composed of* 10 000 *new observations. In Table 15.23, we report the confusion matrix of each scoring system for different cut-off values (*100, 200 *and* 500*).*

TABLE 15.23: Confusion matrix of three scoring systems and three cut-off values s

Score	$s = 100$		$s = 200$		$s = 500$	
S_1	386	616	698	1 304	1 330	3 672
	1 614	7 384	1 302	6 696	670	4 328
S_2	372	632	700	1 304	1 386	3 616
	1 628	7 368	1 300	6 696	614	4 384
S_3	382	616	656	1 344	1 378	3 624
	1 618	7 384	1 344	6 656	622	4 376
Perfect	1 000	0	2 000	0	2 000	3 000
	1 000	8 000	0	8 000	0	5 000

Using confusion matrices given in Table 15.23, we calculate the different classification ratios and report them in Table 15.24. In addition to the three scoring systems, we have also considered a perfect score in order to show what the best value is for each classification ratio. Finally, we indicate the best scoring system in Table 15.25. We notice that it depends on the ratio and on the value of the cut-off. For instance, if we want to maximize the true positive ratio or minimize the false negative ratio, S_1 is the best scoring system for low value of s while S_2 is better when s is equal to 500. For the other ratios, S_1 seems to be the best system when $s = 100$, otherwise S_2 dominates S_1 and S_3 when $s = 200$ or $s = 500$.

Remark 203 *We recall that* $\mathbf{F}_0(s) = \Pr\{S \leq s \mid Y = 0\}$ *and* $\mathbf{F}_1(s) = \Pr\{S \leq s \mid Y = 1\}$. *We deduce that* $\text{TNR} = \mathbf{F}_0(s)$, $\text{FNR} = \mathbf{F}_1(s)$, $\text{FPR} = 1 - \mathbf{F}_0(s)$ *and* $\text{TPR} = 1 - \mathbf{F}_1(s)$. *Therefore, the ROC curve is the parametric curve, where the* x*-coordinates are the false positive rates and the* y*-coordinates are the true positive rates. Generally, we note* α *and* β *the type I and II errors. We may also interpret the ROC curve as the relationship of* $1 - \beta(s)$ *with respect to* $\alpha(s)$.

TABLE 15.24: Binary classification ratios (in %) of the three scoring systems

Score	s	TPR	FNR	TNR	FPR	PPV	ACC	F_1
	100	92.3	7.7	19.3	80.7	82.1	77.7	86.9
S_1	200	83.7	16.3	34.9	65.1	83.7	73.9	83.7
	500	54.1	45.9	66.5	33.5	86.6	56.6	66.6
	100	92.1	7.9	18.6	81.4	81.9	77.4	86.7
S_2	200	83.7	16.3	35.0	65.0	83.7	74.0	83.7
	500	54.8	45.2	69.3	30.7	87.7	57.7	67.5
	100	92.3	7.7	19.1	80.9	82.0	77.7	86.9
S_3	200	83.2	16.8	32.8	67.2	83.2	73.1	83.2
	500	54.7	45.3	68.9	31.1	87.6	57.5	67.3
	100	100.0	0.0	50.0	50.0	88.9	90.0	94.1
Perfect	200	100.0	0.0	100.0	0.0	100.0	100.0	100.0
	500	62.5	37.5	100.0	0.0	100.0	70.0	76.9

TABLE 15.25: Best scoring system

Cut-off	TPR	FNR	TNR	FPR	PPV	ACC	F_1
100	S_1/S_3	S_1/S_3	S_1	S_1	S_1	S_1	S_1
200	S_1/S_2	S_1/S_2	S_2	S_2	S_2	S_2	S_2
500	S_2	S_2	S_2	S_2	S_2	S_2	S_2

15.4 Exercises

15.4.1 Elastic net regression

We consider the standard linear model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

where $\mathbf{Y}$ is a $n \times 1$ vector, $\mathbf{X}$ is a $n \times K$ matrix and $\mathbf{U} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}_n\right)$. Let $\hat{\beta}$ be the estimator of β, that is the solution of the following least squares problem:

$$\hat{\beta} = \arg\min \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + \frac{\lambda}{2} \left(\alpha \|\beta\|_1 + (1-\alpha) \|\beta\|_2^2 \right)$$

where $\lambda \geq 0$ and $\alpha \in [0, 1]$.

1. We consider the case $\alpha = 0$, which corresponds to the ridge regression.
   (a) Find the optimal estimator $\hat{\beta}^{\text{ridge}}$.
   (b) What is the relationship between the ridge estimator $\hat{\beta}^{\text{ridge}}$ and the ordinary least squares $\hat{\beta}^{\text{ols}}$?
   (c) Deduce the expression of $\mathbb{E}\left[\hat{\beta}^{\text{ridge}}\right]$. Show that $\hat{\beta}^{\text{ridge}}$ is a biased estimator except if $\lambda = 0$.
   (d) Demonstrate that the covariance matrix of $\hat{\beta}^{\text{ridge}}$ is equal to:

$$\text{var}\left(\hat{\beta}^{\text{ridge}}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X} + Q\right)^{-1}$$

where Q is a matrix to determine. Deduce that:

$$\text{var}\left(\hat{\beta}^{\text{ols}}\right) \succeq \text{var}\left(\hat{\beta}^{\text{ridge}}\right)$$

where $\succeq$ is the positive definite ordering.

(e) Let $\hat{\mathbf{Y}}$ be the predicted values of $\mathbf{Y}$. If $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$, the model degree of freedom is equal to the trace of $\mathbf{H}$. Show that the degree of freedom of the ridge model is equal to:

$$\text{df}^{\text{model}} = \sum_{k=1}^{K} \frac{s_k^2}{s_k^2 + \lambda}$$

where $(s_1, \ldots, s_K)$ are the singular values of $\mathbf{X}$ (Hastie *et al.*, 2009).

(f) What does the previous results become when $\mathbf{X}$ is an orthonormal matrix?

2. We consider the case $\alpha > 0$, which corresponds to the elastic net regression (Zou and Hastie, 2005).

(a) Write the corresponding QP program.

(b) Consider the data of Example 164 on page 936. Compare the estimates $\hat{\beta}$ when α is respectively equal to 0, 0.25, 0.5 and 1.0.

15.4.2 Cross-validation of the ridge linear regression

We consider the ridge estimator:

$$\hat{\beta} = \arg\min \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + \frac{\lambda}{2}\beta^\top \beta$$

where $\mathbf{Y}$ is a $n \times 1$ vector, $\mathbf{X}$ is a $n \times K$ matrix and β is a $K \times 1$ vector.

1. Compute the ridge estimator $\hat{\beta}$.

2. We note $\hat{\beta}_{-i}$ the ridge estimator when leaving out the i^{th} observation:

$$\hat{\beta}_{-i} = \arg\min \frac{1}{2} (\mathbf{Y}_{-i} - \mathbf{X}_{-i}\beta)^\top (\mathbf{Y}_{-i} - \mathbf{X}_{-i}\beta) + \frac{\lambda}{2}\beta^\top \beta$$

where $\mathbf{Y}_{-i}$ and $\mathbf{X}_{-i}$ correspond to $\mathbf{Y}$ and $\mathbf{X}$ with the i^{th} row removed. By using the relationships $\mathbf{X}^\top\mathbf{X} = \mathbf{X}_{-i}^\top\mathbf{X}_{-i} + x_i x_i^\top$ and $\mathbf{X}^\top\mathbf{Y} = \mathbf{X}_{-i}^\top\mathbf{Y}_{-i} + x_i y_i$, show that:

$$\hat{\beta}_{-i} = \hat{\beta} - \frac{\left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1} x_i \hat{u}_i}{1 - h_i}$$

where $\hat{u}_i = y_i - x_i^\top \hat{\beta}$ and $h_i = x_i^\top \left(\mathbf{X}^\top\mathbf{X} + \lambda I_K\right)^{-1} x_i$.

3. We note $\hat{y}_{i,-i} = x_i^\top \hat{\beta}_{-i}$ and $\hat{u}_{i,-i} = y_i - \hat{y}_{i,-i}$. Demonstrate that:

$$\hat{u}_{i,-i} = \frac{\hat{u}_i}{1 - h_i}$$

4. Calculate the predicted residual error sum of squares (PRESS) statistic:

$$\mathcal{P}\text{ress} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2$$

where $\hat{y}_{i,-i}$ is the estimate of y_i based on the ridge model when leaving out the i^{th} observation.

5. In the OLS regression, we reiterate that $\mathrm{df}^{(\mathrm{model})} = \mathrm{trace}\,\mathbf{H} = K$ where $\mathbf{H}$ is the hat matrix for the OLS regression. Define the corresponding hat matrix $\mathbf{H}(\lambda)$ for the ridge regression. Show that:

$$\mathrm{df}^{(\mathrm{model})}(\lambda) = \sum_{k=1}^{K} \frac{s_k^2}{s_k^2 + \lambda}$$

where $(s_1, \ldots, s_K)$ are the singular values of $\mathbf{X}$.

6. The generalized cross-validation (GCV) statistic is defined by:

$$GCV = nK^2 \left(\sum_{k=1}^{K} \frac{(n-K)\, s_k^2 + n\lambda}{s_k^2 + \lambda} \right)^{-2} \mathrm{RSS}\left(\hat{\beta}(\lambda)\right)$$

where $\bar{h} = n^{-1} \sum_{i=1}^{n} \mathbf{H}(\lambda)_{i,i}$ and $\mathrm{RSS}\left(\hat{\beta}(\lambda)\right)$ is the residual sum of squares calculated with the ridge estimator $\hat{\beta}(\lambda)$. What is the relationship between GCV and PRESS statistics? What is the impact of λ?

7. Show that another expression of the GCV statistic is:

$$GCV = n \left(n - K + \sum_{k=1}^{K} \frac{\lambda}{s_k^2 + \lambda} \right)^{-2} \mathrm{RSS}\left(\hat{\beta}(\lambda)\right)$$

8. Using the data of Example 165 on page 940, calculate the estimates $\hat{\beta}_{-i}$ when λ is equal to 3.0. Compute also $\hat{y}_{i,-i}$, $\hat{u}_{i,-i}$, $\hat{u}_i$ and h_i. Deduce then the value of PRESS and GCV statistics.

15.4.3 K-means and the Lloyd's algorithm

1. We consider n observations with K attributes $x_{i,k}$ ($i = 1, \ldots, n$ and $k = 1, \ldots, K$). We note x_i the $K \times 1$ vector $(x_{i,1}, \ldots, x_{i,K})$. Show that:

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i - x_j\|^2 = n \sum_{i=1}^{n} \|x_i - \bar{x}\|^2$$

where:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

2. We recall that the loss function of the K-means clustering method is:

$$\mathcal{L}(\mathcal{C}) = \frac{1}{2} \sum_{j=1}^{n_C} \sum_{\mathcal{C}(i)=j} \sum_{\mathcal{C}(i')=j} \|x_i - x_{i'}\|^2$$

Deduce that:

$$\mathcal{L}(\mathcal{C}) = \sum_{j=1}^{n_C} n_j \sum_{\mathcal{C}(i)=j} \|x_i - \bar{x}_j\|^2$$

where $\bar{x}_j$ and n_j are two quantities to define.

3. We consider the following optimization function:

$$\left\{\mu_1^\star, \ldots, \mu_{n_C}^\star\right\} = \arg\min \sum_{j=1}^{n_C} n_j \sum_{\mathcal{C}(i)=j} \left\|x_i - \mu_j\right\|^2$$

Show that $\mu_j^\star = \bar{x}_j$. Comment on this result.

4. Apply the K-means analysis to Example 169 and compare the results with those obtained with the discriminant analysis.

15.4.4 Derivation of the principal component analysis

The following exercise is taken from Chapters 1 and 2 of Jolliffe (2002). Let X be a $K \times 1$ random vector, whose covariance matrix is equal to Σ. We consider the linear transform $Z_j = \beta_j^\top X$ where β_j is a $K \times 1$ vector.

1. Calculate $\text{var}(Z_1)$ and define the PCA objective function to estimate β_1. Show that β_1 is the eigenvector associated to the largest eigenvalue of Σ.

2. Calculate $\text{var}(Z_2)$ and $\text{cov}(Z_1, Z_2)$. Define then the PCA objective function to estimate β_2. Show that β_2 is the eigenvector associated to the second eigenvalue of Σ.

15.4.5 Two-class separation maximization

We note x_i the $K \times 1$ vector of exogenous variables X for the i^{th} observation.

1. We consider the case of J classes. We note $\hat{\mu}_j$ the mean vector for class $\mathcal{C}_j$:

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} x_i$$

and $\hat{\mu}$ the mean vector for the entire sample:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{j=1}^{J} n_j \hat{\mu}_j$$

Calculate the scatter matrices $\mathbf{S}$, $\mathbf{S}_W$ and $\mathbf{S}_B$. Show that:

$$\mathbf{S} = \mathbf{S}_W + \mathbf{S}_B$$

2. We now consider the two-class problem, and we note $y_i = \beta^\top x_i$. Show that:

$$\beta^\top \mathbf{S}_B \beta = \frac{n_1 n_2}{n_1 + n_2} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2$$

where:

$$\tilde{\mu}_j = \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} y_i$$

3. Show that:

$$\beta^\top \mathbf{S}_W \beta = \tilde{s}_1^2 + \tilde{s}_2^2$$

where:

$$\tilde{s}_j^2 = \sum_{i \in \mathcal{C}_j} \left(y_i - \tilde{\mu}_j\right)^2$$

4. Deduce that the Fisher optimization program is:

$$\beta^\star = \arg\max \frac{\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

   What is the interpretation of this statistical problem?

5. Find the optimal value $\beta^\star$ and verify that the decision boundary is linear.

6. Using Example 170 on page 967, calculate $\mathbf{S}_W$ and $\mathbf{S}_B$. Find the optimal value $\beta^\star$ and compute the score for each observation. Propose an assignment decision based on the mid-point rule. Comment on these results.

15.4.6 Maximum likelihood estimation of the probit model

1. Given a sample $\{(x_i, y_i), i = 1, \ldots, n\}$, find the log-likelihood function of the probit model.

2. Let $J(\beta)$ be the Jacobian matrix of the log-likelihood vector. Show that:

$$J_{i,k}(\beta) = \frac{\left(y_i - \Phi\left(x_i^\top \beta\right)\right) \phi\left(x_i^\top \beta\right)}{\Phi\left(x_i^\top \beta\right)\left(1 - \Phi\left(x_i^\top \beta\right)\right)} \cdot x_{i,k}$$

   for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Define the score vector $\mathcal{S}(\beta)$.

3. Let $H(\beta)$ be the Hessian matrix of the log-likelihood function. Show that:

$$H(\beta) = -\sum_{i=1}^{n} H_i \cdot \left(x_i x_i^\top\right)$$

   where:

$$\begin{aligned} H_i &= y_i \frac{\left(\phi\left(x_i^\top \beta\right) + x_i^\top \beta \Phi\left(x_i^\top \beta\right)\right)}{\Phi\left(x_i^\top \beta\right)^2} \phi\left(x_i^\top \beta\right) + \\ &\quad (1 - y_i) \frac{\left(\phi\left(x_i^\top \beta\right) - x_i^\top \beta\left(1 - \Phi\left(x_i^\top \beta\right)\right)\right)}{\left(1 - \Phi\left(x_i^\top \beta\right)\right)^2} \phi\left(x_i^\top \beta\right) \end{aligned}$$

4. Propose a Newton-Raphson algorithm to find the ML estimate.

15.4.7 Computation of feed-forward neural networks

We consider the canonical neural network without constant and direct link.

1. We note X the input matrix of dimension $n \times n_x$ and Y the output matrix of dimension $n \times n_y$. Let $\hat{Y}$ be the prediction of Y. Find the matrix relationship between X and $\hat{Y}$ with respect to the parameter matrices β and γ of dimension $n_z \times n_x$ and $n_y \times n_z$.

2. We assume that the activation functions $f_{x,z}$ and $f_{z,y}$ are the identity function. Demonstrate that the neural network is equivalent to an overidentified linear model or a constrained linear regression.

3. We consider the additive loss function:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n_y} \mathcal{L}_{i,j}(\theta)$$

where:

$$\mathcal{L}_{i,j}(\theta) = \xi(y_j(x_i), y_{i,j})$$

Calculate the matrices $\partial_\gamma \mathcal{L}(\theta)$ and $\partial_\beta \mathcal{L}(\theta)$ of dimension $n_y \times n_z$ and $n_z \times n_x$.

4. We assume that the activation functions $f_{x,z}$ and $f_{z,y}$ correspond to the logistic function and the loss is the least squares error function. Find the matrices $\partial_\gamma \mathcal{L}(\theta)$ and $\partial_\beta \mathcal{L}(\theta)$.

5. Same question if we consider the cross-entropy error loss:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{n} (y_i \ln y(x_i) + (1 - y_i) \ln (1 - y(x_i)))$$

6. Explain why we cannot use the property of additivity in the case of the softmax function.

7. Calculate the matrices $\partial_\gamma \mathcal{L}(\theta)$ and $\partial_\beta \mathcal{L}(\theta)$ when $f_{z,y}$ is the softmax function, $f_{x,z}$ is the identity function, and the loss function is the multi-class error function:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{n} \sum_{j=1}^{n_C} y_{i,j} \ln y_j(x_i)$$

where n_C is the number of classes[79].

8. Extend the previous results when we consider a constant between the x's and the z's, a constant between the z's and the y's and a direct link between the x's and the y's.

15.4.8 Primal and dual problems of support vector machines

The goal of this exercise is to determine the primal and dual problems of the different SVM models. For each problem, we ask to write the primal problem into a quadratic programming (QP) format:

$$\begin{aligned} \hat{\theta} &= \arg\min \frac{1}{2}\theta^\top Q\theta - \theta^\top R \\ \text{s.t.} &\begin{cases} A\theta = B \\ C\theta \geq D \\ \theta^- \leq \theta \leq \theta^+ \end{cases} \end{aligned}$$

where θ is the vector of parameters. Then, we ask to find the corresponding dual problem and also the associated QP matrix form.

[79] Hint: Use the following decomposition $\mathcal{L}(\theta) = \sum_{i=1}^{n} \mathcal{L}_i(\theta)$.

Hard margin classification

We first begin with the hard margin classifier. We recall that the primal optimization problem is:

$$\begin{aligned}\left\{\hat{\beta}_0, \hat{\beta}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 \\ \text{s.t.} \quad & y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 \qquad \text{for } i = 1, \ldots, n\end{aligned}$$

1. By noting θ the vector of parameters, write the primal problem in the QP form.
2. We note $\alpha = (\alpha_1, \ldots, \alpha_n)$ the vector of Lagrange coefficients associated to the constraints $y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1$. Write the Lagrange function and find the first-order conditions.
3. Deduce that the dual problem is:

$$\begin{aligned}\hat{\alpha} &= \arg\max \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j x_i^\top x_j^\top \\ \text{s.t.} \quad & \alpha \geq \mathbf{0}\end{aligned}$$

4. Write this dual problem as a QP problem.
5. Determine the dual QP problem directly by applying Equation (A.12) on page 1047. What do you observe? How to fix this issue?

Soft margin classification with binary hinge loss

We now consider the soft margin classification problem:

$$\begin{aligned}\left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \qquad \text{for } i = 1, \ldots, n\end{aligned}$$

1. Write the primal problem as a QP problem.
2. Show that the objective function of the dual problem does not change compared to the hard margin classifier. What does the dual QP problem become?
3. How can we characterize the support vectors?
4. Find the optimal values of ξ_i.
5. We consider the training data set given in Table 15.18 on page 989. Represent the optimal values of β_0, β_1, β_2, $\sum_{i=1}^n \xi_i$ and the margin M with respect to C. Compare the optimal hyperplane when $C = 0.07$ with the optimal hyperplane obtained with the hard margin classifier.

Soft margin classification with squared hinge loss

We replace the binary hinge loss by the squared hinge loss:

$$\begin{aligned}\left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 \\ \text{s.t.} \quad & \begin{cases} y_i\left(\beta_0 + x_i^\top \beta\right) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \qquad \text{for } i = 1, \ldots, n\end{aligned}$$

1. Write the primal problem as a QP problem.
2. Find the dual problem. What do you observe?
3. We consider the training data set given in Table 15.18 on page 989. Study the convergence of the optimal values of β_0, β_1, β_2, $\sum_{i=1}^n \xi_i$ and the margin M with respect to C. What is the main difference between binary and squared hinge loss functions?
4. We introduce in the training set two new points $(6.0, 5.0, +1)$ $(i = 16)$ and $(2.0, 2.0, -1)$ $(i = 17)$. Calculate $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\alpha}_i$ and $\hat{\xi}_i$ when the constant C is equal to 1.

Soft margin classification with ramp loss

1. Compare $0-1$, binary hinge, squared hinge and ramp loss functions.
2. Using the property $\min(1, \max(0, a)) = \max(0, a) - \max(0, a-1)$, show that $\mathcal{L}^{\text{ramp}}(x_i, y_i)$ is the difference of two convex functions. Comment on this result.

LS-SVM regression

We consider the following optimization problem:

$$\begin{aligned}\left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i^2 \\ &\text{s.t.} \quad y_i = \beta_0 + x_i^\top \beta + \xi_i \qquad \text{for } i = 1, \ldots, n\end{aligned}$$

1. Write the primal problem as a QP problem.
2. Find the dual QP problem.
3. Deduce the expression of $\hat{\beta}_0$ and $\hat{\beta}$. Show that the residuals are centered.

ε-SVM regression

We consider the following optimization problem:

$$\begin{aligned}\left\{\hat{\beta}_0, \hat{\beta}, \hat{\xi}^-, \hat{\xi}^+\right\} &= \arg\min \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \left(\xi_i^- + \xi_i^+\right) \\ &\text{s.t.} \quad \begin{cases} \beta_0 + x_i^\top \beta - y_i \leq \varepsilon + \xi_i^- \\ y_i - \beta_0 - x_i^\top \beta \leq \varepsilon + \xi_i^+ \\ \xi_i^- \geq 0 \\ \xi_i^+ \geq 0 \end{cases} \qquad \text{for } i = 1, \ldots, n\end{aligned}$$

where $\varepsilon \geq 0$.

1. Write the primal problem as a QP problem.
2. Find the dual problem.
3. Write the dual problem as a QP problem.
4. Deduce the expression of $\hat{\beta}_0$ and $\hat{\beta}$.
5. Calculate the optimal values $\hat{\xi}^-$ and $\hat{\xi}^+$.
6. What does the optimization problem becomes when $\varepsilon = 0$?

15.4.9 Derivation of the AdaBoost algorithm as the solution of the additive logit model

We consider a special case of additive models, where the loss function is specified as follows:

$$\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right) = \sum_{i=1}^{n} \mathcal{L}\left(y_i, \hat{g}_{(s-1)}\left(x_i\right) + \beta_{(s)} f_{(s)}\left(x_i\right)\right)$$

$\hat{g}_{(s)}(x) = \sum_{s'=1}^{s} \hat{\beta}_{(s')} \hat{f}_{(s')}(x)$, $\hat{f}_{(s)}$ is the s^{th} optimal classification model and $\mathcal{L}(y, f(x)) = e^{-yf(x)}$.

1. Show that:
$$\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right) = \sum_{i=1}^{n} w_{i,s} e^{-y_i \beta_{(s)} f_{(s)}(x_i)}$$
where $w_{i,s}$ is a quantity to determine.

2. Find an expression of $\mathcal{L}\left(\beta_{(s)}, f_{(s)}\right)$ that depends on the error rate:
$$\mathcal{L}_{(s)} = \frac{\sum_{i=1}^{n} w_{i,s} \cdot \mathbb{1}\left\{y_i \neq y_{i,s}\right\}}{\sum_{i=1}^{n} w_{i,s}}$$
where $y_{i,s} = f_{(s)}(x_i)$.

3. We assume that $f_{(s)}$ is known. Verify that the optimal value of $\hat{\beta}_{(s)}$:
$$\hat{\beta}_{(s)} = \arg\min \mathcal{L}\left(\beta_{(s)}, f_{(s)}\right)$$
is equal to:
$$\hat{\beta}_{(s)} = \frac{1}{2} \ln\left(\frac{1 - \mathcal{L}_{(s)}}{\mathcal{L}_{(s)}}\right)$$

4. Suppose that $\hat{f}_{(s)}$ has been already estimated. Show that the normalized observation weights are:
$$w_{i,s+1} = \frac{w_{i,s} e^{w_s \cdot \mathbb{1}\left\{y_i \neq \hat{y}_{i,s}\right\}}}{\sum_{i=1}^{n} w_{i',s} e^{w_s \cdot \mathbb{1}\left\{y_{i'} \neq \hat{y}_{i',s}\right\}}}$$
where w_s is a parameter to determine.

5. Conclude on these results.

15.4.10 Weighted estimation

We note $w = (w_1, \ldots, w_n)$ the vector of observation weights.

1. We consider the weighted log-likelihood function:
$$\boldsymbol{\ell}_w(\theta) = \sum_{i=1}^{n} w_i \cdot \boldsymbol{\ell}_i(\theta)$$
   (a) Define the weighted maximum likelihood estimator.
   (b) Find the expression of the Jacobian and Hessian matrices.

2. We consider neural networks (Exercise 15.4.7 on page 1025).

(a) Define the least squares loss function $\mathcal{L}_w(\theta)$. Give the matrix form of the derivatives $\partial_\gamma \mathcal{L}_w(\theta)$ and $\partial_\beta \mathcal{L}_w(\theta)$.

(b) Same question if we consider the cross-entropy loss function.

3. We consider the soft margin SVM classification (Exercise 15.4.8 on page 1027).

(a) Define the optimization problem.

(b) What is the impact of introducing weights on the primal and dual problems.

(c) Why weighted hard margin classification does not make sense?

Conclusion

In the past forty years, risk management has considerably changed in the banking industry and more broadly in the financial sector. There are certainly two main reasons. The first one is due to the development of financial markets, the innovation of financial solutions and the competitiveness between financial agents. Since products and operations are more complex than in the past, it is quite normal that risk management has followed the trend. However, this first reason explains partially the development of risk management. The second reason is that carrying on banking and financial business requires a strong risk management, because risk pricing has become the essential element to ensure the sustainability of the financial institution. On top of that, regulation has perfectly understood the role of the banking sector on the economy, which is positive for boosting the economic growth but may also be a problem of systemic risk. In particular, the 2008 Global Financial Crisis has completely changed the approach of regulators, and the place of risk management in the financial sector. Before 2008, risk management was a tool for banks for managing their own business risk. Since 2008, risk management has become a tool for regulators and supervisors for managing the systemic risk of the whole financial sector. This is particularly true for banks, but this phenomenon is now expanding to other financial institutions such as insurance companies and asset managers.

This handbook reflects the evolution of risk management of these last forty years. Besides the presentation of statistical and mathematical tools that are necessary for measuring and managing financial risks, it gives the guidelines of the banking and financial regulation and introduces the different methods for computing capital requirements. This handbook illustrates that there is no other industry in the world, where the regulation is so complex and strong. It also illustrates that financial risk management is highly mathematical and technical. The combination of these two dimensions makes the practice of risk management a difficult exercise. In this handbook, we use many examples, provide many illustrations and propose many exercises in order for the student to gain a strong knowledge in the practice of risk measurement. Measuring the risk is the first step before managing it. Therefore, this handbook does not claim to give recipes in order to take the right risk management decisions. This skill will develop with work experience and real life situations. But this handbook claims that the student has the essential background of risk measurement and regulatory rules to become a risk manager.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Appendix A

Technical Appendix

A.1 Numerical analysis

A.1.1 Linear algebra

Following Horn and Johnson (2012), we recall some definitions about matrices[1]:

- the square matrix A is symmetric if it is equal to its transpose $A^\top$;
- the square matrix A is hermitian if it is equal to its own conjugate transpose A^*, implying that we have $A_{i,j} = \text{conj}\, A_{j,i}$;
- we say that A is an orthogonal matrix if we have $AA^\top = A^\top A = I$ and an unitary matrix if we have $A^* = A^{-1}$;
- A^+ is the Moore-Penrose inverse or pseudo-inverse of A if $AA^+A = A$, $A^+AA^+ = A^+$ and, AA^+ and A^+A are hermitian matrices; in the case where A is invertible, we have $A^+ = A^{-1}$; when A has linearly independent columns, we have $A^+ = \left(A^\top A\right)^{-1} A^\top$.

A.1.1.1 Eigendecomposition

The value λ is an eigenvalue of the $n \times n$ matrix A if there exists a non-zero eigenvector v such that we have $Av = \lambda v$. Let V denote the matrix composed of the n eigenvectors. We have:

$$AV = V\Lambda$$

where $\Lambda = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues. We finally obtain the eigendecomposition of the matrix A:

$$A = V\Lambda V^{-1} \tag{A.1}$$

If A is an hermitian matrix, then the matrix V of eigenvectors is unitary. It follows that:

$$A = V\Lambda V^*$$

In particular, if A is a symmetric real matrix, we obtain[2]:

$$A = V\Lambda V^\top \tag{A.2}$$

[1] To go further, the reader may consult the book of Meucci (2005), which contains an extensive presentation of linear algebra tools used in risk management.

[2] We have:

$$\begin{aligned} A^\top &= \left(V\Lambda V^{-1}\right)^\top \\ &= \left(V^{-1}\right)^\top \Lambda V^\top \end{aligned}$$

We deduce that $V^{-1} = V^\top$.

Remark 204 *A related decomposition is the singular value decomposition. Let A be a rectangular matrix with dimension $m \times n$. We have:*

$$A = U\Sigma V^* \tag{A.3}$$

where U is a $m \times m$ unitary matrix, Σ is a $m \times n$ diagonal matrix with elements $\sigma_i \geq 0$ and V is a $n \times n$ unitary matrix. σ_i are the singular values of A, u_i are the left singular vectors of A, and v_i are the right singular vectors of A.

A.1.1.2 Generalized eigendecomposition

The generalized eigenvalue problem is $Av = \lambda Bv$ where A and B are two $n \times n$ matrices. In a matrix form, we have:

$$AV = BV\Lambda$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of ordered generalized eigenvalues. The generalized eigenvalue problem is related to the maximum/minimum of the Rayleigh quotient:

$$R(x) = \frac{x^\top Ax}{x^\top Bx}$$

Indeed, we get $x^\star = v_1$ and $R(x^\star) = \lambda_1$ for the maximization problem $x = \arg\max R(x)$ and $x^\star = v_n$ and $R(x^\star) = \lambda_n$ for the minimization problem $x = \arg\min R(x)$.

A.1.1.3 Schur decomposition

The Schur decomposition of the $n \times n$ matrix A is equal to:

$$A = QTQ^* \tag{A.4}$$

where Q is a unitary matrix and T is an upper triangular matrix[3]. This decomposition is useful to calculate matrix functions.

Let us consider the matrix function in the space $\mathbb{M}$ of square matrices:

$$\begin{aligned} f: \quad & \mathbb{M} \longrightarrow \mathbb{M} \\ & A \longmapsto B = f(A) \end{aligned}$$

For instance, if $f(x) = \sqrt{x}$ and A is positive, we can define the matrix B such that:

$$BB^* = B^*B = A$$

B is called the square root of A and we note $B = A^{1/2}$. This matrix function generalizes the scalar-valued function to the set of matrices. Let us consider the following Taylor expansion:

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \ldots$$

We can show that if the series converge for $|x - x_0| < \alpha$, then the matrix $f(A)$ defined by the following expression:

$$f(A) = f(x_0) + (A - x_0 I) f'(x_0) + \frac{(A - x_0 I)^2}{2!} f''(x_0) + \ldots$$

[3] Q and T are also called the transformation matrix and the Schur form of A.

converges to the matrix B if $|A - x_0 I| < \alpha$ and we note $B = f(A)$. In the case of the exponential function, we have:

$$f(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

We deduce that the exponential of the matrix A is equal to:

$$B = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

In a similar way, the logarithm of A is the matrix B such that $e^B = A$ and we note $B = \ln A$.

Let A and B be two $n \times n$ square matrices. Using the Taylor expansion, Golub and Van Loan (2013) showed that $f\left(A^\top\right) = f(A)^\top$, $Af(A) = f(A)A$ and $f\left(B^{-1}AB\right) = B^{-1}f(A)B$. It follows that:

$$e^{A^\top} = \left(e^A\right)^\top$$

and:

$$e^{B^{-1}AB} = B^{-1}e^A B$$

If $AB = BA$, we can also prove that $Ae^B = e^B A$ and $e^{A+B} = e^A e^B = e^B e^A$.

Remark 205 *There are different ways to compute numerically $f(A)$. For transcendental functions, we have:*

$$f(A) = Qf(T)Q^*$$

where $A = QTQ^*$ is the Schur decomposition of A. Because T is an upper diagonal matrix, $f(T)$ is also a diagonal matrix whose elements can be calculated with Algorithm 9.1.1 of Golub and Van Loan (2013). This algorithm is reproduced below[4]*.*

A.1.2 Approximation methods

A.1.2.1 Spline functions

We consider a set of data points (x_i, y_i) where $x_1 < x_2 < \cdots < x_n$. $S(x)$ is the associated cubic spline if $S(x)$ is a C^2 function, $S(x_i) = y_i$ and $S(x)$ is a polynomial of degree 3 on each interval:

$$S(x) = a_i + b_i x + c_i x^2 + d_i x^3 \qquad \text{if } x \in [x_i, x_{i+1}]$$

The C^2 property implies that:

$$\begin{aligned} a_{i-1} + b_{i-1}x_i + c_{i-1}x_i^2 + d_{i-1}x_i^3 &= a_i + b_i x_i + c_i x_i^2 + d_i x_i^3 \\ b_{i-1} + 2c_{i-1}x_i + 3d_{i-1}x_i^2 &= b_i + 2c_i x_i + 3d_i x_i^2 \\ 2c_{i-1} + 6d_{i-1}x_i &= 2c_i + 3d_i x_i \end{aligned}$$

Therefore, we obtain a linear system of $4n$ equations with $4n$ unknowns. By assuming that[5] $c_0 = d_0 = c_n = d_n = 0$, the linear system is tridiagonal and easy to solve. The main interest of cubic splines is its tractability, because it is straightforward to calculate the quantities $S(x)$, $S'(x)$, $S''(x)$, $S^{-1}(x)$ and $\int_{x_0}^{x} S(u)\,\mathrm{d}u$ for any value x. This explains that it is extensively used in finance.

[4] For the exponential matrix, we may prefer to use the Pade approximation method, which is described in Algorithm 9.3.1 (scaling and squaring) of Golub and Van Loan (2013). See also the survey of Moler and Van Loan (2003).

[5] This is equivalent to impose that the cubic spline is linear if $x < x_1$ and $x > x_n$.

Algorithm 3 Schur-Parlett matrix function $f(A)$

Compute the Schur decomposition $A = QTQ^*$
Initialize F to the matrix $\mathbf{0}_{n\times n}$
for $i = 1 : n$ **do**
  $f_{i,i} \leftarrow f(t_{i,i})$
end for
for $p = 1 : n-1$ **do**
  for $i = 1 : n-p$ **do**
    $j \leftarrow i+p$
    $s \leftarrow t_{i,j}(f_{j,j} - f_{i,i})$
    for $k = i+1 : j-1$ **do**
      $s \leftarrow s + t_{i,k}f_{k,j} - f_{i,k}t_{k,j}$
    end for
    $f_{i,j} \leftarrow s/(t_{j,j} - t_{i,i})$
  end for
end for
$B \leftarrow QFQ^*$
return B

Source: Golub and Van Loan (2013), page 519.

Remark 206 *The interpolation method can be extended to the smoothing problem:*

$$\min p \cdot \sum_{i=1}^{n} (y_i - S(x_i))^2 + (1-p) \cdot \int_{x_1}^{x_n} S''(u)^2 \, \mathrm{d}u$$

where p is the smoothing parameter. We obtain the cubic spline solution when p is equal to 1, whereas we obtain the least squares solution when p is equal to 0. In the general case, the first-order condition consists in solving a band linear system.

A.1.2.2 Positive definite matrix approximation

The computation of Gaussian risk measures involves the use of covariance or correlation matrices. Since we can manipulate many instruments and securities, we generally observe missing values in the dataset. Therefore, several approaches can be used to estimate the covariance matrix Σ. The two most popular approaches are listwise and pairwise methods. Listwise deletion removes all the observations that have one or more missing values. Since this approach is popular, it cannot be implemented from a practical point of view. For instance, deleting all the public holidays dramatically reduces the number of valid dates in a global universe of stocks. This is why pairwise deletion is used in practice. It consists of deleting the observations by considering each pair of observations. However, the estimated covariance matrix $\hat{\Sigma}$ is generally not positive definite. Another issue occurs when the number of observations is lower than the number of variables. In this case, $\hat{\Sigma}$ is only positive semi-definite.

Computing the nearest covariance matrix We assume that Σ is not a positive semi-definite matrix. We consider the square root decomposition $\Sigma = A^2$ where $A = A_1 + iA_2$. We have $A^2 = A_1^2 - A_2^2$ because $A_1A_2 = \mathbf{0}$ (Horn and Johnson, 2012). We deduce that the

eigenvalues $\lambda_i(\Sigma)$ of Σ are related to the eigenvalues of A_1 and A_2:

$$\lambda_i(\Sigma) = \begin{cases} \lambda_i^2(A_1) & \text{if } \lambda_i\left(A_1^2\right) > 0 \\ -\lambda_i^2(A_2) & \text{otherwise} \end{cases}$$

Therefore, Σ can be approximated by $\tilde{\Sigma} = A_1^2$, which is a positive semi-definite matrix. Moreover, we have:

$$\begin{aligned} x^\top \Sigma x &= x^\top (A_1 + iA_2)^2 x \\ &= x^\top A_1^2 x - x^\top A_2^2 x \\ &\leq x^\top A_1^2 x \end{aligned}$$

This means that any quadratic form $x^\top \Sigma x$ is bounded by $x^\top \tilde{\Sigma} x$. This means that $\tilde{\Sigma}$ is a conservative estimator when computing the Gaussian value-at-risk.

Remark 207 *We can transform any positive semi-definite matrix $\tilde{\Sigma}$ into a (strict) positive definite matrix $\breve{\Sigma}$ by considering the eigenvalue thresholding method $\breve{\Sigma} = V\breve{\Lambda}V^\top$ where $\tilde{\Sigma} = V\Lambda V^\top$, $\breve{\Lambda}_{i,i} = \max(\Lambda_{i,i}, \varepsilon)$ and $\varepsilon > 0$ is a small number.*

Computing the nearest correlation matrix Given an arbitrary symmetric matrix A, the nearest correlation matrix is defined as follows

$$\rho(A) = \min\left\{\|A - X\|_2 : X \text{ is a correlation matrix}\right\}$$

For solving this problem, Higham (2002) proposed to use the method of alternating projections, which consists of iterating $A \leftarrow \mathcal{P}_{\mathcal{U}}(\mathcal{P}_{\mathcal{S}}(A))$ where $\mathcal{P}_{\mathcal{U}}$ and $\mathcal{P}_{\mathcal{S}}$ are the projections on the sets $\mathcal{S} = \left\{X = X^\top : X \geq 0\right\}$ and $\mathcal{U} = \left\{X = X^\top : \text{diag}(X) = \mathbf{1}_n\right\}$. There are different approaches to achieve the convergence. Higham (2002) considered the Dykstra's method given in Algorithm 4. For the projections, we have $\mathcal{P}_{\mathcal{S}}(R) = QT^+Q^*$ where QTQ^* is the Schur decomposition of R and $T_{i,j}^+ = \max(T_{i,j}, 0)$, and $\mathcal{P}_{\mathcal{U}}(Y) = X$ where $X_{i,i} = 1$ and $X_{i,j} = Y_{i,j}$ if $i \neq j$.

Algorithm 4 Computing the nearest correlation matrix

The goal is to compute $\rho(A)$
We set $\Delta S_0 = 0$ and $X_0 = A$
We note ε the convergence criterion of the algorithm
repeat
  $R_k \leftarrow X_{k-1} - \Delta S_{k-1}$
  $Y_k = \mathcal{P}_{\mathcal{S}}(R_k)$
  $\Delta S_k = Y_k - R_k$
  $X_k = \mathcal{P}_{\mathcal{U}}(Y_k)$
until $\|X_k - X_{k-1}\|_2 \leq \varepsilon$
return $\rho(A) \leftarrow X_k$

A.1.2.3 Numerical integration

Trapezoidal and Simpson's rules The general approach to calculate $I(a,b) = \int_a^b f(x)\,\mathrm{d}x$ is to approximate the integral by a sum $\hat{I}(a,b) = \sum_{i=1}^n w_i \cdot f(x_i)$. Let $x_i = a + i \cdot h$

where $h = (b-a)/n$ and n is the number of knots. We have:

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &\approx \sum_{i=1}^{n} \frac{h}{2}\left(f(x_{i-1}) + f(x_i)\right) \\ &= h\left(\frac{1}{2}f(a) + \sum_{i=1}^{n-1} f(x_i) + \frac{1}{2}f(b)\right)\end{aligned}$$

There is no difficulty to implement the trapezoidal method. Moreover, we can show that the quadrature error is:

$$\int_a^b f(x)\,\mathrm{d}x - \hat{I}(a,b) = -\frac{h^2}{12}(b-a) f''(c)$$

where $c \in [a,b]$. The error decreases with the discretization step h and depends on the curvature f''. If the curvature is high, it would be better to use the Simpson's rule. This method consists in replacing the function $f(x)$ on the interval $[x_{i-1}, x_{i+1}]$ by the parabolic function that matches the points $f(x_{i-1})$, $f(x_i)$ and $f(x_{i+1})$. For that, we estimate the curvature by the finite difference:

$$f''(x_i) \approx \frac{f(x_{i-1}) - 2f(x_i) + f(x_{i+1})}{h^2}$$

We obtain:

$$\int_{x_{i-1}}^{x_{i+1}} f(x)\,\mathrm{d}x \approx \frac{h}{3}\left(f(x_{i-1}) + 4f(x_i) + f(x_i)\right)$$

The Simpson's rule is then:

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &\approx \frac{h}{3}\left(f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \ldots + f(b)\right) \\ &= \frac{h}{3}\left(f(a) + 4\sum_{i=1}^{n/2-1} f(x_{2i-1}) + 2\sum_{i=1}^{n/2-1} f(x_{2i}) + f(b)\right)\end{aligned}$$

In this case, the quadrature error becomes:

$$\int_a^b f(x)\,\mathrm{d}x - \hat{I}(a,b) = -\frac{h^4}{180}(b-a) f^{(4)}(c)$$

Gaussian Quadratures One of the most popular methods of numerical integration is the quadrature method with irregular steps when we approximate the function by a polynomial. In the case of Gaussian quadratures, Golub and Welsch (1969) showed that, if $f(x) = B(x)P(x)$ where $P \in \mathcal{P}_{2n-1}$ and $\mathcal{P}_n$ is the set of polynomials of order n, then there exists a set of knots $0 < x_1 < x_2 < \cdots < x_n < 1$ such that[6]:

$$I(0,1) = \int_0^1 f(x)\,\mathrm{d}x = \sum\nolimits_{i=1}^{n} w_i f(x_i)$$

[6] If the support is $[a,b]$, we use the change of variable $y = (x-a)/(b-a)$:

$$\int_a^b f(x)\,\mathrm{d}x = \frac{1}{b-a}\int_0^1 f(a + (b-a)y)\,\mathrm{d}y$$

where w_i are positive weights. If the function $f(x)$ is not a polynomial but sufficiently regular with respect to $P(x)$, $\mathcal{G}(f) = \sum_{i=1}^{n} w_i f(x_i)$ is an approximation of the integral $I(0,1)$. To compute the weights and knots, we have to specify the basis function $B(x)$ and the support. (w_i, x_i) is then the eigenvalue solution of a Jacobi matrix. For example, Figure A.1 shows (w_i, x_i) in the case of Gauss-Legendre quadratures, which are used with functions with a finite support and $B(x) = 1$. An important point is that extension to dimension larger than one is straightforward (Davis and Rabinowitz, 1984).

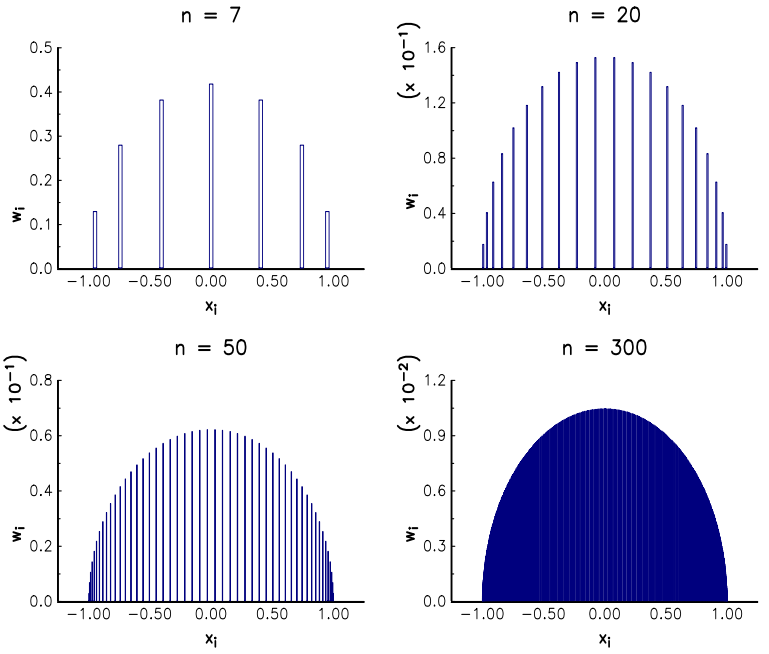


FIGURE A.1: Weights and knots of the Gauss-Legendre quadrature

We consider the function $f(x) = 2\pi\omega\cos(2\pi\omega x) + 2x$. In Figure A.2, we represent $f(x)$, the analytical value $x^2 + \sin(2\pi\omega x)$ of $I(0,x) = \int_0^x f(t)\,\mathrm{d}t$ and the numerical approximation $\hat{I}(0,x)$ when $\omega = 1$ and $\omega = 8$. We notice that the approximation depends on the order n of the quadrature and the upper bound x of the integral. For a fixed value n, the error generally increases with x. In order to understand the accuracy of the numerical solution, we must verify that $f(x)$ is sufficiently regular with respect to the polynomial $P \in \mathcal{P}_{2n-1}$. In Figure A.3, we observe that the adjustment of $f(x)$ for $x \in [0,10]$ is bad when $n = 10$ and $n = 16$, but it is better when $n = 36$ and $n = 200$.

A difficulty concerns functions, whose support is not finite. In Table A.1, we report the value x_n of the last knot for Gauss-Laguerre and Gauss-Hermite quadratures. The use of these methods implies that the approximations $\int_0^{\infty} f(x)\,\mathrm{d}x \simeq \int_0^{x_n} f(x)\,\mathrm{d}x$ and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \simeq \int_{-x_n}^{x_n} f(x)\,\mathrm{d}x$ are valid.

Remark 208 *We can show that the knots corresponds to the roots of the polynomial*[7]*. Once the roots are determined, we can calculate the weights with the following condition (Stoer*

[7]For example, the Legendre polynomial is defined as:

$$P_n(x) = \frac{1}{2^n}\sum_{i=0}^{n}\binom{n}{i}^2 (x-1)^{n-i}(x+1)^n$$

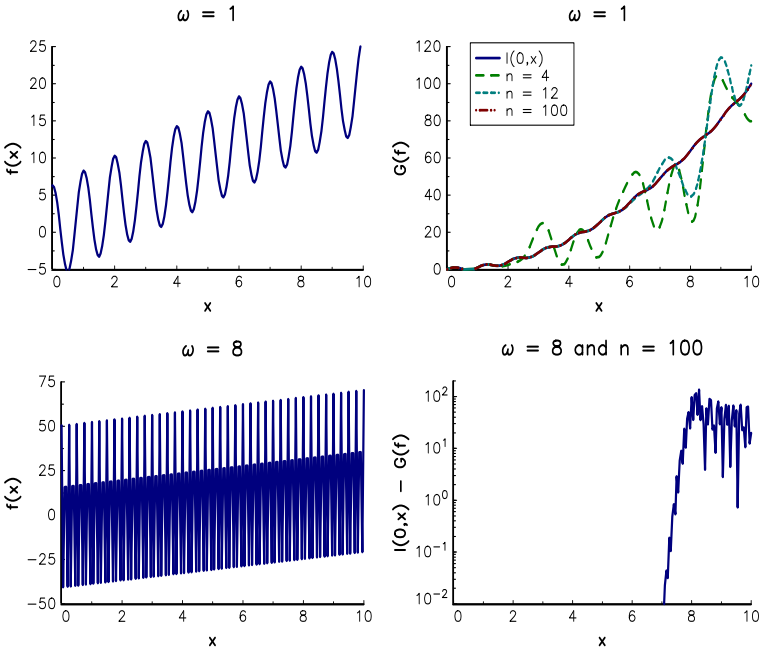


FIGURE A.2: Gauss-Legendre numerical integration

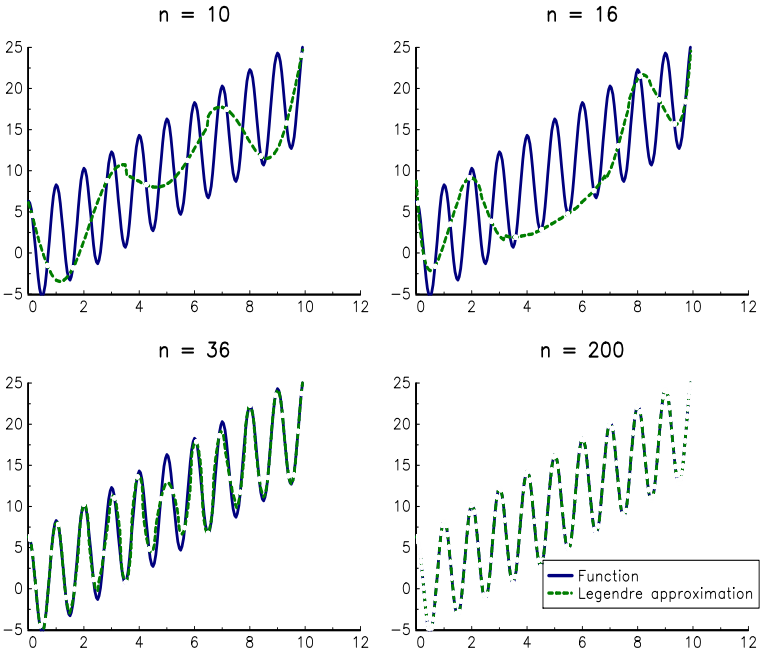


FIGURE A.3: Legendre approximation of $f(x) = 2\pi\cos(2\pi x) + 2x$

TABLE A.1: Value x_n of the last knot in Gauss-Laguerre and Gauss-Hermite quadratures

n	Laguerre	Hermite
4	9.3951	1.6507
8	22.8631	2.9306
16	51.7012	4.6887
32	111.7514	7.1258
100	374.9841	13.4065
200	767.8000	19.3300

and Bulirsch, 1993):

$$\int_a^b w(x)\, x^k P_n(x)\, \mathrm{d}x = 0$$

for all $k = 0, 1, \ldots, n-1$. *The weights are then the solution of a linear system. Abramowitz and Stegun (1970) have tabulated the solution* (w_i, x_i) *for the most known quadratures (Legendre, Laguerre, Hermite, etc.) and different values*[8] *of* n. *These methods are now standard and widely implemented in numerical softwares.*

Quadratures methods can be extended when we consider functions with several variables:

$$\int_a^b \int_c^d f(x,y)\, \mathrm{d}x\, \mathrm{d}y \simeq \sum_{i=1}^n \sum_{j=1}^n w_i w_j f(x_i, x_j)$$

We can also consider non-constant bounds:

$$I = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\, \mathrm{d}x\, \mathrm{d}y$$

and we have:

$$I \simeq \sum_{i=1}^n w_i^{[a,b]} \left(\sum_{j=1}^n w_j^{[g_1(x_i), g_2(x_i)]} f\left(x_i^{[a,b]}, x_j^{[g_1(x_i), g_2(x_i)]} \right) \right)$$

where $\left(w_i^{\Delta}, x_i^{\Delta}\right)$ indicates the weights and knots of the Gauss-Legendre quadrature associated to the support Δ.

A.1.2.4 Finite difference methods

We follow Kurpiel and Roncalli (2000) and consider the linear parabolic equation:

$$\frac{\partial u(t,x)}{\partial t} + c(t,x)\, u(t,x) = \mathcal{A}_t u(t,x) + d(t,x) \tag{A.5}$$

where $\mathcal{A}_t$ is the elliptic differential equation:

$$\mathcal{A}_t u(t,x) = a(t,x) \frac{\partial^2 u(t,x)}{\partial x^2} + b(t,x) \frac{\partial u(t,x)}{\partial x} \tag{A.6}$$

The goal is to solve Equation (A.5) for $t \in [t^-, t^+]$ and $x \in [x^-, x^+]$. In this case, we use the finite difference method, well-adapted for 2-order parabolic equations in x. For that,

[8] Generally, n takes the values 2, 4, 8, 16, 32, 64 and 128.

we introduce a uniform finite-difference mesh for t and x. Let N_t and N_x be the number of discretization points for t and x respectively. We denote by k and h the mesh spacings. We have:

$$\begin{aligned} k &= \frac{t^+ - t^-}{N_t - 1} \\ h &= \frac{x^+ - x^-}{N_x - 1} \end{aligned}$$

and:

$$\begin{aligned} t_m &= t^- + m \cdot k \\ x_i &= x^- + i \cdot h \end{aligned}$$

Let u_i^m be the approximate solution to (A.5) at the grid point (t_m, x_i) and $u(t_m, x_i)$ the exact solution of the partial differential equation at this point.

Discretization in space If we consider the central difference method to approximate the derivatives, we have:

$$\frac{\partial u(t,x)}{\partial x} \approx \frac{u_{i+1}^m - u_{i-1}^m}{2h}$$

and:

$$\frac{\partial^2 u(t,x)}{\partial x^2} \approx \frac{u_{i+1}^m - 2u_i^m + u_{i-1}^m}{h^2}$$

Equation (A.5) becomes:

$$\frac{\partial u(t,x)}{\partial t} + c_i^m u_i^m = \mathbf{A}_i^m + d_i^m$$

where:

$$\mathbf{A}_i^m = a_i^m \frac{u_{i+1}^m - 2u_i^m + u_{i-1}^m}{h^2} + b_i^m \frac{u_{i+1}^m - u_{i-1}^m}{2h}$$

We finally obtain:

$$\frac{\partial u(t,x)}{\partial t} = \mathbf{B}_i^m$$

where:

$$\mathbf{B}_i^m = \mathbf{A}_i^m + d_i^m - c_i^m u_i^m$$

Discretization in time The most classical method to solve Equation (A.5) is to use the Euler scheme. We have:

$$\frac{\partial u(t,x)}{\partial t} \approx \frac{u_i^m - u_i^{m-1}}{k}$$

We also notice that Equation (A.5) becomes:

$$\frac{u_i^m - u_i^{m-1}}{k} + c_i^m u_i^m = \mathcal{A}_t u(t,x) + d_i^m$$

However, the function $\mathcal{A}_t u(t,x)$ depends both on time t and space x. That's why we could not employ the traditional Euler algorithm:

$$u_i^m = u_i^{m-1} + k\left(\mathcal{A}_t u(t,x) + d_i^m - c_i^m u_i^m\right)$$

In this case, we replace the function $\mathcal{A}_t u(t,x)$ by its numerical approximation $\mathbf{A}_i^m$. Therefore, we have:

$$\begin{aligned} u_i^m &= u_i^{m-1} + k\left(\mathbf{A}_i^m + d_i^m - c_i^m u_i^m\right) \\ &= u_i^{m-1} + k\mathbf{B}_i^m \end{aligned}$$

The θ-scheme method In the previous paragraph, we have used the single-sided forward difference to approximate the derivatives $\partial_t u(t,x)$. The θ-scheme method is a combination of left-sided and right-sided differences. Let $\theta \in [0,1]$. We have:

$$u_i^m = u_i^{m-1} + k\left((1-\theta)\,\mathbf{B}_i^{m-1} + \theta \mathbf{B}_i^m\right)$$

Using the expression of $\mathbf{B}_i^m$, we obtain:

$$\begin{aligned}
u_{i-1}^{m-1}\left(a_i^{m-1}(1-\theta)\frac{k}{h^2} - b_i^{m-1}(1-\theta)\frac{k}{2h}\right)\\
+u_i^{m-1}\left(1-2a_i^{m-1}(1-\theta)\frac{k}{h^2} - c_i^{m-1}(1-\theta)k\right)\\
+u_{i+1}^{m-1}\left(a_i^{m-1}(1-\theta)\frac{k}{h^2} + b_i^{m-1}(1-\theta)\frac{k}{2h}\right)\\
+u_{i-1}^{m}\left(a_i^{m}\theta\frac{k}{h^2} - b_i^{m}\theta\frac{k}{2h}\right)\\
+u_{i}^{m}\left(-1-2a_i^{m}\theta\frac{k}{h^2} - c_i^{m}\theta k\right)\\
+u_{i+1}^{m}\left(a_i^{m}\theta\frac{k}{h^2} + b_i^{m}\theta\frac{k}{2h}\right) \quad = \quad -\psi_i^m
\end{aligned}$$

where:

$$\psi_i^m = d_i^{m-1}(1-\theta)k + d_i^m\theta k$$

The different numerical algorithms We introduce the following notations:

$$\begin{aligned}
\alpha_i^m &= a_i^m\frac{k}{h^2} - b_i^m\frac{k}{2h}\\
\beta_i^m &= 1 - 2a_i^m\frac{k}{h^2} - c_i^m k\\
\gamma_i^m &= a_i^m\frac{k}{h^2} + b_i^m\frac{k}{2h}
\end{aligned}$$

The explicit scheme corresponds to $\theta = 0$. We have then:

$$u_i^m = \alpha_i^{m-1}u_{i-1}^{m-1} + \beta_i^{m-1}u_i^{m-1} + \gamma_i^{m-1}u_{i+1}^{m-1} + d_i^{m-1}k \tag{A.7}$$

We obtain the numerical solution by iterating Equation (A.7) from the initial condition and using Dirichlet conditions. The implicit scheme corresponds to $\theta = 1$. We have then:

$$\alpha_i^m u_{i-1}^m + (\beta_i^m - 2)\,u_i^m + \gamma_i^m u_{i+1}^m = -\left(u_i^{m-1} + d_i^m k\right) \tag{A.8}$$

We obtain the numerical solution by solving the linear system (A.8) and using Neumann conditions. In mixed schemes, we have $\theta \in]0,1[$. In particular, we can show that the algorithm is stable if $\theta \geq \frac{1}{2}$. In the general case, the stability assumption is verified if:

$$k \to 0 \bigwedge h \to 0 \bigwedge \frac{k}{h^2} \to 0$$

For instance, the well-famous Crank-Nicholson scheme corresponds to $\theta = \frac{1}{2}$. By introducing the following notations:

$$\begin{aligned}
\varsigma_i^m &= (1-\theta)\,\alpha_i^m \\
\tau_i^m &= 1 + (1-\theta)\,(\beta_i^m - 1) \\
\upsilon_i^m &= (1-\theta)\,\gamma_i^m \\
\phi_i^m &= \theta\alpha_i^m \\
\varphi_i^m &= -1 + \theta\,(\beta_i^m - 1) \\
\chi_i^m &= \theta\gamma_i^m \\
\psi_i^m &= (1-\theta)\,d_i^{m-1}k + \theta d_i^m k
\end{aligned}$$

The linear system to solve becomes:

$$\phi_i^m u_{i-1}^m + \varphi_i^m u_i^m + \chi_i^m u_{i+1}^m = -\left(\varsigma_i^{m-1} u_{i-1}^{m-1} + \tau_i^{m-1} u_i^{m-1} + \upsilon_i^{m-1} u_{i+1}^{m-1} + \psi_i^m\right)$$

The corresponding matrix form is:

$$\Lambda_m \mathbf{u}_m = -\left(\Xi_{m-1}\mathbf{u}_{m-1} + \Psi_m\right) + \varepsilon_m \tag{A.9}$$

where:

$$\mathbf{u}_m = \begin{bmatrix} u_1^m \\ u_2^m \\ \vdots \\ u_i^m \\ \vdots \\ u_{N_x-3}^m \\ u_{N_x-2}^m \end{bmatrix}$$

The Λ_m and Ξ_m matrices are defined in the following manner:

$$\Lambda_m = \begin{bmatrix}
\varphi_1^m & \chi_1^m & 0 & & & \\
\phi_2^m & \varphi_2^m & \chi_2^m & 0 & & \\
\ddots & \ddots & \ddots & \ddots & \ddots & \\
 & 0 & \phi_i^m & \varphi_i^m & \chi_i^m & 0 \\
 & & \ddots & \ddots & \ddots & \ddots \\
 & & & 0 & \phi_{N_x-2}^m & \varphi_{N_x-2}^m
\end{bmatrix}$$

and:

$$\Xi_m = \begin{bmatrix}
\tau_1^m & \upsilon_1^m & 0 & & & \\
\varsigma_2^m & \tau_2^m & \upsilon_2^m & 0 & & \\
\ddots & \ddots & \ddots & \ddots & \ddots & \\
 & 0 & \varsigma_i^m & \tau_i^m & \upsilon_i^m & 0 \\
 & & \ddots & \ddots & \ddots & \ddots \\
 & & & 0 & \varsigma_{N_x-2}^m & \tau_{N_x-2}^m
\end{bmatrix}$$

whereas ε_m is the residual absorbtion vector:

$$\varepsilon_m = \begin{bmatrix} -\left(\phi_1^m u_0^m + \varsigma_1^{m-1} u_0^{m-1}\right) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ -\left(\chi_{N_x-2}^m u_{N_x-1}^m + \upsilon_{N_x-2}^{m-1} u_{N_x-1}^{m-1}\right) \end{bmatrix}$$

Integrating the boundary conditions A new form of the system of equations (A.9) is:

$$\Lambda_m \mathbf{u}_m = \mathbf{v}_m + \varepsilon_m$$

where:

$$\mathbf{v}_m = -\left(\Xi_{m-1}\mathbf{u}_{m-1} + \Psi_m\right)$$

The use of boundary conditions (Dirichlet or/and Neumann) leads us to modify this equation:

$$\Lambda_m^\star \mathbf{u}_m = \mathbf{v}_m^\star$$

where:

$$\begin{aligned} \Lambda_m^\star &\longleftarrow \Lambda_m \\ \mathbf{v}_m^\star &\longleftarrow \mathbf{v}_m \\ \left(\mathbf{v}_m^\star\right)_1 &\longleftarrow -\varsigma_1^{m-1} u_0^{m-1} \\ \left(\mathbf{v}_m^\star\right)_{N_x-1} &\longleftarrow -\upsilon_{N_x-2}^{m-1} u_{N_x-1}^{m-1} \end{aligned}$$

- Conditions on x^-
  - Dirichlet: $u\left(t, x^-\right) = u_{x^-}\left(t\right)$

$$\left(\mathbf{v}_m^\star\right)_1 \longleftarrow -\phi_1^m u_{x^-}\left(t_m\right)$$

  - Neumann: $\partial_x u\left(t, x^-\right) = u'_{x^-}\left(t\right)$

$$\begin{aligned} \left(\Lambda_m^\star\right)_{1,1} &\longleftarrow \phi_1^m \\ \left(\mathbf{v}_m^\star\right)_1 &\longleftarrow \phi_1^m u'_{x^-}\left(t_m\right) h \end{aligned}$$

- Conditions on x^+
  - Dirichlet: $u\left(t, x^+\right) = u_{x^+}\left(t\right)$

$$\left(\mathbf{v}_m^\star\right)_{N_x-2} \longleftarrow -\chi_{N_x-2}^m u_{x^+}\left(t_m\right)$$

  - Neumann: $\partial_x u\left(t, x^+\right) = u'_{x^+}\left(t\right)$

$$\begin{aligned} \left(\Lambda_m^\star\right)_{N_x-2,N_x-2} &\longleftarrow \chi_{N_x-2}^m \\ \left(\mathbf{v}_m^\star\right)_{N_x-2} &\longleftarrow -\chi_{N_x-2}^m u'_{x^+}\left(t_m\right) h \end{aligned}$$

A.1.3 Numerical optimization

A.1.3.1 Quadratic programming problem

A quadratic programming (QP) problem is an optimization problem with a quadratic objective function and linear inequality constraints:

$$\begin{aligned} x^\star &= \arg\min \frac{1}{2}x^\top Qx - x^\top R \\ &\text{s.t.} \quad Sx \leq T \end{aligned} \tag{A.10}$$

where x is a $n \times 1$ vector, Q is a $n \times n$ matrix and R is a $n \times 1$ vector. We note that the system of constraints $Sx \leq T$ allows specifying linear equality constraints[9] $Ax = B$ or weight constraints $x^- \leq x \leq x^+$. Most numerical packages then consider the following formulation:

$$\begin{aligned} x^\star &= \arg\min \frac{1}{2}x^\top Qx - x^\top R \\ &\text{s.t.} \quad \begin{cases} Ax = B \\ Cx \leq D \\ x^- \leq x \leq x^+ \end{cases} \end{aligned} \tag{A.11}$$

because the problem (A.11) is equivalent to the canonical problem (A.10) with the following system of linear inequalities:

$$\begin{bmatrix} -A \\ A \\ C \\ -I_n \\ I_n \end{bmatrix} x \leq \begin{bmatrix} -B \\ B \\ D \\ -x^- \\ x^+ \end{bmatrix}$$

If the space Ω defined by $Sx \leq T$ is non-empty and if Q is a symmetric positive definite matrix, the solution exists because the function $f(x) = \frac{1}{2}x^\top Qx - x^\top R$ is convex. In the general case where Q is a square matrix, the solution may not exist.

The Lagrange function is also:

$$\mathcal{L}(x;\lambda) = \frac{1}{2}x^\top Qx - x^\top R + \lambda^\top (Sx - T)$$

We deduce that the dual problem is defined by:

$$\begin{aligned} \lambda^\star &= \arg\max \left\{ \inf_x \mathcal{L}(x;\lambda) \right\} \\ &\text{s.t.} \quad \lambda \geq 0 \end{aligned}$$

We note that $\partial_x \mathcal{L}(x;\lambda) = Qx - R + S^\top \lambda$. The solution to the problem $\partial_x \mathcal{L}(x;\lambda) = 0$ is then $x = Q^{-1}\left(R - S^\top \lambda\right)$. We obtain:

$$\begin{aligned} \inf_x \mathcal{L}(x;\lambda) &= \frac{1}{2}\left(R^\top - \lambda^\top S\right) Q^{-1}\left(R - S^\top \lambda\right) - \left(R^\top - \lambda^\top S\right) Q^{-1} R + \\ &\quad \lambda^\top \left(SQ^{-1}\left(R - S^\top \lambda\right) - T\right) \\ &= \frac{1}{2}R^\top Q^{-1}R - \lambda^\top SQ^{-1}R + \frac{1}{2}\lambda^\top SQ^{-1}S^\top \lambda - R^\top Q^{-1}R + \\ &\quad 2\lambda^\top SQ^{-1}R - \lambda^\top SQ^{-1}S^\top \lambda - \lambda^\top T \\ &= -\frac{1}{2}\lambda^\top SQ^{-1}S^\top \lambda + \lambda^\top \left(SQ^{-1}R - T\right) - \frac{1}{2}R^\top Q^{-1}R \end{aligned}$$

[9] This is equivalent to imposing that $Ax \geq B$ and $Ax \leq B$.

The dual program is another quadratic program:

$$\begin{aligned} \lambda^\star &= \arg\min \frac{1}{2}\lambda^\top \bar{Q}\lambda - \lambda^\top \bar{R} \\ &\text{s.t.} \quad \lambda \geq 0 \end{aligned} \tag{A.12}$$

where $\bar{Q} = SQ^{-1}S^\top$ and $\bar{R} = SQ^{-1}R - T$.

A.1.3.2 Non-linear unconstrained optimization

We consider the minimization problem:

$$x^\star = \arg\min f(x) \tag{A.13}$$

where $x \in \mathbb{R}^n$. Let $G(x)$ and $H(x)$ be the gradient vector and the Hessian matrix of $f(x)$. The optimum verifies:

$$G(x^\star) = \mathbf{0} \tag{A.14}$$

The first-order Taylor expansion of $G(x)$ around the point x_0 is given by:

$$G(x) = G(x_0) + H(x_0)(x - x_0)$$

If x is the solution of Equation (A.14), we obtain $G(x_0) + H(x_0)(x - x_0) = \mathbf{0}$. The Newton-Raphson algorithm uses an iterative process to find this root:

$$x_{k+1} = x_k - H_k^{-1} G_k$$

where k is the iteration index, $G_k = G(x_k)$ and $H_k = H(x_k)$. Starting from an initial point x_0, we find the solution $x^\star$ if the algorithm converges[10]. However, we generally prefer to use the following process:

$$\begin{aligned} x_{k+1} &= x_k - \lambda_k H_k^{-1} G_k \\ &= x_k + \lambda_k d_k \end{aligned}$$

where $\lambda_k > 0$ is a scalar. The difference comes from the introduction of the step length λ_k. Starting from the point x_k, the vector $d_k = -H_k^{-1} G_k$ indicates the direction to reach the maximum. Nevertheless, using a step length equal to 1 is not always optimal. For instance, we could exceed the optimum[11] or the convergence may be very slow. This is why numerical optimization methods use two types of algorithms:

1. an algorithm to approximate the Hessian matrix H_k and to compute the descent d_k;
2. a second algorithm to define the optimal step length λ_k:

$$\lambda_k = \arg\min_{\lambda > 0} f(x_k + \lambda d_k)$$

The Hessian approximation avoids singularity problems which are frequent in the neighborhood of the optimum. Press *et al.* (2007) distinguished two algorithm families to define the descent, namely conjugate gradient and quasi-Newton methods.

[10] We stop the algorithm when the gradient is close to zero. For example, the stopping rule may be $\max_i \left| G_{k,i} \right| \leq \varepsilon$ where ε is the allowed tolerance.

[11] This means that f does not necessarily decrease at each iteration.

- In the case of the conjugate gradient approach, we have:

$$d_{k+1} = -\left(G_{k+1} - \varrho_{k+1} d_k\right)$$

For the Polak-Ribiere algorithm, the scalar ϱ is given by:

$$\varrho_{k+1} = \frac{G_{k+1}^\top G_{k+1}}{G_k^\top G_k}$$

whereas for the Fletcher-Reeves algorithm, we have:

$$\varrho_{k+1} = \frac{\left(G_{k+1} - G_k\right)^\top G_{k+1}}{G_k^\top G_k}$$

- For quasi-Newton methods, the direction is defined as follows:

$$d_{k+1} = -\tilde{H}_{k+1} G_{k+1}$$

where $\tilde{H}$ is an approximation of the inverse of the Hessian matrix. Its expression is:

$$\begin{aligned}\tilde{H}_{k+1} &= \tilde{H}_k - \frac{\tilde{H}_k y_k y_k^\top \tilde{H}_k}{y_k^\top \tilde{H}_k y_k} + \frac{s_k s_k^\top}{s_k^\top y_k} + \\ &\quad \beta \left(\tilde{H}_k y_k - \theta_k s_k\right)\left(\tilde{H}_k y_k - \theta_k s_k\right)^\top\end{aligned}$$

where $y_k = G_{k+1} - G_k$, $s_k = x_{k+1} - x_k$ and:

$$\theta_k = \frac{y_k^\top \tilde{H}_k y_k}{s_k^\top y_k}$$

The Davidon, Fletcher and Powell (DFP) algorithm corresponds to $\beta = 0$, whereas the Broyden, Fletcher, Goldfarb and Shanno (BFGS) algorithm is given by:

$$\beta = \frac{1}{y_k^\top \tilde{H}_k y_k}$$

To find the optimal value of λ_k, we employ a simple one-dimension minimization algorithm[12] such as the golden section, Brent's method or the cubic spline approximation (Press *et al.*, 2007).

Remark 209 *Newton's method may also be used to solve non-linear optimization problems with linear constraints:*

$$\begin{aligned}x^\star &= \arg\min f(x) \\ &s.t. \quad Ax = B\end{aligned}$$

Indeed, this constrained problem is equivalent to the following unconstrained problem:

$$y^\star = \arg\min g(y)$$

where $g(y) = f(Cy + D)$, C *is an orthonormal basis for the nullspace of* A, $D = \left(A^\top A\right)^+ A^\top B$ *and* $\left(A^\top A\right)^+$ *is the Moore-Penrose pseudo-inverse of* $A^\top A$. *The solution is then:*

$$x^\star = Cy^\star + D$$

[12] Computing the optimal value of λ_k may be time consuming. In this case, we may also prefer the half method which consists in dividing the test value by one half each time the function fails to decrease – λ then takes the respective values 1, 1/2, 1/4, 1/8, etc. – and to stop when the criterion $f(x_k + \lambda_k d_k) < f(x_k)$ is satisfied.

A.1.3.3 Sequential quadratic programming algorithm

The sequential quadratic programming (or SQP) algorithm solves this constrained non-linear programming problem:

$$\begin{array}{rcl} x^\star & = & \arg\min f(x) \\ & \text{s.t.} & \left\{ \begin{array}{l} A(x) = \mathbf{0} \\ B(x) \geq \mathbf{0} \end{array} \right. \end{array} \tag{A.15}$$

where $A(x)$ and $B(x)$ are two multidimensional non-linear functions. Like Newton's methods, this algorithm is an iterative process:

$$x_{k+1} = x_k + \lambda_k d_k$$

where:

$$\begin{array}{rcl} d_k & = & \arg\min \dfrac{1}{2} d^\top H_k d + d^\top G_k \\ & \text{s.t.} & \left\{ \begin{array}{l} \partial_x A(x_k)\, d + A(x_k) = \mathbf{0} \\ \partial_x B(x_k)\, d + B(x_k) \geq \mathbf{0} \end{array} \right. \end{array}$$

It consists in replacing the non-linear optimization problem by a sequence of quadratic programming problems (Boggs and Tolle, 1995). The QP problem corresponds to the second-order Taylor expansion of $f(x)$:

$$f(x_k + \delta) = f(x_k) + \delta^\top G_k + \frac{1}{2}\delta^\top H_k \delta$$

where:

$$\left\{ \begin{array}{l} A(x_k + \delta) = A(x_k) + \partial_x A(x_k)\, \delta = \mathbf{0} \\ B(x_k + \delta) = B(x_k) + \partial_x B(x_k)\, \delta \geq \mathbf{0} \end{array} \right.$$

and $\delta = \lambda d$. We can use quasi-Newton methods to approximate the Hessian matrix H_k. However, if we define λ_k as previously:

$$\lambda_k = \min_{\lambda > 0} f(x_k + \lambda d_k)$$

we may face some problems because the constraints $A(x) = \mathbf{0}$ and $B(x) \geq \mathbf{0}$ are not necessarily satisfied. This is why we prefer to specify λ_k as the solution to the one-dimensional minimization problem:

$$\lambda_k = \min_{\lambda > 0} m(x_k + \lambda d_k)$$

where $m(x)$ is the merit function:

$$m(x) = f(x) + p_A \sum_j |A_j(x)| - p_B \sum_j \min(0, B_j(x))$$

We generally choose the penalization weights p_A and p_B as the infinite norm of Lagrange coefficients associated with linear and non-linear constraints (Nocedal and Wright, 2006).

A.1.3.4 Dynamic programming in discrete time with finite states

We note k the discrete time where $k \in \{1, \ldots, K\}$. Let $s(k)$ and $c(k)$ be the state and control variables. We assume that the state variable evolves according to the dynamics:

$$s(k+1) = g(k, s(k), c(k))$$

The state variable $s(k+1)$ only depends on the previous value $s(k)$ and not on the entire path of the state. However, it can be controlled thanks to $c(k)$. Knowing the initial value of the state $s(1) = s$, we would like to find the optimal control $c^\star(k)$ that maximizes the additive gain function:

$$\{c^\star(k)\}_{k=1}^{K-1} = \arg\max \mathcal{J}(s, c(1), \ldots, c(K-1)) \tag{A.16}$$

where:

$$\mathcal{J}(s, c(1), \ldots, c(K-1)) = \sum_{k=1}^{K-1} f(k, s(k), c(k)) + f(K, s(K))$$

$f(k, s(k), c(k))$ is the gain function at time k, whereas $f(K, s(K))$ is the terminal gain. We impose that the state and control variables satisfy some constraints: $s(k) \in \mathcal{S}(k)$ and $c(k) \in \mathcal{C}(k)$. The optimization problem becomes:

$$\begin{aligned} \{c^\star(k)\}_{k=1}^{K-1} &= \arg\max \sum_{k=1}^{K-1} f(k, s(k), c(k)) + f(K, s(K)) \\ \text{s.t.} &\begin{cases} s(k+1) = g(k, s(k), c(k)) \\ s(k) \in \mathcal{S}(k) \\ c(k) \in \mathcal{C}(k) \\ s(1) = s \end{cases} \end{aligned} \tag{A.17}$$

A policy $\pi = \{\mu(1), \ldots, \mu(K-1)\}$ is described by functions $\mu(k) = \mu(k, s(k))$ that map states into controls (Bertsekas, 2005). The optimal policy $\pi^\star$ is then defined as follows:

$$\begin{aligned} \pi^\star &= \arg\max \sum_{k=1}^{K-1} f(k, s(k), \mu(k, s(k))) + f(K, s(K)) \\ \text{s.t.} &\begin{cases} s(k+1) = g(k, s(k), \mu(k, s(k))) \\ s(k) \in \mathcal{S}(k) \\ \mu(k, s(k)) \in \mathcal{C}(k) \\ s(1) = s \end{cases} \end{aligned} \tag{A.18}$$

This problem may be solved with the method of dynamic programming introduced by Bellman (1957). Let $\kappa \in \mathbb{N}$ such that $1 \leq \kappa \leq K-1$. We consider the tail subproblem:

$$\pi^\star(\kappa) = \arg\max \sum_{k=\kappa}^{K-1} f(k, s(k), \mu(k, s(k))) + f(K, s(K)) \tag{A.19}$$

with the same set of constraints that those used for Problem (A.18) Bellman's optimality principle states that if $\pi^\star = (\mu^\star(1), \ldots, \mu^\star(K-1))$ is an optimal policy for Problem (A.18), then the tail policy $\pi^\star(\kappa) = \{\mu^\star(\kappa), \ldots, \mu^\star(K-1)\}$ is an optimal policy for Problem (A.19). Therefore, we can solve Problem (A.18) using a backward algorithm, which is characterized by a set of recursive optimizations (Bertsekas, 2005):

1. at the terminal date, we have:

$$\mathcal{J}(K, s(K)) = f(K, s(K))$$

2. at the intermediate date $k < K$, we have:

$$\mathcal{J}(k, s(k)) = \sup_{c(k) \in \mathcal{C}(k)} \{f(k, s(k), c(k)) + \mathcal{J}(k+1, g(k, s(k), c(k)))\} \tag{A.20}$$

In the finite case where there are n_S states and n_C controls, the previous algorithm is simplified. Let $\{s_1, \ldots, s_{n_S}\}$ and $\{c_1, \ldots, c_{n_C}\}$ be the values taken by $s(k)$ and $c(k)$. We note $\mathcal{J}(K, s_i)$ the terminal value when $s(K) = s_i$. We store the solutions $\mathcal{J}(k, s(k))$ and $c^\star(k)$ into the matrices $\boldsymbol{J}$ and $\boldsymbol{C}$. The DP algorithm becomes:

1. We initialize the algorithm by $k = K$.
2. At the time K, we know the terminal value $\mathcal{J}(K, s(K))$. Therefore, we initialize the element (i, K) of the matrix $\boldsymbol{J}$ to $\mathcal{J}(K, s_i)$.
3. We set $k \longleftarrow k - 1$.
4. At the date $k < K$, we calculate for each state s_i the value taken by $\mathcal{J}(k, s_i)$:

$$\mathcal{J}(k, s_i) = \sup_{1 \leq j \leq n_C} \{f(k, s_i, c_j) + \mathcal{J}(k+1, s')\}$$

   where $s' = g(k, s_i, c_j)$. By construction, s' corresponds to a state $s_{i'}$. We deduce that $\mathcal{J}(k+1, s')$ is equal to the element $(i', k+1)$ of the matrix $\boldsymbol{J}$. Moreover, the optimal control $c^\star(k)$ is given by:

$$c^\star(k) = \arg\max_{1 \leq j \leq n_C} \{f(k, s_i, c_j) + \mathcal{J}(k+1, s')\}$$

   We deduce that the element (i, k) of the matrices $\boldsymbol{J}$ and $\boldsymbol{C}$ are $\mathcal{J}(k, s_i)$ and $c^\star(k)$.
5. If $k = 1$, we stop the algorithm. Otherwise, we go to step 3.

A.2 Statistical and probability analysis

A.2.1 Probability distributions

A.2.1.1 The Bernoulli distribution

The Bernoulli random variable X takes the value 1 with success probability of p and the value 0 with failure probability of $q = 1 - p$. We note $X \sim \mathcal{B}(p)$. The probability mass function may also be expressed as follows:

$$\Pr\{X = k\} = p^k (1-p)^{1-k} \qquad \text{with } k = 0, 1$$

We have $\mathbb{E}[X] = p$ and $\text{var}(X) = p(1-p)$.

A.2.1.2 The binomial distribution

The binomial random variable X is the sum of n independent Bernoulli random variables with the same probability of success p:

$$X = \sum_{i=1}^{n} \mathcal{B}_i(p)$$

We note $X \sim \mathcal{B}(n, p)$. The probability mass function is equal to:

$$\Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{with } k = 0, 1, \ldots, n$$

We have $\mathbb{E}[X] = np$ and $\text{var}(X) = np(1-p)$.

A.2.1.3 The geometric distribution

The geometric random variable X is the number of Bernoulli trials needed to get one success. We note $X \sim \mathcal{G}(p)$. The probability mass function is equal to:

$$\Pr\{X = k\} = (1-p)^{k-1} p \qquad \text{with } k \in \mathbb{N}^\star$$

We have $\mathbb{E}[X] = 1/p$ and $\text{var}(X) = (1-p)/p^2$.

Remark 210 *If we define X as the number of failures before the first success, we have $\Pr\{X = k\} = (1-p)^k p$ with $k \in \mathbb{N}$, $\mathbb{E}[X] = (1-p)/p$ and $\text{var}(X) = (1-p)/p^2$.*

A.2.1.4 The Poisson distribution

The Poisson random variable X is the number of times an event occurs in the unit interval of time. We note $X \sim \mathcal{P}(\lambda)$ where λ is the parameter of the Poisson distribution. The probability mass function is equal to:

$$\Pr\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{with } k \in \mathbb{N}$$

We have $\mathbb{E}[X] = \text{var}(X) = \lambda$. The parameter λ is then the expected number of events occurring in the unit interval of time.

A.2.1.5 The negative binomial distribution

The negative binomial distribution is another probability distribution for modeling the frequency of an event. We note $X \sim \mathcal{NB}(r, p)$ where $r > 0$ and $p \in [0, 1]$. The probability mass function is equal to:

$$\Pr\{X = k\} = \binom{r+k-1}{k} (1-p)^r p^k \qquad \text{with } k \in \mathbb{N}$$

We have $\mathbb{E}[X] = pr/(1-p)$ and $\text{var}(X) = pr/(1-p)^2$.

A.2.1.6 The gamma distribution

The gamma distribution is a two-parameter family of continuous probability distributions, whose support is $[0, \infty)$. We note $X \sim \mathcal{G}(\alpha, \beta)$ where $\alpha > 0$ and $\beta > 0$. α and β are called the shape parameter and the rate parameter. The probability density function is equal to:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

where $\Gamma(\alpha)$ is the gamma function defined as:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, \mathrm{d}t$$

The cumulative distribution function is the regularized gamma function:

$$\mathbf{F}(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

where $\gamma(\alpha, x)$ is the lower incomplete gamma function defined as:

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} \, \mathrm{d}t$$

We have $\mathbb{E}[X] = \alpha/\beta$ and $\text{var}(X) = \alpha/\beta^2$. We verify the following properties:

- $\mathcal{G}(1,\beta) \sim \mathcal{E}(\beta)$;
- if $X \sim \mathcal{G}(\alpha,\beta)$, then $cX \sim \mathcal{G}(\alpha,\beta/c)$ when $c>0$;
- $\sum_{i=1}^{n} \mathcal{G}(\alpha_i,\beta) \sim \mathcal{G}\left(\sum_{i=1}^{n}\alpha_i,\beta\right)$.

Remark 211 *The standard gamma distribution corresponds to* $\mathcal{G}(\alpha,1)$ *and is denoted by* $\mathcal{G}(\alpha)$.

A.2.1.7 The beta distribution

The beta distribution is a two-parameter family of continuous probability distributions defined on the interval $[0,1]$. We note $X \sim \mathcal{B}(\alpha,\beta)$ where $\alpha>0$ and $\beta>0$. The probability density function is equal to:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathfrak{B}(\alpha,\beta)}$$

where $\mathfrak{B}(\alpha,\beta)$ is the gamma function defined as:

$$\begin{aligned}\mathfrak{B}(\alpha,\beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\end{aligned}$$

The cumulative distribution function is the regularized incomplete beta function.

$$\begin{aligned}\mathbf{F}(x) &= \mathcal{IB}(x;\alpha,\beta) \\ &= \frac{\mathfrak{B}(x;\alpha,\beta)}{\mathfrak{B}(\alpha,\beta)}\end{aligned}$$

where $\mathfrak{B}(x;\alpha,\beta)$ is the incomplete beta function defined as:

$$\mathfrak{B}(x;\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$$

We have $\mathbb{E}[X] = \alpha/(\alpha+\beta)$ and:

$$\mathrm{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

A.2.1.8 The noncentral chi-squared distribution

Let $(X_1,\ldots,X_\nu)$ be a set of independent Gaussian random variables such that $X_i \sim \mathcal{N}\left(\mu_i,\sigma_i^2\right)$. The noncentral chi-squared random variable is defined as follows:

$$Y = \sum_{i=1}^{\nu} \frac{X_i^2}{\sigma_i^2}$$

We write $Y \sim \chi_\nu^2(\zeta)$ where ν is the number of degrees of freedom and ζ is the noncentrality parameter:

$$\zeta = \sum_{i=1}^{\nu} \frac{\mu_i^2}{\sigma_i^2}$$

The cumulative distribution function of Y is defined as:

$$\mathbf{F}(y;\nu,\zeta) = \Pr\{Y \leq y\} = \sum_{j=0}^{\infty} \frac{e^{-\zeta/2}\zeta^j}{2^j j!}\mathbf{F}(y;\nu+2j,0)$$

where $\mathbf{F}(y;\nu,0)$ is the cumulative distribution function of the chi-squared distribution with ν degrees of freedom. We deduce that the probability density function is:

$$f(y;\nu,\zeta) = \sum_{j=0}^{\infty} \frac{e^{-\zeta/2}\zeta^j}{2^j j!} f(y;\nu+2j,0)$$

where $f(y;\nu,0)$ is the probability density function of the chi-squared distribution. We may also show that the mean and the variance of Y are $\nu+\zeta$ and $2(\nu+2\zeta)$. For the skewness and excess kurtosis coefficients, we obtain:

$$\begin{aligned}\gamma_1 &= (\nu+3\zeta)\sqrt{\frac{2^3}{(\nu+2\zeta)^3}} \\ \gamma_2 &= \frac{12(\nu+4\zeta)}{(\nu+2\zeta)^2}\end{aligned}$$

Remark 212 *When μ_i is equal to zero, Y becomes a (central) chi-squared distribution $\chi^2_\nu(0)$. The density function is equal to:*

$$f(y;\nu,0) = \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}$$

whereas the cumulative distribution function has the following expression:

$$\mathbf{F}(y;\nu,0) = \frac{\gamma(\nu/2, x/2)}{\Gamma(\nu/2)}$$

A.2.1.9 The exponential distribution

X is an exponential random variable $\mathcal{E}(\lambda)$ if the density function is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. We deduce that $\mathbf{F}(x) = 1 - e^{-\lambda x}$. We have $\mathbb{E}[X] = 1/\lambda$ and $\text{var}(X) = 1/\lambda^2$. More generally, we can show that $\mathbb{E}[X^n] = n!/\lambda^n$. This distribution verifies the lack of memory property:

$$\Pr\{X \geq s+t \mid X \geq s\} = \Pr\{X \geq t\}$$

for all $s \geq 0$ and $t \geq 0$.

A.2.1.10 The normal distribution

Let $\mathbb{C}$ be a correlation matrix. We consider the standardized Gaussian random vector $X \sim \mathcal{N}(\mathbf{0}, \mathbb{C})$ of dimension n. We note $\phi_n(x;\mathbb{C})$ the associated density function defined as:

$$\phi_n(x;\mathbb{C}) = (2\pi)^{-n/2}|\mathbb{C}|^{-1/2}\exp\left(-\frac{1}{2}x^\top \mathbb{C}^{-1}x\right)$$

We deduce that the expression of cumulative distribution function is:

$$\Phi_n(x;\mathbb{C}) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_2}\phi_n(u;\mathbb{C})\,\mathrm{d}u$$

By construction, we have $\mathbb{E}[X] = \mathbf{0}$ and $\text{cov}(x) = \mathbb{C}$. In the bivariate case, we use the notations $\phi_2(x_1, x_2; \rho) = \phi_2(x; \mathbb{C})$ and $\Phi_2(x_1, x_2; \rho) = \Phi_2(x; \mathbb{C})$ where $\rho = \mathbb{C}_{1,2}$ is the correlation between the components X_1 and X_2. In the univariate case, we also consider the alternative notations $\phi(x) = \phi_1(x; 1)$ and $\Phi(x) = \Phi_1(x; 1)$. The density function reduces then to:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

Concerning the moments, we have $\mu(X) = 0$, $\sigma(X) = 1$, $\gamma_1(X) = 0$ and $\gamma_2(X) = 0$.

Adding a mean vector μ and a covariance matrix Σ is equivalent to apply the linear transformation to X:

$$Y = \mu + \sigma X$$

where $\sigma = \text{diag}^{1/2}(\Sigma)$.

A.2.1.11 The Student's t distribution

Let $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and $V \sim \chi^2_\nu / \nu$ be independent of X. We define the multivariate Student's t distribution as the one corresponding to the linear transformation:

$$Y = V^{-1/2} X$$

The corresponding density function is:

$$\mathbf{t}_n(y; \Sigma, \nu) = \frac{\Gamma\left((\nu+n)/2\right)}{\Gamma(\nu/2)(\nu\pi)^{n/2}} |\Sigma|^{-1/2} \left(1 + \frac{1}{\nu} y^\top \Sigma^{-1} y\right)^{-(\nu+n)/2}$$

We note $\mathbf{T}_n(y; \Sigma, \nu)$ the cumulative density function:

$$\mathbf{T}_n(y; \Sigma, \nu) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_2} \mathbf{t}_n(u; \Sigma, \nu)\, \mathrm{d}u$$

The first two moments[13] of Y are $\mathbb{E}[Y] = \mathbf{0}$ and $\text{cov}(Y) = \nu(\nu-2)^{-1}\Sigma$. Adding a mean μ is equivalent to consider the random vector $Z = \mu + Y$. We also verify that Y tends to the Gaussian random vector X when the number of degrees of freedom tends to ∞.

In the univariate case, the standardized density function becomes:

$$\mathbf{t}_1(y; \nu) = \frac{\Gamma\left((\nu+1)/2\right)}{\Gamma(\nu/2)\sqrt{\nu\pi}} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}$$

We also use the alternative notations $\mathbf{t}_\nu(y) = \mathbf{t}_1(y; \nu)$ and $\mathbf{T}_\nu(y) = \mathbf{T}_1(y; \nu)$. Concerning the moments[14], we obtain $\mu(Y) = 0$, $\sigma^2(Y) = \nu/(\nu-2)$, $\gamma_1(Y) = 0$ and $\gamma_2(Y) = 6/(\nu-4)$.

A.2.1.12 The log-normal distribution

Let $Z \sim \mathcal{N}(\mu, \sigma^2)$ be a normal-distributed random variable. $X = e^Z$ is a log-normal random variable and we note $X \sim \mathcal{LN}(\mu, \sigma^2)$. The probability distribution function is equal to:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

[13] The second moment is not defined if $\nu \leq 2$.

[14] The skewness is not defined if $\nu \leq 3$ whereas the excess kurtosis is infinite if $2 < \nu \leq 4$.

whereas the cumulative distribution function has the following expression:

$$\mathbf{F}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

We have:

$$\mathbb{E}[X] = e^{\mu+\frac{1}{2}\sigma^2}$$

and:

$$\text{var}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

A.2.1.13 The Pareto distribution

The Pareto distribution is denoted by $\mathcal{P}(\alpha, x_-)$. We have:

$$f(x) = \frac{\alpha}{x}\left(\frac{x}{x_-}\right)^{-\alpha}$$

and:

$$\mathbf{F}(x) = 1 - \left(\frac{x}{x_-}\right)^{-\alpha}$$

where $x \geq x_-$, $\alpha > 0$ and $x_- > 0$. Concerning the first two moments, we obtain:

$$\mathbb{E}[X] = \frac{\alpha x_-}{\alpha - 1}$$

if $\alpha > 1$ and:

$$\text{var}(X) = \frac{\alpha x_-^2}{(\alpha-1)^2(\alpha-2)}$$

if $\alpha > 2$.

Remark 213 *The Pareto distribution: can be parameterized as follows;*

$$\mathbf{F}(x) = 1 - \left(\frac{\theta + x}{\theta}\right)^{-\alpha}$$

where $x \geq 0$, $\alpha > 0$ and $\theta > 0$. In this case, it is denoted by $\mathcal{P}(\alpha, \theta)$.

A.2.1.14 The generalized extreme value distribution

The generalized extreme value distribution is denoted by $\mathcal{GEV}(\mu, \sigma, \xi)$. We have:

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-(1+1/\xi)} \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

and:

$$\mathbf{F}(x) = \exp\left(-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

where $x > \mu - \sigma/\xi$, $\sigma > 0$ and $\xi > 0$. Concerning the first two moments, we obtain:

$$\mathbb{E}[X] = \mu + \frac{\sigma}{\xi}(\Gamma(1-\xi) - 1)$$

if $\xi < 1$ and:

$$\text{var}(X) = \frac{\sigma^2}{\xi^2}\left(\Gamma(1-2\xi) - \Gamma^2(1-\xi)\right)$$

if $\xi < \frac{1}{2}$.

A.2.1.15 The generalized Pareto distribution

The generalized Pareto distribution is denoted by $\mathcal{GPD}(\sigma, \xi)$. We have:

$$f(x) = \frac{1}{\sigma}\left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi - 1}$$

and:

$$\mathbf{F}(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}$$

where $x \geq 0$, $\sigma > 0$ and $\xi > 0$. Concerning the first two moments, we obtain:

$$\mathbb{E}[X] = \frac{\sigma}{1 - \xi}$$

if $\xi < 1$ and:

$$\text{var}(X) = \frac{\sigma^2}{(1 - \xi)^2 (1 - 2\xi)}$$

if $\xi < \frac{1}{2}$.

A.2.1.16 The skew normal distribution

The seminal work of Azzalini (1985) has led to a rich development on skew distributions with numerous forms, parameterizations and extensions[15]. We adopt here the construction of Azzalini and Dalla Valle (1996).

The multivariate case Azzalini and Dalla Valle (1996) define the density function of the skew normal (or SN) distribution as follows:

$$f(x) = 2\phi_n(x - \xi; \Omega)\, \Phi_1\left(\eta^\top \omega^{-1}(x - \xi)\right)$$

with $\omega = \text{diag}^{1/2}(\Omega)$. We say that X follows a multivariate skew normal distribution with parameters ξ, Ω and η and we write $X \sim \mathcal{SN}(\xi, \Omega, \eta)$. We notice that the distribution of $X \sim \mathcal{SN}_n(\xi, \Omega, \mathbf{0})$ is the standard normal distribution $\mathcal{N}(\xi, \Omega)$. We verify the property $X = \xi + \omega Y$ where $Y \sim \mathcal{SN}(\mathbf{0}, \mathbb{C}, \eta)$ and $\mathbb{C} = \omega^{-1}\Omega\omega^{-1}$ is the correlation matrix of Ω. Azzalini and Dalla Valle (1996) demonstrated that the first two moments are:

$$\begin{aligned} \mathbb{E}[X] &= \xi + \sqrt{\frac{2}{\pi}}\omega\delta \\ \text{cov}(X) &= \omega\left(\mathbb{C} - \frac{2}{\pi}\delta\delta^\top\right)\omega^\top \end{aligned}$$

where $\delta = \left(1 + \eta^\top \mathbb{C}\eta\right)^{-1/2} \mathbb{C}\eta$.

Azzalini and Capitanio (1999) showed that $Y \sim \mathcal{SN}(\mathbf{0}, \mathbb{C}, \eta)$ has the following stochastic representation:

$$Y = \begin{cases} U & \text{if } U_0 > 0 \\ -U & \text{otherwise} \end{cases}$$

where:

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbb{C}_+(\delta)), \qquad \mathbb{C}_+(\delta) = \begin{pmatrix} 1 & \delta^\top \\ \delta & \mathbb{C} \end{pmatrix}$$

[15] See for instance Arellano-Valle and Genton (2005) and Lee and Wang (2013) for a review.

and $\delta = \left(1 + \eta^\top \mathbb{C}\eta\right)^{-1/2} \mathbb{C}\eta$. We deduce that:

$$\begin{aligned}
\Pr\{X \leq x\} &= \Pr\left\{Y \leq \omega^{-1}(x-\xi)\right\} \\
&= \Pr\left\{U \leq \omega^{-1}(x-\xi) \mid U_0 > 0\right\} \\
&= \frac{\Pr\left\{U \leq \omega^{-1}(x-\xi), U_0 > 0\right\}}{\Pr\{U_0 > 0\}} \\
&= 2\left(\Pr\left\{U \leq \omega^{-1}(x-\xi)\right\} - \Pr\left\{U \leq \omega^{-1}(x-\xi), U_0 \leq 0\right\}\right) \\
&= 2\left(\Phi_n\left(\omega^{-1}(x-\xi); \mathbb{C}\right) - \Phi_{n+1}\left(u_+; \mathbb{C}_+(\delta)\right)\right) \\
&= 2\Phi_{n+1}\left(u_+; \mathbb{C}_+(-\delta)\right)
\end{aligned}$$

where $u_+ = \left(0, \omega^{-1}(x-\xi)\right)$. We can therefore use this representation to simulate the random vector $X \sim \mathcal{SN}(\xi, \Omega, \eta)$ and compute the cumulative distribution function.

Let A be a $m \times n$ matrix and $X \sim \mathcal{SN}(\xi, \Omega, \eta)$. Azzalini and Capitanio (1999) demonstrated that the linear transformation of a skew normal vector is still a skew normal vector:

$$AX \sim \mathcal{SN}(\xi_A, \Omega_A, \eta_A)$$

where:

$$\begin{aligned}
\xi_A &= A\xi \\
\Omega_A &= A\Omega A^\top \\
\eta_A &= \frac{\omega_A \Omega_A^{-1} B^\top \eta}{\left(1 + \eta^\top \left(\mathbb{C} - B\Omega_A^{-1} B^\top\right)\eta\right)^{1/2}}
\end{aligned}$$

with $\omega = \text{diag}^{1/2}(\Omega)$, $\mathbb{C} = \omega^{-1}\Omega\omega$, $\omega_A = \text{diag}^{1/2}(\Omega_A)$ and $B = \omega^{-1}\Omega A^\top$. This property also implies that the marginal distributions of a subset of X is still a skew normal distribution.

The univariate case When the dimension n is equal to 1, the density function of $X \sim \mathcal{SN}\left(\xi, \omega^2, \eta\right)$ becomes:

$$f(x) = \frac{2}{\omega} \cdot \phi\left(\frac{x-\xi}{\omega}\right) \cdot \Phi\left(\eta\left(\frac{x-\xi}{\omega}\right)\right)$$

Using the previous stochastic representation, we have:

$$\begin{aligned}
\Pr\{X \leq x\} &= 2\left(\Phi\left(\frac{x-\xi}{\omega}\right) - \Phi_2\left(0, \frac{x-\xi}{\omega}; \delta\right)\right) \\
&= 2\Phi_2\left(0, \frac{x-\xi}{\omega}; -\delta\right)
\end{aligned}$$

where:

$$\delta = \frac{\eta}{\sqrt{1+\eta^2}}$$

We note $m_0 = \delta\sqrt{2/\pi}$. The moments of the univariate SN distribution are:

$$\begin{aligned}
\mu(X) &= \xi + \omega m_0 \\
\sigma^2(X) &= \omega^2\left(1 - m_0^2\right) \\
\gamma_1(X) &= \left(\frac{4-\pi}{2}\right)\frac{m_0^3}{\left(1-m_0^2\right)^{3/2}} \\
\gamma_2(X) &= 2(\pi-3)\frac{m_0^4}{\left(1-m_0^2\right)^2}
\end{aligned}$$

A.2.1.17 The skew t distribution

The multivariate case Let $X \sim \mathcal{SN}(\mathbf{0}, \Omega, \eta)$ and $V \sim \chi^2_\nu / \nu$ be independent of X. Following Azzalini and Capitanio (2003), the mixture transformation $Y = \xi + V^{-1/2} X$ has a skew t distribution and we write $Y \sim \mathcal{ST}(\xi, \Omega, \eta, \nu)$. The density function of Y is related to the multivariate t distribution as follows:

$$f(y) = 2t_n(y - \xi; \Omega, \nu)\, \mathbf{T}_1\left(\eta^\top \omega^{-1}(y - \xi)\sqrt{\frac{\nu + n}{Q + \nu}}; \nu + n\right)$$

where $Q = (y - \xi)^\top \Omega^{-1}(y - \xi)$. We notice that we have:

$$\begin{aligned}
\Pr\{Y \le y\} &= \Pr\left\{V^{-1/2} X \le \omega^{-1}(y - \xi)\right\} \\
&= \Pr\left\{V^{-1/2} U \le \omega^{-1}(y - \xi) \mid U_0 > 0\right\} \\
&= 2\Pr\left\{V^{-1/2}\begin{pmatrix} -U_0 \\ U \end{pmatrix} \le \begin{pmatrix} 0 \\ \omega^{-1}(y - \xi) \end{pmatrix}\right\} \\
&= 2\left(\mathbf{T}_n\left(\omega^{-1}(y - \xi); \mathbb{C}, \nu\right) - \mathbf{T}_{n+1}\left(u_+; \mathbb{C}_+(\delta), \nu\right)\right) \\
&= 2\mathbf{T}_{n+1}\left(u_+; \mathbb{C}_+(-\delta), \nu\right)
\end{aligned}$$

where $u_+ = \left(0, \omega^{-1}(y - \xi)\right)$.

Like the multivariate skew normal distribution, the skew t distribution satisfies the closure property under linear transformation. Let A be a $m \times n$ matrix and $Y \sim \mathcal{ST}(\xi, \Omega, \eta)$. We have:

$$AY \sim \mathcal{SN}(\xi_A, \Omega_A, \eta_A, \nu_A)$$

where:

$$\begin{aligned}
\xi_A &= A\xi \\
\Omega_A &= A\Omega A^\top \\
\eta_A &= \frac{\omega_A \Omega_A^{-1} B^\top \eta}{\left(1 + \eta^\top\left(\mathbb{C} - B\Omega_A^{-1} B^\top\right)\eta\right)^{1/2}} \\
\nu_A &= \nu
\end{aligned}$$

with $\omega = \mathrm{diag}^{1/2}(\Omega)$, $\mathbb{C} = \omega^{-1}\Omega\omega$, $\omega_A = \mathrm{diag}^{1/2}(\Omega_A)$ and $B = \omega^{-1}\Omega A^\top$. This property also implies that the marginal distributions of a subset of Y is still a skew t distribution.

The univariate case The density function becomes:

$$f(y) = \frac{2}{\omega} \cdot t_1\left(\frac{y - \xi}{\omega}; \nu\right) \cdot \mathbf{T}_1\left(\eta\left(\frac{y - \xi}{\omega}\right)\sqrt{\frac{\nu + 1}{Q + \nu}}; \nu + 1\right)$$

where $Q = (y - \xi)^2 / \omega^2$. To compute the cumulative density function, we use the following result:

$$\Pr\{Y \le y\} = 2\mathbf{T}_2\left(0, \frac{y - \xi}{\omega}; -\delta; \nu\right)$$

Let m_0 and v_0 be two scalars defined as follows[16]:

$$\begin{aligned}
m_0 &= \delta\sqrt{\frac{\nu}{\pi}} \exp\left(\ln\Gamma\left(\frac{\nu - 1}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right)\right) \\
v_0 &= \frac{\nu}{\nu - 2} - \mu_0^2
\end{aligned}$$

[16] We recall that $\delta = \alpha / \sqrt{1 + \alpha^2}$.

As shown by Azzalini and Capitanio (2003), the moments of the univariate ST distribution are:

$$\begin{aligned}
\mu(Y) &= \xi + \omega m_0 \\
\sigma^2(Y) &= \omega^2 v_0 \\
\gamma_1(Y) &= m_0 v_0^{-3/2}\left(\frac{\nu\left(3-\delta^2\right)}{\nu-3} - \frac{3\nu}{\nu-2} + 2m_0^2\right) \\
\gamma_2(Y) &= m_0 v_0^{-2}\left(\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4m_0^2\nu\left(3-\delta^2\right)}{\nu-3} + \frac{6m_0^2\nu}{\nu-2} - 3m_0^4\right) - 3
\end{aligned}$$

A.2.1.18 The Wishart distribution

Let $X_1, \ldots, X_n$ be n independent Gaussian random vectors $\mathcal{N}_p(0, \Sigma)$. If we note $X = (X_1, \ldots, X_n)$ the $n \times p$ matrix, then $S = X^\top X$ is positive definite and follows a Wishart distribution $\mathcal{W}_p(\Sigma, n)$ with n degrees of freedom and covariance matrix Σ. Its probability density function is:

$$f(S) = \frac{|S|^{(n-p-1)/2}}{2^{np/2}\Gamma_p\left(n/2\right)|\Sigma|^{n/2}} \exp\left(-\frac{1}{2}\operatorname{trace}\left(\Sigma^{-1}S\right)\right)$$

where $\Gamma_p(\alpha)$ is the multivariate gamma function:

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma_p\left(\alpha + \frac{1-j}{2}\right)$$

We have $\mathbb{E}[S] = n\Sigma$ and $\operatorname{var}(S_{i,j}) = n\left(\Sigma_{i,j}^2 + \Sigma_{i,i}\Sigma_{j,j}\right)$. Here are the main properties:

1. if A is a $q \times p$ matrix with rank q, then $ASA^\top \sim \mathcal{W}_q\left(A\Sigma A^\top, n\right)$;
2. if $\Sigma > 0$, then $\Sigma^{-1/2}S\Sigma^{-1/2} \sim \mathcal{W}_p(I_p, n)$;
3. if S_i are independent random matrices $\mathcal{W}_p(\Sigma, n_i)$, then $\sum_{i=1}^m S_i \sim \mathcal{W}_p\left(\Sigma, \sum_{i=1}^m n_i\right)$;
4. if a is a $p \times 1$ vector, we have:
$$\frac{a^\top S a}{a^\top \Sigma a} \sim \chi^2(n)$$
5. S^{-1} follows an inverse Wishart distribution $\mathcal{W}_p^{-1}(\Sigma, n)$ and we have:
$$\frac{a^\top \Sigma^{-1} a}{a^\top S^{-1} a} \sim \chi^2(n-p+1)$$

A.2.2 Special results

A.2.2.1 Affine transformation of random vectors

The univariate case Let X be a random variable with probability distribution $\mathbf{F}$. We consider the affine transformation $Y = a + bX$. If $b > 0$, the cumulative distribution function $\mathbf{H}$ of Y is:

$$\begin{aligned}
\mathbf{H}(y) &= \Pr\{Y \le y\} \\
&= \Pr\left\{X \le \frac{y-a}{b}\right\} \\
&= \mathbf{F}\left(\frac{y-a}{b}\right)
\end{aligned}$$

and its density function is:

$$h(y) = \partial_y \mathbf{H}(y) = \frac{1}{b} f\left(\frac{y-a}{b}\right)$$

If $b < 0$, we obtain:

$$\begin{aligned}\mathbf{H}(y) &= \Pr\{Y \leq y\} \\ &= \Pr\left\{X \geq \frac{y-a}{b}\right\} \\ &= 1 - \mathbf{F}\left(\frac{y-a}{b}\right)\end{aligned}$$

and:

$$h(y) = \partial_y \mathbf{H}(y) = -\frac{1}{b} f\left(\frac{y-a}{b}\right)$$

The mean and the variance of Y are respectively equal to $a + b \cdot \mu(X)$ and $b^2 \cdot \text{var}(X)$. The centered moments are:

$$\mathbb{E}\left[(Y - \mu(Y))^r\right] = b^r \cdot \mathbb{E}\left[(X - \mu(X))^r\right]$$

We deduce that the excess kurtosis of Y is the same as for X whereas the skewness is equal to:

$$\gamma_1(Y) = \text{sign}(b) \cdot \gamma_1(X)$$

As an illustration, we consider the random variable $Y = \mu + \sigma X$ with $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\sigma > 0$. We obtain:

$$\mathbf{H}(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

and

$$h(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2$$

We also deduce that

$$\mathbf{H}^{-1}(\alpha) = \mu + \sigma\Phi^{-1}(\alpha)$$

For the moments, we obtain $\mu(Y) = \mu$, $\sigma^2(Y) = \sigma^2$, $\gamma_1(Y) = 0$ and $\gamma_2(Y) = 0$.

The multivariate case Let X be a random vector of dimension n, A a $m \times 1$ vector and B a $m \times n$ matrix. We consider the affine transformation $Y = A + BX$. The moments verify $\mu(Y) = A + B\mu(X)$ and $\text{cov}(Y) = B\,\text{cov}(X)\,B^\top$. In the general case, it is not possible to find the distribution of Y. However, if $X \sim \mathcal{N}(\mu, \Sigma)$, Y is also a normal random vector with $Y \sim \mathcal{N}\left(A + B\mu, B\Sigma B^\top\right)$.

A.2.2.2 Change of variables

Let X be a random variable, whose probability density function is $f(x)$. We consider the change of variable $Y = \varphi(X)$. If the function φ is monotonic, the probability density function $g(y)$ of Y is equal to:

$$g(y) = f(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$$

In the multivariate case, we note $(X_1, \ldots, X_n)$ the random vector with density function $f(x_1, \ldots, x_n)$. If the function φ is bijective, we can show that the probability density function of $(Y_1, \ldots, Y_n) = \varphi(X_1, \ldots, X_n)$ is equal to:

$$g(y_1, \ldots, y_n) = f(x_1, \ldots, x_n) \left| \frac{1}{\det J_\varphi} \right|$$

where J_φ is the Jacobian associated to the change of variables.

A.2.2.3 Relationship between density and quantile functions

Let $\mathbf{F}(x)$ be a cumulative distribution function. The density function is $f(x) = \partial_x \mathbf{F}(x)$. We note $\alpha = \mathbf{F}(x)$ and $x = \mathbf{F}^{-1}(\alpha)$. We have:

$$\frac{\partial \mathbf{F}^{-1}(\mathbf{F}(x))}{\partial x} = \frac{\partial \mathbf{F}^{-1}(\alpha)}{\partial \alpha} \left(\frac{\partial \mathbf{F}(x)}{\partial x} \right) = 1$$

We deduce that:

$$\frac{\partial \mathbf{F}^{-1}(\alpha)}{\partial \alpha} = \left(\frac{\partial \mathbf{F}(x)}{\partial x} \right)^{-1} = \frac{1}{f(\mathbf{F}^{-1}(\alpha))}$$

and:

$$f(x) = \frac{1}{\partial_\alpha \mathbf{F}^{-1}(\mathbf{F}(x))}$$

For instance, we can use this result to compute the moments of the random variable X with the quantile function instead of the density function:

$$\mathbb{E}[X^r] = \int_{-\infty}^{\infty} x^r f(x)\, \mathrm{d}x = \int_0^1 \left(\mathbf{F}^{-1}(\alpha)\right)^r \mathrm{d}\alpha$$

A.2.2.4 Conditional expectation in the case of the normal distribution

Let us consider a Gaussian random vector defined as follows:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{x,x} & \Sigma_{x,y} \\ \Sigma_{y,x} & \Sigma_{y,y} \end{pmatrix} \right)$$

The conditional probability distribution of Y given $X = x$ is a multivariate normal distribution. We have:

$$\begin{aligned} \mu_{y|x} &= \mathbb{E}[Y \mid X = x] \\ &= \mu_y + \Sigma_{y,x}\Sigma_{x,x}^{-1}(x - \mu_x) \end{aligned}$$

and:

$$\begin{aligned} \Sigma_{y,y|x} &= \sigma^2[Y \mid X = x] \\ &= \Sigma_{y,y} - \Sigma_{y,x}\Sigma_{x,x}^{-1}\Sigma_{x,y} \end{aligned}$$

We deduce that:

$$Y = \mu_y + \Sigma_{y,x}\Sigma_{x,x}^{-1}(x - \mu_x) + u$$

where u is a centered Gaussian random variable with variance $\sigma^2 = \Sigma_{y,y|x}$. It follows that:

$$Y = \underbrace{\left(\mu_y - \Sigma_{y,x}\Sigma_{x,x}^{-1}\mu_x\right)}_{\beta_0} + \underbrace{\Sigma_{y,x}\Sigma_{x,x}^{-1}}_{\beta^\top} x + u$$

We recognize the linear regression of Y on a constant and a set of exogenous variables X:

$$Y = \beta_0 + \beta^\top X + u$$

Moreover, we have:

$$\begin{aligned} R^2 &= 1 - \frac{\sigma^2}{\Sigma_{y,y}} \\ &= \frac{\Sigma_{y,x}\Sigma_{x,x}^{-1}\Sigma_{x,y}}{\Sigma_{y,y}} \end{aligned}$$

A.2.2.5 Calculation of a useful integral function in credit risk models

We consider the following integral:

$$I = \int_{-\infty}^{c} \Phi(a + bx)\,\phi(x)\,\mathrm{d}x$$

We have:

$$\begin{aligned} I &= \int_{-\infty}^{c} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a+bx} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \right) \phi(x)\,\mathrm{d}x \\ &= \frac{1}{2\pi} \int_{-\infty}^{c} \int_{-\infty}^{a+bx} \exp\left(-\frac{y^2+x^2}{2}\right) \mathrm{d}y\,\mathrm{d}x \end{aligned}$$

By considering the change of variables $(x, z) = \varphi(x, y)$ such that $z = y - bx$, we obtain[17]:

$$I = \frac{1}{2\pi} \int_{-\infty}^{c} \int_{-\infty}^{a} \exp\left(-\frac{z^2 + 2bzx + b^2x^2 + x^2}{2}\right) \mathrm{d}z\,\mathrm{d}x$$

If we consider the new change of variable $t = \left(1+b^2\right)^{-1/2} z$ and use the notation $\delta = 1 + b^2$, we have:

$$\begin{aligned} I &= \frac{\sqrt{\delta}}{2\pi} \int_{-\infty}^{c} \int_{-\infty}^{\frac{a}{\sqrt{1+b^2}}} \exp\left(-\frac{\delta t^2 + 2b\sqrt{\delta}tx + \delta x^2}{2}\right) \mathrm{d}t\,\mathrm{d}x \\ &= \frac{\sqrt{\delta}}{2\pi} \int_{-\infty}^{c} \int_{-\infty}^{\frac{a}{\sqrt{1+b^2}}} \exp\left(-\frac{\delta}{2}\left(t^2 + \frac{2b}{\sqrt{\delta}}xt + x^2\right)\right) \mathrm{d}t\,\mathrm{d}x \end{aligned}$$

We recognize the expression of the cumulative bivariate normal distribution[18], whose correlation parameter ρ is equal to $-b/\sqrt{\delta}$:

$$\int_{-\infty}^{c} \Phi(a + bx)\,\phi(x)\,\mathrm{d}x = \Phi_2\left(c, \frac{a}{\sqrt{1+b^2}}; \frac{-b}{\sqrt{1+b^2}}\right)$$

[17] We use the fact that the Jacobian of $\varphi(x, y)$ has the following expression:

$$J_\varphi = \begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix}$$

and its determinant $|J_\varphi|$ is equal to 1.

[18] We recall that $\Phi_2(x, y; \rho)$ is the cumulative distribution function of the bivariate Gaussian vector (X, Y) with correlation ρ on the space $[-\infty, x] \times [-\infty, y]$.

A.3 Stochastic analysis

In what follows, we recall the main results of stochastic analysis related to Brownian motions and stochastic differential equations. Most of them can be found in Gikhman and Skorokhod (1972), Liptser and Shiryaev (1974), Friedman (1975), Karatzas and Shreve (1991) and Øksendal (2010). But before that, we introduce some definitions and notations.

- The probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the sample space, $\mathcal{F}$ is the σ-algebra representing the collection of all events and $\mathbb{P}$ is the probability measure.
- A random (or stochastic) process $X = \{X(t) : t \in T\}$ is a collection of random variables $X(t)$ where $T = [0, \infty)$ is the index set.
- A filtration $\mathcal{F}_t$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing sequence of σ-algebras included in $\mathcal{F}$:
$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F} \qquad \forall\, t \geq s$$
The filtration represents the time evolution of the information produced by the stochastic process X.
- The random process X is $\mathcal{F}_t$-adapted if $X(t)$ is $\mathcal{F}_t$-measurable for all fixed $t \in T$, meaning that the value of $X(t)$ depends only on $\mathcal{F}_t$. In other words, the value of $X(t)$ cannot depend on unknown future data.
- The random process X is a martingale with respect to the filtration $\mathcal{F}_t$ if $\mathbb{E}|X(t)| < \infty$ (or $\mathbb{E}\left|X^2(t)\right| < \infty$) and $\mathbb{E}[X(t)|\mathcal{F}_s] = x_s$ where x_s is the realization of $X(s)$.
- The random process X is stationary if:
$$\mathbb{P}\{X(s) \in A\} = \mathbb{P}\{X(t) \in A\} \qquad \forall\, (s,t) \in T^2$$
It is weak-sense stationary if the first moment and the autocovariance do not vary with respect to time:
$$\begin{cases} \mathbb{E}\left[X(t)^2\right] < \infty \\ \mathbb{E}[X(s)] = \mathbb{E}[X(t)] \\ \mathbb{E}[X(s)\,X(t)] = \mathbb{E}[X(s+u)\,X(t+u)] \end{cases}$$
where $u \geq 0$. In the case where X is a Gaussian random process, the two definitions are equivalent.
- The stochastic process X is Markov if the probability distribution of $X(t)$ conditionally to the filtration $\mathcal{F}_s$ is equal to the probability distribution of $X(t)$ conditionally to the realization x_s:
$$\mathbb{P}\{X(t) \in A|\,\mathcal{F}_s\} = \mathbb{P}\{X(t) \in A|\,X(s) = x_s\}$$
This implies that we don't need to know all the information before s, but only the last value taken by the process.
- The random process X is continuous at time t if, for all $\varepsilon > 0$,
$$\lim_{t \to s} \mathbb{P}\{|X(t) - X(s)| > \varepsilon\} = 0$$
X is said to be a continuous stochastic process on T if X is continuous for any fixed $t \in T$.

A.3.1 Brownian motion and Wiener process

The stochastic process $B = \{B(t) : t \in T\}$ is a Brownian motion if:

1. $B(0) = 0$;
2. for all partition $0 = t_0 < t_1 < t_2 < \cdots < t_n$, the random variables $B(t_i) - B(t_{i-1})$ are independent;
3. $B(t) - B(s)$ is normally distributed with $\mathbb{E}[B(t) - B(s)] = \mu(t-s)$ and $\mathbb{E}\left[(B(t) - B(s))^2\right] = \sigma^2(t-s)$ for $t \geq s$.

If $\mu = 0$ and $\sigma = 1$, we obtain the standard Brownian motion. It corresponds to the Wiener process[19] and we denote it by $W(t)$. Here, we list the four main properties of $W(t)$:

1. $\mathbb{E}[W(t)] = 0$;
2. $\text{cov}(W(s) W(t)) = \mathbb{E}[W(s) W(t)] = \min(s,t)$;
3. $W(t)$ is a martingale;
4. the process W is continuous.

Notice that Wiener paths are not differentiable (Friedman, 1975), meaning that $\partial_t W(t)$ has no sense. We can also show that the Wiener process is invariant in law under various transformation: $c^{-1/2} W(ct) \overset{\mathcal{L}}{=} W(t)$ if $c > 0$ (rescaling), $tW\left(t^{-1}\right) \overset{\mathcal{L}}{=} W(t)$ (inversion) and $W(1) - W(1-t) \overset{\mathcal{L}}{=} W(t)$ if $t \in [0,1]$ (time reversibility).

The multidimensional Wiener process $W(t) = (W_1(t), \ldots, W_n(t))$ satisfies the following properties:

- each component $W_i(t)$ is a Brownian motion;
- the different components are correlated:

$$\mathbb{E}[W_i(s) W_j(t)] = \rho_{i,j} \cdot \min(s,t)$$

We note ρ the correlation matrix of $W(t)$:

$$\mathbb{E}\left[W(t) W(t)^\top\right] = \rho\, t$$

It implies that the density function of $W(t)$ is the multivariate normal pdf:

$$\phi_n(x, \rho\, t) = \frac{(2\pi)^{-n/2} |\rho|^{-1/2}}{\sqrt{t}} \exp\left(-\frac{1}{2t} x^\top \rho^{-1} x\right)$$

Remark 214 *Let $W(t)$ be a multivariate Wiener process with correlation matrix ρ. We have:*

$$W(t) = A W^\star(t)$$

where $AA^\top = \rho$ and $W^\star(t)$ is a multivariate independent Wiener process.

[19] From a historical point of view, the Brownian motion and the Wiener process were derived in a different manner. Today, the two terms are equivalent.

A.3.2 Stochastic integral

Let $f(t)$ be a stochastic function that defines the stochastic process[20] X. We assume that $f(t)$ is a random step function on $[a,b]$ and we denote by $\Delta = \{a = t_0 < t_1 < \cdots < t_n = b\}$ the associated partition. We have $f(t) = f(t_i)$ if $t \in [t_i, t_{i+1}[$. We introduce the notation:

$$\int_a^b f(t)\,\mathrm{d}W(t) = \sum_{i=0}^{n-1} f(t_i)\left(W(t_{i+1}) - W(t_i)\right)$$

$\int_a^b f(t)\,\mathrm{d}W(t)$ is called the stochastic integral of the random process $X = \{f(t), t \geq 0\}$ on $[a,b]$ with respect to the Wiener process. If $f(t)$ is non-random, we have an integration by parts (IPP) formula:

$$\int_a^b f(t)\,\mathrm{d}W(t) = f(b)\,W(b) - f(a)\,W(a) - \int_a^b f'(t)\,W(t)\,\mathrm{d}t$$

We deduce that:

$$\int_a^b \mathrm{d}W(t) = W(b) - W(a)$$

In the case of a general function $f(t)$, the stochastic integral is defined as the limit in probability of the Riemann sum[21]:

$$\int_a^b f(t)\,\mathrm{d}W(t) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i)\left(W(t_{i+1}) - W(t_i)\right)$$

Like the Riemann–Stieltjes integral, it satisfies the linearity property:

$$\int_a^b (\alpha f + \beta g)(t)\,\mathrm{d}W(t) = \alpha \int_a^b f(t)\,\mathrm{d}W(t) + \beta \int_a^b g(t)\,\mathrm{d}W(t)$$

and the Chasles decomposition:

$$\int_a^c f(t)\,\mathrm{d}W(t) = \int_a^b f(t)\,\mathrm{d}W(t) + \int_b^c f(t)\,\mathrm{d}W(t)$$

where $(\alpha,\beta) \in \mathbb{R}^2$ and $a < b < c$. We also have:

$$\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W(t)\right] = 0$$

Another important result is the Itô isometry that is useful for computing the variance of the stochastic integral:

$$\mathbb{E}\left[\left(\int_a^b f(t)\,\mathrm{d}W(t)\right)^2\right] = \mathbb{E}\left[\int_a^b f^2(t)\,\mathrm{d}t\right]$$

More generally, we have:

$$\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W(t)\int_a^b g(t)\,\mathrm{d}W(t)\right] = \mathbb{E}\left[\int_a^b f(t)\,g(t)\,\mathrm{d}t\right]$$

[20] This is equivalent that $X = \{X(t) = f(t), t \geq 0\}$ is a stochastic process.
[21] This construction is valid only if the random process X is $\mathcal{F}_t$-adapted.

Remark 215 *If $f(t)$ is a non-random function, then $\int_a^b f(t)\,\mathrm{d}W(t)$ is a Gaussian random variable, whose mean is zero and variance is equal to $\int_a^b f^2(t)\,\mathrm{d}t$.*

The Itô integral is a special case of the stochastic integral:

$$I(t) = \int_0^t f(s)\,\mathrm{d}W(s)$$

It is also called the '*indefinite integral*'. An important property is that any Itô integral is a martingale:

$$\mathbb{E}\left[I(t)|\mathcal{F}_s\right] = I(s)$$

A related result is the martingale representation theorem. Assuming that the filtration $\mathcal{F}_t$ is generated by a Wiener process, the theorem states that any $\mathcal{F}_t$-martingale $M(t)$ can be written as an Itô integral:

$$M(t) = \mathbb{E}\left[M(0)\right] + \int_0^t f(s)\,\mathrm{d}W(s)$$

A.3.3 Stochastic differential equation and Itô's lemma

An Itô process is an adapted stochastic process that can be expressed as follows:

$$X(t) = X(0) + \int_0^t \mu(s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W(s)$$

The stochastic differential equation (SDE) of $X(t)$ is:

$$\mathrm{d}X(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t)$$

The conditional process $\mathrm{d}X(t)$ with respect to $\mathcal{F}_t$ is Gaussian with mean $\mu(t)\,\mathrm{d}t$ and variance $\sigma^2(t)\,\mathrm{d}t$. $\mu(t)$ is called the drift coefficient, while $\sigma(t)$ is the diffusion coefficient.

A.3.3.1 Existence and uniqueness of a stochastic differential equation

Let $\mu(t,x)$ and $\sigma(t,x)$ be two measurable functions where $(t,x) \in T \times \mathbb{R}$. If $X(t)$ is a random process such that:

$$\begin{cases} \mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t) & \text{(A.21a)} \\ X(0) = x_0 & \text{(A.21b)} \end{cases}$$

we say that $X(t)$ satisfies the stochastic differential equation (A.21a) with the initial condition (A.21b).

Friedman (1975) showed that the system (A.21) has a unique solution if there exist two scalar K_1 and K_2 such that $\forall (x,y) \in \mathbb{R}^2$, we verify the following inequalities:

$$\begin{cases} |\mu(t,x) - \mu(t,y)| \leq K_1 |x-y| \\ |\sigma(t,x) - \sigma(t,y)| \leq K_1 |x-y| \end{cases}$$

and

$$\begin{cases} |\mu(t,x)| \leq K_2 (1+|x|) \\ |\sigma(t,x)| \leq K_2 (1+|x|) \end{cases}$$

The previous theorem is not the unique way to show the existence of a solution. For instance, a variant of this theorem is given by Karatzas and Shreve (1991). If there exist a scalar

$K \in \mathbb{R}$ that verifies the inequality $|\mu(t,x) - \mu(t,y)| \leq K|x-y|$ for all $(x,y) \in \mathbb{R}^2$, and a strictly increasing function $h : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ that satisfies the conditions[22] $h(0) = 0$ and $\int_0^\varepsilon h^{-2}(u)\,\mathrm{d}u = \infty$ for all $\varepsilon > 0$ such that $|\sigma(t,x) - \sigma(t,y)| = h(|x-y|)$, then the solution of the SDE exists and is unique.

A.3.3.2 Relationship with diffusion processes

A Markov process $X(t)$ is called a diffusion process if the transition probability function[23] $p(s,x;t,A)$ satisfies the two following properties:

1. For all $\varepsilon > 0$, $t \in T$ and $x \in \mathbb{R}$, we have:
$$\lim_{h \to 0} \frac{1}{h} \int_\Delta p(t,x;t+h,\mathrm{d}y) = 0$$
where $\Delta = \{y \in \mathbb{R} : |x-y| > \varepsilon\}$.

2. For all $\varepsilon > 0$, $t \in [0,T]$ and $x \in \mathbb{R}$, there exist two functions $a(t,x)$ and $b(t,x)$ such that:
$$\lim_{h \to 0} \frac{1}{h} \int_{\bar{\Delta}} (y-x)\, p(t,x;t+h,\mathrm{d}y) = a(t,x)$$
and:
$$\lim_{h \to 0} \frac{1}{h} \int_{\bar{\Delta}} (y-x)^2\, p(t,x;t+h,\mathrm{d}y) = b(t,x)$$
where $\bar{\Delta} = \{y \in \mathbb{R} : |x-y| \leq \varepsilon\}$.

This definition is given by Gikhman and Skorokhod (1972). They also showed that the unique solution of the SDE (A.21) is a diffusion process with $a(t,x) = \mu(t,x)$ and $b(t,x) = \sigma^2(t,x)$.

A.3.3.3 Itô calculus

To find the explicit solution of a SDE, we can use Itô calculus. We consider the following differential:
$$\mathrm{d}X(t) = \mu(t,X(t))\,\mathrm{d}t + \sigma(t,X(t))\,\mathrm{d}W(t)$$
Let $f(t,x)$ be a $\mathcal{C}^2$ function. The stochastic differential equation of $Y(t) = f(t,X(t))$ is equal to:
$$\begin{aligned}
\mathrm{d}Y(t) &= \mathrm{d}f(t,X(t)) \\
&= \frac{\partial f}{\partial t}(t,X(t))\,\mathrm{d}t + \frac{\partial f}{\partial x}(t,X(t))\,\mathrm{d}X(t) + \\
&\quad \frac{1}{2}\sigma^2(t,X(t))\frac{\partial^2 f}{\partial x^2}(t,X(t))\,\mathrm{d}t \\
&= \left(\frac{\partial f}{\partial t}(t,X(t)) + \mu(t,X(t))\frac{\partial f}{\partial x}(t,X(t)) + \right. \\
&\quad \left. \frac{1}{2}\sigma^2(t,X(t))\frac{\partial^2 f}{\partial x^2}(t,X(t))\right)\mathrm{d}t + \\
&\quad \sigma(t,X(t))\frac{\partial f}{\partial x}(t,X(t))\,\mathrm{d}W(t)
\end{aligned}$$

[22] For instance, we can take $h(u) = u^\alpha$ where $\alpha \geq \frac{1}{2}$.

[23] The transition probability function of the Markov process $X(t)$ is defined as:
$$p(s,x;t,A) = P(X(t) \in A \mid X(s) = x)$$

The previous result is called the Itô formula. It can be viewed as Taylor series with the following Itô rules: $\mathrm{d}t \cdot \mathrm{d}t = 0$, $\mathrm{d}t \cdot \mathrm{d}W(t) = 0$ and $\mathrm{d}W(t) \cdot \mathrm{d}W(t) = \mathrm{d}t$.

Remark 216 *In compact form, we have:*

$$\begin{aligned} \mathrm{d}Y &= \left(\frac{\partial f}{\partial t} + \mu(\cdot) \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(\cdot) \frac{\partial^2 f}{\partial x^2} \right) \mathrm{d}t + \frac{\partial f}{\partial x}\sigma(\cdot)\, \mathrm{d}W \\ &= \left(\frac{\partial f}{\partial t} + \mathcal{A}_t f \right) \mathrm{d}t + \frac{\partial f}{\partial x}\sigma(\cdot)\, \mathrm{d}W \end{aligned}$$

where $\mathcal{A}_t f$ is the infinitesimal generator of X:

$$\mathcal{A}_t f = \frac{1}{2}\sigma^2(\cdot) \frac{\partial^2 f}{\partial x^2} + \mu(\cdot) \frac{\partial f}{\partial x}$$

In the case where $X(t) = W(t)$, we obtain:

$$\mathrm{d}f(t, W(t)) = \partial_t f(t, W(t))\, \mathrm{d}t + \frac{1}{2}\partial_x^2 f(t, W(t))\, \mathrm{d}t + \partial_x f(t, W(t))\, \mathrm{d}W(t)$$

If we now consider two stochastic processes $X_1(t)$ and $X_2(t)$ that depend on the same Wiener process:

$$\begin{cases} \mathrm{d}X_1(t) = \mu_1(t, X_1(t))\, \mathrm{d}t + \sigma_1(t, X_1(t))\, \mathrm{d}W(t) \\ \mathrm{d}X_2(t) = \mu_2(t, X_2(t))\, \mathrm{d}t + \sigma_2(t, X_2(t))\, \mathrm{d}W(t) \end{cases}$$

the Itô formula becomes:

$$\mathrm{d}(X_1(t) X_2(t)) = X_1(t)\, \mathrm{d}X_2(t) + X_2(t)\, \mathrm{d}X_1(t) + \mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t)$$

where:

$$\mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t) = \sigma_1(t, X_1(t))\, \sigma_2(t, X_2(t))\, \mathrm{d}t$$

A.3.3.4 Extension to the multidimensional case

Let $X(t) = (X_1(t), \ldots, X_m(t))$ be a random vector process. We consider the functions $\mu : T \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$ and $\sigma : T \times \mathbb{R}^m \longrightarrow \mathbb{R}^{m \times n}$. The multidimensional SDE is defined as:

$$\begin{cases} \mathrm{d}X(t) = \mu(t, X(t))\, \mathrm{d}t + \sigma(t, X(t))\, \mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

where x_0 is a vector of dimension m. The Itô formula applied to $Y(t) = f(t, X(t))$ is:

$$\begin{aligned} \mathrm{d}Y(t) &= \mathrm{d}f(t, X(t)) \\ &= \frac{\partial f}{\partial t}(t, X(t))\, \mathrm{d}t + \frac{\partial f}{\partial x}(t, X(t))^\top\, \mathrm{d}X(t) + \\ &\quad \frac{1}{2}\mathrm{d}X(t)^\top \frac{\partial^2 f}{\partial x^2}(t, X(t))\, \mathrm{d}X(t) \end{aligned}$$

We finally obtain[24]:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{\partial f}{\partial t}(t, X(t)) + \frac{\partial f}{\partial x}(t, X(t))^\top \mu(t, X(t)) + \right. \\ &\quad \left. \frac{1}{2} \operatorname{trace}\left(\sigma(t, X(t))^\top \frac{\partial^2 f}{\partial x^2}(t, X(t))\, \sigma(t, X(t))\, \rho \right) \right) \mathrm{d}t + \\ &\quad \frac{\partial f}{\partial x}(t, X(t))^\top \sigma(t, X(t))\, \mathrm{d}W(t) \end{aligned}$$

[24] The Itô rules are $\mathrm{d}t \cdot \mathrm{d}t = 0$, $\mathrm{d}t \cdot \mathrm{d}W(t) = \mathbf{0}$ and $\mathrm{d}W(t) \cdot \mathrm{d}W(t)^\top = \rho\, \mathrm{d}t$.

If we apply the Itô formula to $Y(t) = X_1(t) X_2(t)$, we obtain:

$$\begin{aligned} \mathrm{d}Y(t) &= \mu_1(t, X_1(t)) X_2(t)\ \mathrm{d}t + \mu_2(t, X_2(t)) X_1(t)\ \mathrm{d}t + \\ &\quad \rho_{1,2}\sigma_1(t, X_1(t))\sigma_2(t, X_2(t)) + \\ &\quad \sigma_1(t, X_1(t)) X_2(t)\ \mathrm{d}W_1(t) + \sigma_2(t, X_2(t)) X_1(t)\ \mathrm{d}W_2(t) \\ &= X_1(t)\ \mathrm{d}X_2(t) + X_2(t)\ \mathrm{d}X_1(t) + \mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t) \end{aligned}$$

where:

$$\mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t) = \rho_{1,2}\sigma_1(t, X_1(t))\sigma_2(t, X_2(t))\ \mathrm{d}t$$

In the case where $W_1(t) = W_2(t) = W(t)$, the correlation $\rho_{1,2}$ is equal to one and we retrieve the previous result.

Using the previous framework, we also deduce that the integration by parts formula becomes:

$$\begin{aligned} \int_a^b X_1(t)\ \mathrm{d}X_2(t) &= X_1(b) X_2(b) - X_1(a) X_2(a) - \\ &\quad \int_a^b X_2(t)\ \mathrm{d}X_1(t) - \int_a^b \mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t) \end{aligned}$$

In the case where $X_1(t) = f(t)$ is a non-random process and $X_2(t) = W(t)$, we retrieve the classical IPP:

$$\begin{aligned} \int_a^b f(t)\,\mathrm{d}W(t) &= f(b) W(b) - f(a) W(a) - \\ &\quad \int_a^b W(t)\ \mathrm{d}f(t) - \int_a^b \mathrm{d}f(t) \cdot \mathrm{d}W(t) \\ &= f(b) W(b) - f(a) W(a) - \int_a^b f'(t) W(t)\ \mathrm{d}t \end{aligned}$$

because $\int_a^b f'(t)\ \mathrm{d}t\,\mathrm{d}W(t) = 0$.

Remark 217 $\mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t)$ *is also called the quadratic variation and we note:*

$$\mathrm{d}X_1(t) \cdot \mathrm{d}X_2(t) = \langle X_1(t), X_2(t) \rangle$$

Using the notation $\langle X_1(t) \rangle = \langle X_1(t), X_1(t) \rangle$*, the quadratic variation satisfies the bilinearity property and the polarization identity:*

$$\langle X_1(t), X_2(t) \rangle = \frac{\langle X_1(t) + X_2(t) \rangle - \langle X_1(t) \rangle - \langle X_2(t) \rangle}{2}$$

A.3.4 Feynman-Kac formula

We consider the state variable $X(t)$ defined by:

$$\mathrm{d}X(t) = \mu(t, X(t))\ \mathrm{d}t + \sigma(t, X(t))\ \mathrm{d}W(t)$$

and $\mathcal{A}_t V$ the infinitesimal generator of the diffusion process:

$$\mathcal{A}_t V(t, x) = \frac{1}{2}\sigma^2(t, x) \frac{\partial^2 V(t, x)}{\partial x^2} + \mu(t, x) \frac{\partial V(t, x)}{\partial x}$$

Under the following assumptions:

1. $\mu(t,x)$, $\sigma(t,x)$, $g(t,x)$ and $h(t,x)$ are Lipschitz and bounded on $[0,T]\times\mathbb{R}$;
2. $f(x)$ is a continuous function of class $\mathcal{C}^2$;
3. $f(x)$ and $g(x)$ grows exponentially[25].

the solution of the Cauchy problem:

$$\begin{cases} -\partial_t V(t,x) + h(t,x)\,V(t,x) = \mathcal{A}_t V(t,x) + g(t,x) \\ V(T,x) = f(x) \end{cases} \tag{A.22}$$

is unique and given by:

$$V(t,x) = \mathbb{E}\left[\beta(T) f(X(T)) + \int_t^T \beta(s)\, g(s, X(s))\, \mathrm{d}s \,\middle|\, X(t) = x\right] \tag{A.23}$$

where:

$$\beta(s) = \exp\left(-\int_t^s h(u, X(u))\, \mathrm{d}u\right)$$

The Feynman-Kac formula states that the solution of the parabolic PDE (A.22) can be found by calculating the conditional expectation (A.23).

In the case where $h(t,x) = g(t,x) = 0$, we obtain the backward Chapman-Kolmogorov equation:

$$\partial_t V(t,x) = -\mu(t,x)\frac{\partial V(t,x)}{\partial x} - \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 V(t,x)}{\partial x^2}$$

where $V(T,x) = f(x)$ and:

$$V(t,x) = \mathbb{E}\left[f(X(T)) | X(t) = x\right]$$

If $f(x) = \mathbb{1}\{x \leq x_T\}$, we obtain the probability distribution:

$$\begin{aligned} V(t,x) &= \mathbb{E}\left[\mathbb{1}\{X(T) \leq x_T\} | X(t) = x\right] \\ &= \mathbb{P}\left\{X(T) \leq x_T | X(t) = x\right\} \end{aligned}$$

To obtain the density function, we set $f(x) = \mathbb{1}\{x = x_T\}$ and we have:

$$V(t,x) = \mathbb{P}\left\{X(T) = x_T | X(t) = x\right\}$$

Remark 218 *The Feynman-Kac formula is valid in the multivariate case by considering the following infinitesimal generator of the diffusion process:*

$$\begin{aligned} \mathcal{A}_t V(t,x) &= \frac{1}{2}\operatorname{trace}\left(\sigma(t,X(t))^\top \frac{\partial^2 f}{\partial x^2}(t,X(t))\,\sigma(t,X(t))\,\rho\right) + \\ &\quad \frac{\partial f}{\partial x}(t,X(t))^\top \mu(t,X(t)) \end{aligned}$$

[25] This implies that there exist two scalars $K \geq 0$ and $\xi \geq 0$ such that $|f(x)| \leq Ke^{\xi x^2}$.

A.3.5 Girsanov theorem

Let W be a Wiener process on the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$. If the process $g(t)$ satisfies the Novikov condition:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s\right)\right] < \infty$$

then $W^{\mathbb{Q}}$ defined by $W^{\mathbb{Q}}(t) = W(t) - \int_0^t g(s)\,\mathrm{d}s$ is a Wiener process on the probability space $\{\Omega, \mathcal{F}, \mathbb{Q}\}$. The change of measure is given by the Radon-Nikodym derivative:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} &= M(t)\\ &= \exp\left(\int_0^t g(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t g^2(s)\,\mathrm{d}s\right)\end{aligned}$$

Moreover, $M(t)$ is an $\mathcal{F}_t$-martingale.

Remark 219 *If we consider the state variable $X(t)$ defined by:*

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t)$$

the Girsanov theorem states that the change of measure is equivalent to change the drift of the diffusion:

$$\mathrm{d}X(t) = (\mu(t, X(t)) + g(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W^{\mathbb{Q}}(t)$$

The Girsanov theorem can be extended to the multidimensional Wiener process. In this case, $g(t)$ and $W^{\mathbb{Q}}(t) = W(t) - \int_0^t g(s)\,\mathrm{d}s$ are two vector processes and we have:

$$M(t) = \exp\left(\int_0^t g(s)^\top\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t g(s)^\top g(s)\,\mathrm{d}s\right)$$

A.3.6 Fokker-Planck equation

With the backward Chapman-Kolmogorov equation, we can compute the probability of the event $\{X(T) = x_T\}$ conditionally to $X(t) = x$. From a numerical point of view, this approach is generally not efficient because we need to solve one PDE for each value of x_T. Another way to compute this probability is to consider the forward Chapman-Kolmogorov equation:

$$\begin{cases}\partial_t U(t,x) = -\partial_x\left[\mu(t,x)\,U(t,x)\right] + \frac{1}{2}\partial_x^2\left[\sigma^2(t,x)\,U(t,x)\right]\\ U(s,x) = \mathbf{1}\{x = x_s\}\end{cases}$$

where $s < t$. This PDE is known as the Fokker-Planck equation and its solution is:

$$\begin{aligned}U(t,x) &= \mathbb{P}\{X(t) = x \mid X(s) = x_s\}\\ &= p(s, x_s; t, x)\end{aligned}$$

In particular, we can calculate the density function $p(0, x_0; T, x_T)$.

In the multidimensional case, the Fokker-Planck equation becomes:

$$\begin{aligned}\partial_t U(t, x_1, \ldots, x_m) &= -\sum_{i=1}^m \partial_{x_i}\left[\mu_i(t, x_1, \ldots, x_m)\,U(t, x_1, \ldots, x_m)\right] +\\ &\quad \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \partial^2_{x_i,x_j}\left[\Sigma_{i,j}(t, x_1, \ldots, x_m)\,U(t, x_1, \ldots, x_m)\right]\end{aligned}$$

where:

$$\Sigma_{i,j}(t, x_1, \ldots, x_m) = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \rho_{k_1,k_2} \sigma_{i,k_1}(t, x_1, \ldots, x_m) \sigma_{j,k_2}(t, x_1, \ldots, x_m)$$

In the diagonal bivariate case:

$$\begin{cases} \mathrm{d}X_1(t) = \mu_1(t, X_1(t), X_2(t))\,\mathrm{d}t + \sigma_1(t, X_1(t), X_2(t))\,\mathrm{d}W_1(t) \\ \mathrm{d}X_2(t) = \mu_2(t, X_1(t), X_2(t))\,\mathrm{d}t + \sigma_2(t, X_1(t), X_2(t))\,\mathrm{d}W_2(t) \\ \mathbb{E}[W_1(t) W_2(t)] = \rho_{1,2}\,\mathrm{d}t \end{cases}$$

we obtain:

$$\begin{aligned} \partial_t U(t, x_1, x_2) = & -\partial_{x_1}[\mu_1(t, x_1, x_2) U(t, x_1, x_2)] - \\ & \partial_{x_2}[\mu_2(t, x_1, x_2) U(t, x_1, x_2)] + \\ & \frac{1}{2}\partial^2_{x_1}[\sigma_1^2(t, x_1, x_2) U(t, x_1, x_2)] + \\ & \frac{1}{2}\partial^2_{x_2}[\sigma_2^2(t, x_1, x_2) U(t, x_1, x_2)] + \\ & \rho_{1,2}\partial^2_{x_1,x_2}[\sigma_1(t, x_1, x_2)\sigma_2(t, x_1, x_2) U(t, x_1, x_2)] \end{aligned}$$

A.3.7 Reflection principle and stopping times

A nonnegative random variable τ is a stopping time with respect to the stochastic process W if the event $\{\tau \leq t\}$ depends on $W(s)$ for $s \leq t$. This implies that $\{\tau \leq t\} \in \mathcal{F}_t$, meaning that the event cannot depend on the future path of the stochastic process. A particular case of stopping times is a hitting time. Let $\tau_x = \inf\{t : W(t) = x\}$ denote the first time when the Brownian motion hits the value $x \geq 0$. We can show that the hitting time τ_x is also a stopping time, and satisfies the strong Markov property:

$$W(\tau_x + t) - W(\tau_x) = W(\tau_x + t) - x \overset{\mathcal{L}}{=} W(t)$$

Therefore, $W(\tau_x + t) - W(\tau_x)$ is a Brownian motion that is independent from $W(s)$ for $s \leq \tau_x$. This result generalizes the independent increments property when τ_x is not a fixed time, but a random time. Let us define $\tilde{W}(t)$ as follows:

$$\tilde{W}(t) = \begin{cases} W(t) & \text{if } t \leq \tau_x \\ 2x - W(t) & \text{otherwise} \end{cases}$$

The reflection principle states that $\tilde{W}(t)$ is also a Brownian motion. Then, we can show that[26]:

$$\begin{aligned} \Pr\{\tau_x \leq t\} &= 2\Pr\{W(t) \geq x\} \\ &= 2\left(1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right) \end{aligned}$$

[26] We deduce that the density function of τ_x is equal to:

$$\begin{aligned} f(t) &= \partial_t \Pr\{\tau_x \leq t\} \\ &= \frac{x}{t^{3/2}}\phi\left(\frac{x}{\sqrt{t}}\right) \end{aligned}$$

Let $M(t) = \sup_{s \leq t} W(s)$ be the maximum of a Brownian motion. We have:

$$\{M(t) \geq x\} \Leftrightarrow \{\tau_x \leq t\}$$

It follows that:

$$\begin{aligned} \Pr\{M(t) \geq x\} &= \Pr\{\tau_x \leq t\} \\ &= 2\left(1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right) \end{aligned}$$

Another important result is the joint distribution of $(M(t), W(t))$, which is given by:

$$\begin{aligned} \Pr\{M(t) \geq x, W(t) \leq y\} &= \Pr\{W(t) \geq 2x - y\} \\ &= 1 - \Phi\left(\frac{2x - y}{\sqrt{t}}\right) \end{aligned}$$

where $x \geq 0$ and $y \leq x$. It follows that the joint density of $(M(t), W(t))$ is equal to:

$$f(x, y) = \frac{(2x - y)}{t^{3/2}}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{(2x - y)^2}{2t}\right)$$

If we consider the Brownian motion with drift:

$$X(t) = \mu t + W(t)$$

the distribution of $M(t) = \sup_{s \leq t} X(s)$ becomes:

$$\Pr\{M(t) \leq x\} = \Phi\left(\frac{x - \mu t}{\sqrt{t}}\right) - e^{2\mu x}\Phi\left(\frac{-x - \mu t}{\sqrt{t}}\right) \tag{A.24}$$

whereas the joint density of $(M(t), X(t))$ is equal to:

$$f(x, y) = \frac{(2x - y)}{t^{3/2}}\sqrt{\frac{2}{\pi}}\exp\left(\mu y - \frac{1}{2}\mu^2 t - \frac{(2x - y)^2}{2t}\right)$$

A.3.8 Some diffusion processes

A.3.8.1 Geometric Brownian motion

It is the solution of the following SDE:

$$\begin{cases} \mathrm{d}X(t) = \mu X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

In order to find the explicit solution, we apply the Itô's lemma to the stochastic process $Y(t) = \ln X(t)$ and we have[27]:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(\frac{1}{X(t)}\mu X(t) - \frac{1}{2X(t)^2}(\sigma X(t))^2\right)\mathrm{d}t + \frac{1}{X(t)}\sigma X(t)\,\mathrm{d}W(t) \\ &= \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma\,\mathrm{d}W(t) \end{aligned}$$

[27] We have $f(t, x) = \ln x$, $\partial_t f(t, x) = 0$, $\partial_x f(t, x) = x^{-1}$ and $\partial_x^2 f(t, x) = -x^{-2}$.

where $Y(0) = \ln x_0$. We deduce that $Y(t)$ is a Gaussian random process[28]:

$$\begin{aligned} Y(t) &= \ln x_0 + \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}W(s) \\ &= \ln x_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t) \end{aligned}$$

It follows that:

$$\ln X(t) = \ln x_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)$$

or:

$$\begin{aligned} X(t) &= \exp\left(\ln x_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)\right) \\ &= x_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)} \end{aligned}$$

We obtain a log-normal random process, whose first moments are:

$$\begin{aligned} \mathbb{E}[X(t)] &= \exp\left(\ln x_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \frac{1}{2}\sigma^2 t\right) \\ &= x_0 e^{\mu t} \end{aligned}$$

and:

$$\begin{aligned} \mathrm{var}(X(t)) &= e^{2\ln x_0 + 2\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma^2 t}\left(e^{\sigma^2 t} - 1\right) \\ &= x_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right) \end{aligned}$$

A.3.8.2 Ornstein-Uhlenbeck process

We consider the SDE:

$$\begin{cases} \mathrm{d}X(t) = a(b - X(t)) \, \mathrm{d}t + \sigma \, \mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

where $a > 0$. We notice that $\mathbb{E}[\mathrm{d}X(t) \mid \mathcal{F}_t] = a(b - X(t)) \, \mathrm{d}t$. It follows that:

- if $X(t) < b$, then $\mathbb{E}[\mathrm{d}X(t) \mid \mathcal{F}_t] \geq 0$ implying that $\mathbb{E}[X(t) + \mathrm{d}X(t) \mid \mathcal{F}_t] \geq X(t)$;
- if $X(t) > b$, then $\mathbb{E}[\mathrm{d}X(t) \mid \mathcal{F}_t] \leq 0$ implying that $\mathbb{E}[X(t) + \mathrm{d}X(t) \mid \mathcal{F}_t] \leq X(t)$.

The coefficient b is the mean-reversion parameter (or the long-term mean) whereas a is the speed of reversion. If we apply the Itô's lemma to $Y(t) = (b - X(t)) e^{at}$, we obtain[29]:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(a(b - X(t)) e^{at} - a(b - X(t)) e^{at}\right) \mathrm{d}t - e^{at}\sigma \, \mathrm{d}W(t) \\ &= -\sigma e^{at} \, \mathrm{d}W(t) \end{aligned}$$

and:

$$Y(t) = b - x_0 - \sigma \int_0^t e^{as} \, \mathrm{d}W(s)$$

[28] We have:

$$Y(t) \sim \mathcal{N}\left(\ln x_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right)$$

[29] We have $f(t,x) = (b - x) e^{at}$, $\partial_t f(t,x) = a f(t,x)$, $\partial_x f(t,x) = -e^{at}$ and $\partial_x^2 f(t,x) = 0$.

Therefore, $Y(t)$ is a Gaussian random process, because $\int_0^t e^{as}\,\mathrm{d}W(s)$ is a Gaussian random variable. We deduce that:

$$\begin{aligned} X(t) &= b - e^{-at}Y(t) \\ &= e^{-at}x_0 + b\left(1 - e^{-at}\right) + \sigma\int_0^t e^{-a(t-s)}\,\mathrm{d}W(s) \end{aligned}$$

$X(t)$ is also a Gaussian random process, whose first two moments are:

$$\mathbb{E}[X(t)] = e^{-at}x_0 + b\left(1 - e^{-at}\right)$$

and:

$$\begin{aligned} \mathrm{var}(X(t)) &= \mathbb{E}\left[\sigma^2\int_0^t e^{-2a(t-s)}\,\mathrm{d}s\right] \\ &= \sigma^2\left[\frac{e^{-2a(t-s)}}{2a}\right]_0^t \\ &= \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right) \end{aligned}$$

A.3.8.3 Cox-Ingersoll-Ross process

The CIR process is the solution of the SDE:

$$\begin{cases} \mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$

where $a > 0$. It can be viewed as a modified Ornstein-Uhlenbeck process where the diffusion coefficient is $\sigma\sqrt{X(t)}$. This implies that the CIR process is positive, and explains that this process is frequently used for interest rate modeling. If we apply the Itô's lemma to $Y(t) = (b - X(t))\,e^{at}$, we obtain:

$$\begin{aligned} \mathrm{d}Y(t) &= \left(a(b - X(t))\,e^{at} - a(b - X(t))\,e^{at}\right)\mathrm{d}t - e^{at}\sigma\sqrt{X(t)}\,\mathrm{d}W(t) \\ &= -\sigma e^{at}\sqrt{X(t)}\,\mathrm{d}W(t) \end{aligned}$$

and:

$$\begin{aligned} X(t) &= b - e^{-at}Y(t) \\ &= e^{-at}x_0 + b\left(1 - e^{-at}\right) + \sigma\int_0^t e^{-a(t-s)}\sqrt{X(s)}\,\mathrm{d}W(s) \end{aligned} \tag{A.25}$$

We can show that:

$$X(t) = \frac{1}{c}\chi^2_\nu(\zeta)$$

where:

$$c = \frac{4a}{(1 - e^{-at})\,\sigma^2}$$

and $\chi^2_\nu(\zeta)$ is the noncentral chi-squared random variable where $\nu = 4ab\sigma^{-2}$ is the number of degrees of freedom and $\zeta = cx_0e^{-at}$ is the noncentrality parameter. It follows that the probability density function of $X(t)$ is equal to:

$$f(x) = cf\left(cx; \frac{4ab}{\sigma^2}, cx_0e^{-at}\right)$$

where $f(y;\nu,\zeta)$ is the probability density function of the noncentral chi-squared random variable $\chi^2_\nu(\zeta)$. Using Equation (A.25), we can also show that:

$$\mathbb{E}[X(t)] = e^{-at}x_0 + b\left(1 - e^{-at}\right)$$

and:

$$\begin{aligned}
\operatorname{var}(X(t)) &= \mathbb{E}\left[\sigma^2 \int_0^t e^{-2a(t-s)} X(s)\, \mathrm{d}s\right] \\
&= \sigma^2 e^{-2at} \int_0^t e^{2as} \mathbb{E}[X(s)]\, \mathrm{d}s \\
&= \sigma^2 e^{-2at} \int_0^t (x_0 - b) e^{as} + b e^{2as}\, \mathrm{d}s \\
&= \sigma^2 e^{-2at} \left[\frac{(x_0 - b)}{a} e^{as} + \frac{b}{2a} e^{2as}\right]_0^t \\
&= \sigma^2 e^{-2at} \left(\frac{(x_0 - b)}{a} e^{at} + \frac{b}{2a} e^{2at} - \frac{(x_0 - b)}{a} - \frac{b}{2a}\right) \\
&= \frac{\sigma^2 b}{2a}\left(1 - 2e^{-at} + e^{-2at}\right)
\end{aligned}$$

A.3.8.4 Multidimensional processes

The multidimensional geometric Brownian motion is defined as:

$$\begin{cases} \mathrm{d}X_j(t) = \mu_j X_j(t)\, \mathrm{d}t + \sigma_j X_j(t)\, \mathrm{d}W_j(t) & \text{for } j = 1, \ldots, n \\ X(0) = x_0 \end{cases}$$

where $X(t) = (X_1(t), \ldots, X_n(t))$ and $W(t) = (W_1(t), \ldots, W_n(t))$ is a n-dimensional Brownian motion with $\mathbb{E}\left[W(t) W(t)^\top\right] = \rho t$. The solution of the multidimensional SDE is a multivariate log-normal process with:

$$X_j(t) = X_j(0) \cdot \exp\left(\left(\mu_j - \frac{1}{2}\sigma_j^2\right) t + \sigma_j W_j(t)\right)$$

where $W(t) \sim \mathcal{N}_n(0, \rho t)$.

Other multivariate stochastic processes are not very useful in finance, except stochastic volatility models. For instance, the Heston model is defined as follows:

$$\begin{cases} \mathrm{d}X_1(t) = \mu X_1(t)\, \mathrm{d}t + \sqrt{X_2(t)} X_1(t)\, \mathrm{d}W_1(t) \\ \mathrm{d}X_2(t) = a(b - X_2(t))\, \mathrm{d}t + \sigma \sqrt{X_2(t)}\, \mathrm{d}W_2(t) \end{cases}$$

where $\mathbb{E}[W_1(t) W_2(t)] = \rho t$. Therefore, the process $X_1(t)$ is a geometric Brownian motion with a stochastic volatility $\sigma(t) = \sqrt{X_2(t)}$ and the stochastic variance $X_2(t)$ is a CIR process. Another related process is the SABR model:

$$\begin{cases} \mathrm{d}X_1(t) = X_2(t) X_1^\beta(t)\, \mathrm{d}W_1(t) \\ \mathrm{d}X_2(t) = \nu X_2(t)\, \mathrm{d}W_2(t) \end{cases}$$

where $\mathbb{E}[W_1(t) W_2(t)] = \rho t$ and $\beta \in [0, 1]$. We notice that the stochastic volatility is a geometric Brownian motion and there are two special cases: $X_1(t)$ is a log-normal process ($\beta = 1$) or a normal process ($\beta = 0$).

A.4 Exercises

A.4.1 Discrete-time random process

1. We consider the discrete-time random process X_t defined by:

$$X_t = X_{t-1} + \varepsilon_t$$

where $X_0 = 0$ and ε_t is an *iid* random process.

   (a) We assume that ε_t is a Bernoulli random variable $\mathcal{B}(p)$. Give the filtration $\mathcal{F}_t$ for $t = 0, 1, 2$.

   (b) We assume that $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$. Show that X_t is a martingale.

2. We consider the AR(1) process:

$$X_t = \phi X_{t-1} + \varepsilon_t$$

where $|\phi| < 1$ and ε_t is an *iid* random process with $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$.

   (a) Show that X_t is a weak-sense stationary process.

   (b) Show that X_t is a strong-sense stationary process.

   (c) Is X_t a Markov process?

   (d) Same questions with the MA(1) process:

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

A.4.2 Properties of Brownian motion

We consider the standard Brownian motion $W(t)$.

1. Demonstrate the three properties:

   (a) $\mathbb{E}[W(t)] = 0$;

   (b) $\text{cov}(W(s) W(t)) = \min(s, t)$;

   (c) $W(t)$ is a martingale.

2. Show that $W(t)$ is continuous.

3. Calculate $\mathbb{E}\left[W^2(t)\right]$, $\mathbb{E}\left[W^2(t) \middle| \mathcal{F}_s\right]$, $\mathbb{E}\left[W^3(t)\right]$, $\mathbb{E}\left[W^4(t)\right]$, $\mathbb{E}\left[e^{W(t)}\right]$ and $\mathbb{E}\left[e^{W(t)} \middle| \mathcal{F}_s\right]$.

4. Calculate the mathematical expectation of $W^n(t)$ for $n \in \mathbb{N}^\star$.

A.4.3 Stochastic integral for random step functions

We assume that $f(t)$ and $g(t)$ are two random step functions on $[a, b]$:

$$\begin{cases} f(t) = f(t_i) \\ g(t) = g(t_i) \end{cases} \qquad \text{if } t \in [t_i, t_{i+1}[$$

where $t_0 = a$ and $t_n = b$.

1. Demonstrate the linearity property and Chasles decomposition of the stochastic integral $\int_a^b f(t)\,\mathrm{d}W(t)$.
2. Show that:
$$\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W(t)\right] = 0$$
and:
$$\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}W(t)\int_a^b g(t)\,\mathrm{d}W(t)\right] = \mathbb{E}\left[\int_a^b f(t)\,g(t)\,\mathrm{d}t\right]$$
3. Deduce that:
$$\mathrm{var}\left(\int_a^b f(t)\,\mathrm{d}W(t)\right) = \int_a^b \mathbb{E}\left[f^2(t)\right]\,\mathrm{d}t$$

A.4.4 Power of Brownian motion

Let $W(t)$ be a standard Brownian motion.

1. Show that:
$$\mathrm{d}W^2(t) = \mathrm{d}t + 2W(t)\,\mathrm{d}W(t)$$
2. Deduce the solution of $I(t) = \int_0^t W(s)\,\mathrm{d}W(s)$ and calculate the first two moments of $I(t)$.
3. Let $n \in \mathbb{N}^\star$. Show that:
$$\mathrm{d}W^n(t) = \frac{1}{2}n(n-1)W^{n-2}(t)\,\mathrm{d}t + nW^{n-1}(t)\,\mathrm{d}W(t)$$
4. Calculate the first two moments of the Itô integral $I_n(t) = \int_0^t W^n(s)\,\mathrm{d}W(s)$.
5. Calculate the first two moments of the stochastic process $J_n(t) = W^n(t)$.
6. Calculate the first moment of the random process $K_n(t) = \int_0^t W^n(s)\,\mathrm{d}s$.
7. Explain why it is difficult to calculate the second moment of $K_n(t)$.
8. Find the variance of $K_1(t)$, $K_2(t)$ and $K_3(t)$.
9. What is the relationship between $I_n(t)$, $J_n(t)$ and $K_n(t)$?

A.4.5 Exponential of Brownian motion

Let $W(t)$ be a standard Brownian motion.

1. Find the stochastic differential $\mathrm{d}e^{W(t)}$.
2. Calculate the first two moments of $X(t) = e^{W(t)}$, $Y(t) = \int_0^t e^{W(s)}\,\mathrm{d}s$ and $Z(t) = \int_0^t e^{W(s)}\,\mathrm{d}W(s)$.
3. Deduce the correlation between $Y(t)$ and $Z(t)$.

A.4.6 Exponential martingales

1. Show that $X(t) = e^{W(t)}$ is a not martingale.
2. Find the function $m(t)$ such that $M(t) = m(t) X(t)$ is a martingale.
3. We assume that $X(t) = g(t)$ is non-random. Let Y be a $\mathcal{F}_t$-adapted process with:
$$\mathrm{d}Y(t) = -\frac{1}{2}g^2(t)\,\mathrm{d}t + g(t)\,\mathrm{d}W(t)$$
Find the solution of $M(t) = e^{Y(t)}$. Show that $M(t)$ is a martingale.
4. We now assume that $X(t) = g(t)$ is random. How can we show that $M(t)$ is a martingale?

A.4.7 Existence of solutions to stochastic differential equations

1. We consider the following SDE:
$$\mathrm{d}X(t) = (1 + X(t))\,\mathrm{d}t + 4\,\mathrm{d}W(t)$$
Show that it has a unique solution.
2. Let a, b and c be three scalars. Show that the following SDE has a unique solution:
$$\mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + cX(t)\,\mathrm{d}W(t)$$

A.4.8 Itô calculus and stochastic integration

1. Find the solution of:
$$\mathrm{d}X(t) = -\frac{X(t)}{1+t}\,\mathrm{d}t + \frac{1}{1+t}\,\mathrm{d}W(t)$$
2. Find the solution of[30]:
$$\mathrm{d}X(t) = X(t)\,\mathrm{d}t + X^2(t)\,\mathrm{d}W(t)$$
3. Find the stochastic differential of:
$$X(t) = \int_0^t \frac{1-t}{1-s}\,\mathrm{d}W(s)$$
4. Deduce the stochastic differential of $Y(t) = (1-t)^{-1} X(t)$. Find the solution of $Y(t)$.
5. Let $X(t) = f(t, W(t))$. Using Itô's lemma, find a necessary condition such that $X(t)$ is an $\mathcal{F}_t$-martingale.
6. Verify that the necessary condition is satisfied for the cubic martingale:
$$X(t) = W^3(t) - 3tW(t)$$
and the quartic martingale:
$$X(t) = W^4(t) - 6tW^2(t) + 3t^2$$
7. Show that $X(t) = e^{t/2}\cos W(t)$ is a martingale.

[30] Hint: use the transform function $f(t, X(t)) = 1/X(0) - 1/X(t)$.

A.4.9 Solving a PDE with the Feynman-Kac formula

We assume that:

$$\begin{cases} \mathrm{d}X(t) = \mathrm{d}t + \mathrm{d}W(t) \\ X(0) = x \end{cases}$$

Let $V(t,x)$ be the solution of the following partial differential equation:

$$\begin{cases} -\partial_t V(t,x) + 3V(t,x) = \dfrac{1}{2}\partial_x^2 V(t,x) + 2\partial_x V(t,x) + 4 \\ V(5,x) = x \end{cases} \tag{A.26}$$

1. Using Girsanov theorem, show that:

$$\mathrm{d}X(t) = 2\,\mathrm{d}t + \mathrm{d}Z(t)$$

   where $Z(t) = W(t) - \int_0^t \mathrm{d}s$ is a Brownian motion.

2. Compute $\mathbb{E}[X(5)|\mathcal{F}_t]$ and $\mathbb{E}[X(5)|\mathcal{G}_t]$ where $\mathcal{F}_t$ is the natural filtration and $\mathcal{G}_t$ is the filtration generated by the Brownian motion $Z(t)$. Find $\mathbb{E}[X(5)|\mathcal{G}_0]$.

3. Solve the PDE (A.26) and compute $V(0,x)$. Check that the solution satisfies the PDE.

4. What does the solution become when the terminal value $V(5,x)$ of the PDE is equal to e^x? Check that the solution satisfies the PDE.

A.4.10 Fokker-Planck equation

1. We consider the Ornstein-Uhlenbeck process:

$$\mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)$$

   How can we calculate the density function using the Feynman-Kac representation? Same question if we consider the Fokker-Planck equation. Solve numerically the two PDEs and draw the density function $\mathbb{P}\{X(1) = x \mid X(0) = 0\}$ when $a = 1$, $b = 10\%$ and $\sigma = 20\%$.

2. We consider the geometric Brownian motion:

$$\mathrm{d}X(t) = \mu X(t)\,\mathrm{d}t + \sigma X(t)\,\mathrm{d}W(t)$$

   How can we calculate the density function using the Feynmac-Kac representation? Same question if we consider the Fokker-Planck equation. Solve numerically the two PDEs and draw the density function $\mathbb{P}\{X(1) = x \mid X(0) = 100\}$ when $\mu = 10\%$ and $\sigma = 20\%$.

A.4.11 Dynamic strategy based on the current asset price

We assume that the price process $S(t)$ follows a diffusion process given by the following SDE:

$$\mathrm{d}S(t) = \mu(t, S(t))\,\mathrm{d}t + \sigma(t, S(t))\,\mathrm{d}W(t)$$

where $S(0) = S_0$. We consider a dynamic strategy $V(t)$ that consists in being long or short on the asset $S(t)$. We note $n(t)$ the number of shares at time t and we assume that it only depends on the current asset price:

$$n(t) = f(S(t))$$

1. Define $\mathrm{d}V(t)$.
2. We define the function $F(S)$ as follows:
$$F(S) = \int_c^S f(x)\,\mathrm{d}x$$
where c is a constant. Find the stochastic differential of $Y(t) = F(S(t))$.
3. Deduce an expression of the terminal value $V(T)$.
4. Show that $V(T)$ is composed of two terms:
$$V(T) = G(T) + C(T)$$
where $G(T)$ only depends on the initial and terminal values of S and $C(T)$ depends on the trajectory of S. How to interpret these two terms?
5. We consider the stop-loss strategy: $n(t) = \mathbb{1}\{S(t) > S_\star\}$ where $S_\star$ is the level of the stop[31]. Show that the option profile of this strategy is a long-only exposure to the asset plus a put option. What is the value of the option strike? What is the cost of this trading strategy?
6. We consider the stop-gain strategy: $n(t) = \mathbb{1}\{S(t) < S_\star\}$ where $S_\star$ is the level of the gain[32]. What is the option profile of this strategy? Why the cost of this trading strategy is positive?
7. We assume the following reversal strategy:
$$n(t) = m\frac{S_\star - S(t)}{S(t)}$$
where $S_\star$ is the price target of the asset and $m > 0$ is the leverage.
   (a) Explain the rationale of this strategy.
   (b) Find the value of $V(T)$.
   (c) We assume that the diffusion coefficient $\sigma(t, S(t))$ is equal to $\sigma(t)S(t)$. Show that:
$$C(T) = \frac{m}{2}S_\star \mathrm{IV}(T)$$
where $\mathrm{IV}(T)$ is the integrated variance.
   (d) Explain why the vega of the strategy is positive.

A.4.12 Strong Markov property and maximum of Brownian motion

1. Let $M(t) = \sup_{s\leq t} W(s)$ be the maximum of the Brownian motion. Show that:
$$\Pr\{M(t) \geq x\} = 2\left(1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right)$$
2. Let $x \geq 0$ and $y \leq x$. Show that:
$$\Pr\{W(t) \geq 2x - y\} = \Pr\{M(t) \geq x, W(t) \leq y\}$$

[31] We assume that $S(0) > S_\star$.
[32] We assume that $S(0) < S_\star$.

3. Calculate the joint density of $(M(t), W(t))$.
4. We now consider the maximum $M_X(t)$ of the process $X(t)$:
$$X(t) = \mu t + W(t)$$
Using Girsanov theorem, find the joint distribution of $(M_X(t), X(t))$.
5. Deduce the density function of $M_X(t)$.
6. Verify that the distribution $F(x)$ of the maximum is given by:
$$\Pr\{M_X(t) \leq x\} = \Phi\left(\frac{x-\mu t}{\sqrt{t}}\right) - e^{2\mu x}\Phi\left(\frac{-x-\mu t}{\sqrt{t}}\right)$$

A.4.13 Moments of the Cox-Ingersoll-Ross process

We consider the CIR process defined by:
$$\begin{cases} \mathrm{d}X(t) = a(b - X(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\,\mathrm{d}W(t) \\ X(0) = x_0 \end{cases}$$
We recall that $X(t)$ is related to the noncentral chi-squared distribution. Indeed, we have:
$$X(t) = \frac{1}{c}Y\left(\frac{4ab}{\sigma^2}, cx_0e^{-at}\right)$$
where $Y(\nu, \zeta)$ is a noncentral chi-squared random variable whose number of degrees of freedom is ν and noncentrality parameter is ζ, and:
$$c = \frac{4a}{(1-e^{-at})\,\sigma^2}$$
1. Calculate the mathematical expectation of $X(t)$.
2. Find the variance of $X(t)$.
3. Determine the skewness and excess kurtosis coefficients of $X(t)$.

A.4.14 Probability density function of Heston and SABR models

1. We consider the Heston model:
$$\begin{cases} \mathrm{d}X_1(t) = \mu X_1(t)\,\mathrm{d}t + \sqrt{X_2(t)}X_1(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}X_2(t) = a(b - X_2(t))\,\mathrm{d}t + \sigma\sqrt{X_2(t)}\,\mathrm{d}W_2(t) \end{cases}$$
where $\mathbb{E}[W_1(t)\,W_2(t)] = \rho t$. Write the Fokker-Planck equation.
2. We consider the SABR model:
$$\begin{cases} \mathrm{d}X_1(t) = X_2(t)\,X_1^{\beta}(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}X_2(t) = \nu X_2(t)\,\mathrm{d}W_2(t) \end{cases}$$
where $\mathbb{E}[W_1(t)\,W_2(t)] = \rho t$. Write the Fokker-Planck equation.
3. Solve numerically the Fokker-Planck equation for the Heston and SABR models. The parameters are the following:

(Heston) $\mu = 0$, $a = 2$, $b = 4\%$, $\sigma = 20\%$ and $\rho = -75\%$;
(SABR) $\beta = 1.00$, $\nu = 0.5$ and $\rho = -75\%$.

The initial values are $X_1(0) = 1$ and $X_2(0) = 6\%$. Draw the bivariate probability density function for $t = 1/2$.

A.4.15 Discrete dynamic programming

We assume that:

$$s(k+1) = g(k, s(k), c(k)) = s(k)$$

and:

$$f(k, s(k), c(k)) = -\frac{1}{s(k)} \ln s(k) - \frac{\alpha}{k} (c(k) - s(k))^2 + \beta c(k) + \gamma \sqrt{s(k)} e^{\sin s(k)}$$

The terminal value $f(K, s(K))$ is equal to $2s(K) - 1$. The state variable $s(k)$ takes the values $s_i = 1 + (i-1)/2$ where $i = 1, \ldots, n_S$ while the control variable $c(k)$ takes the value $v_j = j$ where $j = 1, \ldots, n_C$.

1. We set $\alpha = 0.02$, $\beta = 0.1$ and $\gamma = 0.01$.
   (a) Compute the matrices $\boldsymbol{J}$ and $\boldsymbol{C}$ when $K = 5$, $n_S = 4$ and $n_C = 8$.
   (b) Deduce the optimal value $\mathcal{J}(1, 1)$.
   (c) How do you explain that $c^\star(k) \not> 3$?
2. We set $\alpha = 0.02$, $\beta = 0.1$ and $\gamma = 0.01$.
   (a) Draw the values taken by $\mathcal{J}(k, s(k))$ when $K = 100$, $n_S = 100$ and $n_C = 25$.
   (b) What is the optimal state at time $k = 1$?
   (c) Find the optimal control $c^\star(k)$ when $s(k)$ is equal to 3, 13 and 22. Comment on these results.

A.4.16 Matrix computation

1. We consider the following matrix:

$$A = \begin{pmatrix} 1.000 & 0.500 & 0.700 \\ 0.500 & 0.900 & 0.200 \\ 0.700 & 0.200 & 0.300 \end{pmatrix}$$

   (a) Find the Schur decomposition.
   (b) Calculate e^A and $\ln A$.
   (c) How to compute $\cos A$ and $\sin A$? Calculate $\cos^2 A + \sin^2 A$.
   (d) Calculate $A^{1/2}$.
2. We consider the following covariance matrix:

$$\Sigma = \begin{pmatrix} 0.04000 & 0.01500 & 0.00200 & -0.00600 \\ 0.01500 & 0.02250 & -0.00375 & -0.00750 \\ 0.00200 & -0.00375 & 0.00250 & -0.00250 \\ -0.00600 & -0.00750 & -0.00250 & 0.01000 \end{pmatrix}$$

   Compute the nearest covariance matrix $\tilde{\Sigma}$.
3. We consider the matrix $B = \mathbb{C}_5(-50\%)$. Find the nearest correlation matrix $\rho(B)$. What do you observe? Generalize this result.

4. We consider the following matrix:

$$C = \begin{pmatrix} 1.0 & & & & \\ 0.9 & 1.0 & & & \\ 0.5 & 0.6 & 1.0 & & \\ 0.2 & 0.9 & 0.0 & 1.0 & \\ 0.9 & 0.0 & 0.9 & 0.0 & 1.0 \end{pmatrix}$$

Compute the nearest correlation matrix $\rho(C)$.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Bibliography

Abramowitz, M., and Stegun, I.A. (1970), *Handbook of Mathematical Functions*, Ninth edition, Dover.

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015), Systemic Risk and Stability in Financial Networks, *American Economic Review*, 105(2), pp. 564-608.

Acerbi, C., and Tasche, D. (2002), On the Coherence of Expected Shortfall, *Journal of Banking & Finance*, 26(7), pp. 1487-1503.

Acharya, V.V., Cooley, T., Richardson, M., and Walter, I. (2010), Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007-2009, *Foundations and Trends in Finance*, 4(4), pp. 247-325.

Acharya, V.V., Engle, R.F., and Richardson, M. (2012), Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks, *American Economic Review*, 102(3), pp. 59-64.

Acharya, V.V., and Pedersen, L.H. (2005), Asset Pricing with Liquidity Risk, *Journal of Financial Economics*, 77(2), pp. 375-410.

Acharya, V.V., Pedersen, L.H., Philippon, T., and Richardson, M. (2017), Measuring Systemic Risk, *Review of Financial Studies*, 30(1), pp. 2-47.

Acharya, V.V., and Yorulmazer, T. (2008), Information Contagion and Bank Herding, *Journal of Money, Credit and Banking*, 40(1), pp. 215-231.

Admati, A., and Hellwig, M. (2014), *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It*, Princeton University Press.

Adrian, T., and Brunnermeier, M.K. (2016), CoVaR, *American Economic Review*, 106(7), pp. 1705-1741.

Agarwal, S., Driscoll, J.C., and Laibson, D.I. (2013), Optimal Mortgage Refinancing: A Closed-Form Solution, *Journal of Money, Credit and Banking*, 45(4), pp. 591-622.

Ahn, D.H., Dittmar, R.F., and Gallant, A.R. (2002), Quadratic Term Structure Models: Theory and Evidence, *Review of Financial Studies*, 15(1), pp. 243-288.

Aldasoro, I., and Ehlers, T. (2018), The Geography of Dollar Funding of Non-US Banks, *BIS Quarterly Review*, December, pp. 15-26.

Alexander, C., and Sheedy, E. (2008), Developing a Stress Testing Framework Based on Market Risk Models, *Journal of Banking & Finance*, 32(10), pp. 2220-2236.

Allen, D.M. (1971), Mean Square Error of Prediction as a Criterion for Selecting Variables, *Technometrics*, 13(3), pp. 469-475.

Allen, D.M. (1974), The Relationship Between Variable Selection and Data Augmentation and a Method For Prediction, *Technometrics*, 16(1), pp. 125-127.

Allen, F., and Gale, D. (2000), Financial Contagion, *Journal of Political Economy*, 108(1), pp. 1-33.

Altman, E.I. (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, 23(4), pp. 589-609.

Altman, E.I., and Kalotay, E.A. (2014), Ultimate Recovery Mixtures, *Journal of Banking & Finance*, 40, pp. 116-129.

Altman, E.I., Sabato, G., and Wilson, N. (2010), The Value of Non-financial Information in Small and Medium-sized Enterprise Risk Management, *Journal of Credit Risk*, 6(2), pp. 1-33.

Amemiya, T. (1973), Regression Analysis when the Dependent Variable is Truncated Normal, *Econometrica*, 41(6), pp. 997-1016.

Amemiya, T. (1981), Qualitative Response Models: A Survey, *Journal of Economic Literature*, 19(4), pp. 1483-1536.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

Ames, M., Schuermann, T., and Scott, H.S. (2015), Bank Capital for Operational Risk: A Tale of Fragility and Instability, *SSRN*, www.ssrn.com/abstract=2396046.

Amihud, Y. (2002), Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets*, 5(1), pp. 31-56.

Andersen, L., and Sidenius, J. (2005), Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings, *Journal of Credit Risk*, 1(1), pp. 29-70.

Anderson, T.W., Olkin, I., and Underhill, L.G. (1987), Generation of Random Orthogonal Matrices, *SIAM Journal on Scientific and Statistical Computing*, 8(4), pp. 625-629.

Andrieu, C., Doucet, A., and Holenstein, R. (2010), Particle Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3), pp. 269-342.

Arellano-Valle, R.B., and Genton, M.G. (2005), On Fundamental Skew Distributions, *Journal of Multivariate Analysis*, 96(1), pp. 93-116.

Arrow, K.J. (1964), The Role of Securities in the Optimal Allocation of Risk-Bearing, *Review of Economic Studies*, 31(2), pp. 91-96.

Artzner, P., Delbaen, F., Eber, J.M., and Heath, D. (1999), Coherent Measures of Risk, *Mathematical Finance*, 9(3), pp. 203-228.

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. (2002), A Tutorial on Particle Filters for Online Nonlinear/non-Gaussian Bayesian Tracking, *IEEE Transactions on Signal Processing*, 50(2), pp. 174-188.

Association for Financial Markets in Europe (2019), Securtization Data Report Q4 2018 and 2018 Full Year, www.afme.eu/Divisions/Securitisation.

Augustin, P., Subrahmanyam, M.G., Tang, D.Y., and Wang, S.Q. (2014), Credit Default Swaps: A Survey, *Foundations and Trends in Finance*, 9(1-2), pp. 1-196.

Autorité des Marchés Financiers (2017), The Use of Stress Tests as Part of Risk Management — Guide for Asset Management Companies, February 2017.

Avellaneda, M., Levy, A., and Parás, A. (1995), Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities, *Applied Mathematical Finance*, 2(2), pp. 73-88.

Azzalini, A. (1985), A Class of Distributions Which Includes the Normal Ones, *Scandinavian Journal of Statistics*, 12(2), pp. 171-178.

Azzalini, A., and Capitanio, A. (1999), Statistical Applications of the Multivariate Skew Normal Distribution, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 61(3), pp. 579-602.

Azzalini, A., and Capitanio, A. (2003), Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-distribution, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 65(2), pp. 367-389.

Azzalini, A., and Dalla Valle, A. (1996), The Multivariate Skew-Normal Distribution, *Biometrika*, 83(4), pp. 715-726.

Bailey, R.W. (1994), Polar Generation of Random Variates with the t-distribution, *Mathematics of Computation*, 62(206), pp. 779-781.

Baillie, R.T. (1996), Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics*, 73(1), pp. 5-59.

Banasik, J., and Crook, J. (2007), Reject Inference, Augmentation, and Sample Selection, *European Journal of Operational Research*, 183(3), pp. 1582-1594.

Bank for International Settlements (2014), OTC Derivatives Market Activity in the First Half of 2014, www.bis.org/statistics.

Bank for International Settlements (2019), BIS Quarterly Review, March 2019, www.bis.org/forum/research.htm.

Bank of Japan (2014), Survey on Core Deposit Modeling in Japan: Toward Enhancing Asset Liability Management, *Working Paper*.

Bartlett, B. (2006), Hedging under SABR Model, *Wilmott Magazine*, July, pp. 2-4.

Basel Committee on Banking Supervision (1988), *International Convergence of Capital Measurement and Capital Standards*, July 1988.

Basel Committee on Banking Supervision (1995), *An Internal Model-based Approach to Market Risk Capital Requirements*, April 1995.

Basel Committee on Banking Supervision (1996a), *Amendment to the Capital Accord to Incorporate Market Risks*, January 1996.

Basel Committee on Banking Supervision (1996b), *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*, January 1996.

Basel Committee on Banking Supervision (1999), *A New Capital Adequacy Framework*, First consultative paper on Basel II, June 1999.

Basel Committee on Banking Supervision (2001a), *The New Basel Capital Accord*, Second consultative paper on Basel II, January 2001.

Basel Committee on Banking Supervision (2001b), *Results of the Second Quantitative Impact Study*, November 2001.

Basel Committee on Banking Supervision (2003), *The New Basel Capital Accord*, Third consultative paper on Basel II, April 2003.

Basel Committee on Banking Supervision (2004a), *International Convergence of Capital Measurement and Capital Standards — A Revised Framework*, June 2004.

Basel Committee on Banking Supervision (2004b), *Principles for the Management and Supervision of Interest Rate Risk*, Guidelines, July 2004.

Basel Committee on Banking Supervision (2006), *International Convergence of Capital Measurement and Capital Standards — A Revised Framework — Comprehensive version*, June 2006.

Basel Committee on Banking Supervision (2009a), *Enhancements to the Basel II Framework*, July 2009.

Basel Committee on Banking Supervision (2009b), *Revisions to the Basel II Market Risk Framework*, July 2009.

Basel Committee on Banking Supervision (2009c), *Guidelines for Computing Capital for Incremental Risk in the Trading Book*, July 2009.

Basel Committee on Banking Supervision (2009d), *Results from the 2008 Loss Data Collection Exercise for Operational Risk*, July 2009.

Basel Committee on Banking Supervision (2010), *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*, December 2010 (revision June 2011).

Basel Committee on Banking Supervision (2013a), *Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools*, January 2013.

Basel Committee on Banking Supervision (2013b), *Fundamental Review of the Trading Book: A Revised Market Risk Framework*, October 2013.

Basel Committee on Banking Supervision (2013c), *Capital requirements for banks' equity investments in funds*, December 2013.

Basel Committee on Banking Supervision (2014a), *Basel III Leverage Ratio Framework and Disclosure Requirements*, January 2014.

Basel Committee on Banking Supervision (2014b), *The Standardized Approach for Measuring Counterparty Credit Risk Exposures*, March 2014 (revision April 2014).

Basel Committee on Banking Supervision (2014c), *Supervisory Framework for Measuring and Controlling Large Exposures*, April 2014.

Basel Committee on Banking Supervision (2014d), *A Brief History of the Basel Committee*, October 2014.

Basel Committee on Banking Supervision (2014e), *Basel III: The Net Stable Funding Ratio*, October 2014.

Basel Committee on Banking Supervision (2014f), *Operational Risk — Revisions to the Simpler Approaches*, Consultative Document, October 2014.

Basel Committee on Banking Supervision (2014g), *The G-SIB Assessment Methodology — Score Calculation*, November 2014.

Basel Committee on Banking Supervision (2014h), *Fundamental Review of the Trading Book: Outstanding Issues*, Consultative Document, December 2014.

Basel Committee on Banking Supervision (2014i), *Capital Floors: The Design of a Framework based on Standardized Approaches*, Consultative Document, December 2014.

Basel Committee on Banking Supervision (2015a), *Revised Pillar 3 Disclosure Requirements*, January 2015.

Basel Committee on Banking Supervision (2015b), *Eight Progress Report on Adoption of the Basel Regulatory Framework*, April 2015.

Basel Committee on Banking Supervision (2015c), *Review of the Credit Valuation Adjustment Risk Framework*, Consultative Document, July 2015.

Basel Committee on Banking Supervision (2015d), *Revisions to the Standardized Approach for Credit Risk*, Consultative Document, December 2015.

Basel Committee on Banking Supervision (2016a), *Minimum Capital Requirements for Market Risk*, January 2016.

Basel Committee on Banking Supervision (2016b), *Standardised Measurement Approach for Operational Risk*, Consultative Document, March 2016.

Basel Committee on Banking Supervision (2016c), *Reducing Variation in Credit Risk-weighted Assets — Constraints on the Use of Internal Model Approaches*, Consultative Document, March 2016.

Basel Committee on Banking Supervision (2016d), *Interest Rate Risk in the Banking Book*, April 2016.

Basel Committee on Banking Supervision (2016e), *Revisions to the Securitisation Framework*, Consultative Document, July 2016.

Basel Committee on Banking Supervision (2017a), *Supervisory and Bank Stress Testing: Range of Practices*, December 2017.

Basel Committee on Banking Supervision (2017b), *Stress Testing Principles*, Consultative Document, December 2017.

Basel Committee on Banking Supervision (2017c), *Basel III: Finalising Post-crisis Reforms*, December 2017.

Basel Committee on Banking Supervision (2018), *Global Systemically Important Banks: Revised Assessment Methodology and the Higher Loss Absorbency Requirement*, July 2018.

Basel Committee on Banking Supervision (2019), *Minimum Capital Requirements for Market Risk*, January 2019.

Bastos, J.A. (2010), Forecasting Bank Loans Loss-Given-Default, *Journal of Banking & Finance*, 34(10), pp. 2510-2517.

Baud, N., Frachot, A., and Roncalli, T. (2002), Internal Data, External Data and Consortium Data — How to Mix Them for Measuring Operational Risk, *SSRN*, www.ssrn.com/abstract=1032529.

Baud, N., Frachot, A., and Roncalli, T. (2003), How to Avoid Overestimating Capital Charges for Operational Risk, *Operational Risk*, 4(2), pp. 14-19 (February 2003).

Baxter, M., and King, R.G. (1999), Measuring Business Cycles: Approximate Band-pass Filters for Economic Time Series, *Review of Economics and Statistics*, 81(4), pp. 575-593.

Beaver, W.H. (1966), Financial Ratios as Predictors of Failure, *Journal of Accounting Research*, Empirical Research in Accounting: Selected Studies 1966, 4, pp. 71-111.

Bellman, R. (1957), *Dynamic Programming*, Princeton University Press.

Bellotti, T., and Crook, J. (2012), Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards, *International Journal of Forecasting*, 28(1), pp. 171-182.

Bendel, R.B., and Mickey, M.R (1978), Population Correlation Matrices for Sampling Experiments, *Communications in Statistics — Simulation and Computation*, 7(2), pp. 163-182.

Benoit, S., Colliard, J.E., Hurlin, C., and Pérignon, C. (2017), Where the Risks Lie: A Survey on Systemic Risk, *Review of Finance*, 21(1), pp. 109-152.

Berestycki, H., Busca, J., and Florent, I. (2002), Asymptotics and Calibration of Local Volatility Models, *Quantitative Finance*, 2(1), pp. 61-69.

Bernstein, P.L. (1992), *Capital Ideas: The Improbable Origins of Modern Wall Street*, Free Press.

Bernstein, P.L. (2007), *Capital Ideas Evolving*, John Wiley & Sons.

Bertsekas, D.P. (2005), *Dynamic Programming and Optimal Control*, Volume 1, Third edition, Athena Scientific.

Bessis, J. (2015), *Risk Management in Banking*, Fourth edition, John Wiley & Sons.

Billio, M., Getmansky, M., Lo, A.W., and Pelizzon, L. (2012), Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors, *Journal of Financial Economics*, 104(3), pp. 535-559.

Bishop, C.M. (2006), *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer.

Black, F. (1976), The Pricing of Commodity Contracts, *Journal of Financial Economics*, 3(1-2), pp. 167-179.

Black, F. (1989), How We Came Up With The Option Formula, *Journal of Portfolio Management*, 15(2), pp. 4-8.

Black, F., and Cox, J.C. (1976), Valuing Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance*, 31(2), pp. 351-367.

Black, F. and Scholes, M. (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81(3), pp. 637-654.

Blöchlinger, A. (2015). Identifying, Valuing and Hedging of Embedded Options in Non-maturity Deposits, *Journal of Banking & Finance*, 50, pp. 34-51.

Board of Governors of the Federal Reserve System (2017), Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule, www.federalreserve.gov/supervisionreg/dfast-archive.htm.

Board of Governors of the Federal Reserve System (2019), Z.1 Financial Accounts of the United States, *Federal Reserve Statistics Release*, February 2019, www.federalreserve.gov/releases/z1.

Böcker, K., and Klüppelberg, C. (2005), Operational VaR: A Closed-form Approximation, *Risk Magazine*, 18(12), pp. 90-93.

Böcker, K., and Sprittulla, J. (2006), Operational VAR: Meaningful Means, *Risk Magazine*, 19(12), pp. 96-98.

Boggs, P.T. and Tolle, J.W. (1995), Sequential Quadratic Programming, *Acta Numerica*, 4, pp. 1-51.

Bollerslev, T., (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31(3), pp. 307-327.

Borio, C., Drehmann, M., and Tsatsaronis, K. (2014), Stress-testing Macro Stress Testing: Does it Live Up to Expectations?, *Journal of Financial Stability*, 12, pp. 3-15.

Boutsidis, C. and Gallopoulos, E. (2008), SVD Based Initialization: A Head Start for Nonnegative Matrix Factorization, *Pattern Recognition*, 41(4), pp. 1350-1362.

Bouveret, A. (2017), Liquidity Stress Tests for Investment Funds: A Practical Guide, *IMF Working Paper*, 17/226.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000), Copulas for Finance — A Reading Guide and Some Applications, *SSRN*, www.ssrn.com/abstract=1032533.

Box, G.E.P., and Muller, M.E. (1958), A Note on the Generation of Random Normal Deviates, *Annals of Mathematical Statistics*, 29(2), pp. 610-611.

Brace, A., Gatarek, D. and Musiela, M. (1997), The Market Model of Interest Rate Dynamics, *Mathematical Finance*, 7(2), pp. 127-155.

Brady, N.F. (1988), *Report of the Presidential Task Force on Market Mechanisms*, US Government Printing Office, Washington.

Breeden, D.T., and Litzenberger, R.H. (1978), State Contingent Prices Implicit in Option Prices, *Journal of Business*, 51(4), pp. 621-651.

Brei, M., and Gambacorta, L. (2014), The Leverage Ratio Over The Cycle, *BIS Working Paper*, 471.

Breiman, L. (1996), Bagging Predictors, *Machine Learning*, 24(2), pp. 123-140.

Breiman, L. (1998), Arcing Classifier, *Annals of Statistics*, 26(3), pp. 801-849.

Breiman, L. (2001), Random Forests, *Machine Learning*, 45(1), pp. 5-32.

Brennan, M.J., and Schwartz, E.S. (1979), A Continuous Time Approach to the Pricing of Bonds, *Journal of Banking & Finance*, 3(2), pp. 133-155.

Brennan, M.J., and Schwartz, E.S. (1985), Determinants of GNMA Mortgage Prices, *Real Estate Economics*, 13(3), pp. 209-228.

Brigo, D., and Mercurio, F. (2002a), Displaced and Mixture Diffusions for Analytically-Tractable Smile Models, in Geman, H., Madan, D.B., Pliska, S.R., and Vorst, A.C.F. (Eds), *Mathematical Finance — Bachelier Congress 2000*, Springer, pp. 151-174.

Brigo, D., and Mercurio, F. (2002b), Lognormal-mixture Dynamics and Calibration to Market Volatility Smiles, *International Journal of Theoretical and Applied Finance*, 5(4), pp. 427-446.

Brigo, D., and Mercurio, F. (2006), Interest Rate Models — Theory and Practice, Second edition, Springer.

Britten-Jones, M., and Schaefer, S.M. (1999), Non-Linear Value-at-Risk, *European Finance Review*, 2(2), pp. 161-187.

Broto, C., and Ruiz, E. (2004), Estimation Methods for Stochastic Volatility Models: A Survey, *Journal of Economic Surveys*, 18(5), pp. 613-649.

Brown, R.L., Durbin, J. and Evans, J.M. (1975), Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of the Royal Statistical Society B*, 37(2), pp. 149-192.

Brownlees, C.T., and Engle, R.F. (2016), SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, *Review of Financial Studies*, 30(1), pp. 48-79.

Brunnermeier, M.K., and Oehmke, M. (2013), Bubbles, Financial Crises, and Systemic Risk, Chapter 18 in Constantinides, G.M., Harris, M., and Stulz, R.M. (Eds), *Handbook of the Economics of Finance*, Volume 2, Part B, Elsevier, pp. 1221-1288.

Brunnermeier, M.K., and Pedersen, L.H. (2009), Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22(6), pp. 2201-2238.

Bundesanstalt für Finanzdienstleistungsaufsicht (2017), Liquidity Stress Testing by German Asset Management Companies, December 2017.

Burtschell, X., Gregory, J., and Laurent, J-P. (2007), Beyond the Gaussian Copula: Stochastic and Local Correlation, *Journal of Credit Risk*, 3(1), pp. 31-62.

Busch, R., Drescher, C., and Memmel, C. (2017), Bank Stress Testing under Different Balance Sheet Assumptions, *Deutsche Bundesbank Discussion Paper*, 07/2017.

Calomiris, C.W. (2009), The Subprime Turmoil: What's Old, What's New, and What's Next, *Journal of Structured Finance*, 15(1), pp. 6-52.

Canabarro, E., and Duffie, D. (2003), Measuring and Marking Counterparty Risk, Chapter 9 in Tilman, L. (Ed.), *Asset/Liability Management for Financial Institutions*, Institutional Investor Books.

Canabarro, E., Picoult, E., and Wilde, T. (2003), Analysing Counterparty Risk, *Risk Magazine*, 16(9), pp. 117-122.

Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992), A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling, *Journal of the American Statistical Association*, 87(418), pp. 493-500.

Carlton, D.W. (1984), Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures, *Journal of Futures Markets*, 4(3), pp. 237-271.

Carroll, R.B., Perry, T., Yang, H., and Ho, A. (2001), A New Approach to Component VaR, *Journal of Risk*, 3(3), pp. 57-67.

Casella, G., and George, E.I. (1992), Explaining the Gibbs Sampler, *American Statistician*, 46(3), pp. 167-174.

Cazalet, Z., and Roncalli, T. (2014), Facts and Fantasies About Factor Investing, *SSRN*, www.ssrn.com/abstract=2524547.

Cespedes, J.C.G., de Juan Herrero, J.A., Rosen, D., and Saunders, D. (2010), Effective Modeling of Wrong Way Risk, Counterparty Credit Risk Capital, and Alpha in Basel II, *Journal of Risk Model Validation*, 4(1), pp. 71-98.

Çetin, U., Jarrow, R.A., and Protter, P. (2004), Liquidity Risk and Arbitrage Pricing Theory, *Finance and Stochastics*, 8(3), pp. 311-341.

Çetin, U., Jarrow, R.A., Protter, P. and Warachka, M. (2006), Pricing Options in an Extended Black Scholes Economy with Illiquidity: Theory and Empirical Evidence, *Review of Financial Studies*, 19(2), pp. 493-529.

Çetin, U., Soner, H.M., and Touzi, N. (2010), Option Hedging for Small Investors under Liquidity Costs, *Finance and Stochastics*, 14(3), pp. 317-341.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., and Sanders, A.B. (1992), An Empirical Comparison of Alternative Models of the Short-term Interest Rate, *Journal of Finance*, 47(3), pp. 1209-1227.

Chen, S.X. (2007), Nonparametric Estimation of Expected Shortfall, *Journal of Financial Econometrics*, 6(1), pp. 87-107.

Chernov, M., Dunn, B.R., and Longstaff, F.A. (2017), Macroeconomic-driven Prepayment Risk and the Valuation of Mortgage-Backed Securities, *Review of Financial Studies*, 31(3), pp. 1132-1183.

Cherubini, U., and Luciano, E. (2002), Bivariate Option Pricing with Copulas, *Applied Mathematical Finance*, 9(2), pp. 69-85.

Chib, S., and Greenberg, E. (1995), Understanding the Metropolis-Hastings Algorithm, *American Statistician*, 49(4), pp. 327-335.

Chou, R.Y. (1988), Volatility Persistence and Stock Valuations: Some Empirical Evidence using GARCH, *Journal of Applied Econometrics*, 3(4), pp. 279-294.

Christoffersen, P., Jacobs, K., Jin, X., and Langlois, H. (2017), Dynamic Dependence and Diversification in Corporate Credit, *Review of Finance*, 22(2), pp. 521-560.

Cline, B.N., and Brooks, R. (2004), Embedded Options in Enhanced Certificates of Deposit, *Financial Services Review*, 13, pp. 19-32.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, 208, Springer.

Coles, S., Heffernan, J., and Tawn, J.A. (1999), Dependence Measures for Extreme Value Analyses, *Extremes*, 2(4), pp. 339-365.

Committee on the Global Financial System (2001), A Survey of Stress Tests and Current Practice at Major Financial Institutions, *CGFS Papers*, 18.

Committee on the Global Financial System (2005), Stress Testing at Major Financial Institutions: Survey Results and Practice, *CGFS Report*, 24.

Cont, R. (2001), Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues, *Quantitative Finance*, 1(2), pp. 223-236.

Cont, R. (2006), Model Uncertainty and its Impact on the Pricing of Derivative Instruments, *Mathematical Finance*, 16(3), pp. 519-547.

Cont, R. (2018), Margin Requirements for Non-cleared Derivatives, *ISDA Working Paper*.

Cont, R., Santos, E.B., and Moussa, A. (2013), Network Structure and Systemic Risk in Banking Systems, Chapter 13 in Fouque, J.P., and Langsam, J. (Eds), *Handbook of Systemic Risk*, Cambridge University Press, pp. 327-368.

Cortes, C., and Vapnik, V. (1995), Support-vector Networks, *Machine Learning*, 20(3), pp. 273-297.

Cox, D.R. (1958), The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 20(2), pp. 215-242.

Cox, D.R. (1972), Regression Models and Life-tables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2), pp. 187-220.

Cox, J.C., Ingersoll, J.E., and Ross, S.A. (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, 53(2), pp. 385-407.

Crammer, K., and Singer, Y. (2001), On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines, *Journal of Machine Learning Research*, 2, pp. 265-292.

Crépey, S. (2003), Calibration of the Local Volatility in a Generalized Black-Scholes Model Using Tikhonov Regularization, *SIAM Journal on Mathematical Analysis*, 34(5), pp. 1183-1206.

Crockford, G.N. (1982), The Bibliography and History of Risk Management: Some Preliminary Observations, *Geneva Papers on Risk and Insurance*, 7(23), pp. 169-179.

Crook, J., and Banasik, J. (2004), Does Reject Inference Really Improve the Performance of Application Scoring Models?, *Journal of Banking & Finance*, 28(4), pp. 857-874.

Crosbie, P., and Bohn, J. (2002), Modeling Default Risk, *KMV Working Paper*.

Crouhy, M., Galai, D., and Mark, R. (2013), *The Essentials of Risk Management*, Second edition, McGraw-Hill.

Dall'Aglio, G. (1972), Fréchet Classes and Compatibility of Distribution Functions, *Symposia Mathematica*, 9, pp. 131-150.

Daníelsson, J., and Zigrand, J.P. (2006), On Time-scaling of Risk and the Square-root-of-time Rule, *Journal of Banking & Finance*, 30(10), pp. 2701-2713.

Davidson, R., and MacKinnon, J.G. (2004), Econometric Theory and Methods, Oxford University Press.

Davies P.I., and Higham, N.J. (2000), Numerically Stable Generation of Correlation Matrices and Their Factors, *BIT Numerical Mathematics*, 40(4), pp. 640-651.

Davis, P.J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, Second edition, Academic Press.

De Bandt, O., and Hartmann, P. (2000), Systemic Risk: A Survey, *ECB Working Paper*, 35.

De Bandt, O., and Oung, V. (2004), Assessment of Stress Tests Conducted on the French Banking System, *Banque de France, Financial Stability Review*, 5, pp. 55-72.

De Jong, F., and Wielhouwer, J. (2003), The Valuation and Hedging of Variable Rate Savings Accounts, *ASTIN Bulletin*, 33(2), pp. 383-397.

Dees, S., Henry, J., and Martin, R. (2017), STAMP€: Stress-Test Analytics for Macroprudential Purposes in the Euro Area, *European Central Bank*, February 2017.

Deheuvels, P. (1978), Caractérisation Complète des Lois Extrêmes Multivariées et de la Convergence des Types Extrêmes, *Publications de l'Institut de Statistique de l'Université de Paris*, 23(3), pp. 1-36.

Deheuvels, P. (1979), La Fonction de Dépendance Empirique et ses Propriétés. Un Test non Paramétrique d'Indépendance, *Académie Royale de Belgique — Bulletin de la Classe des Sciences — 5ème Série*, 65(6), pp. 274-292.

Deheuvels, P. (1981), Multivariate Tests of Independence, in Dugue, D., Lukacs, E., and Rohatgi, V.K. (Eds), *Analytical Methods in Probability Theory (Proceedings of a Conference held at Oberwolfach in 1980)*, Lecture Notes in Mathematics, Volume 861, Springer.

DeMarzo, P., and Duffie, D. (1999), A Liquidity-based Model of Security Design, *Econometrica*, 67(1), pp. 65-99.

Dembiermont, C., Drehmann, M., and Muksakunratana, S. (2013), How Much does the Private Sector Really Borrow? A New Database for Total Credit to the Private Non-financial Sector, *BIS Quarterly Review*, March, pp. 65-81.

Demey, P., Frachot, A., and Riboulet, G. (2003), *Introduction à la Gestion Actif-Passif Bancaire*, Economica.

Demey, P., Jouanin, J.-F., Roget, C., and Roncalli, T. (2004), Maximum Likelihood Estimation of Default Correlations, *Risk Magazine*, 17(11), pp. 104-108.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society B*, 39(1), pp. 1-38.

Denault, M. (2001), Coherent Allocation of Risk Capital, *Journal of Risk*, 4(1), pp. 1-34.

Deng, Y., Quigley, J.M., and Van Order, R. (2000), Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options, *Econometrica*, 68(2), pp. 275-307.

Dent, K., Westwood, B., and Segoviano, M.A. (2016), Stress Testing of Banks: An Introduction, *Bank of England Quarterly Bulletin*, 56(3), pp. 130-143.

Derman, E. (1996), Markets and Models, *Quantitative Strategies Research Notes*, Goldman Sachs, April 1996.

Derman, E. (2001), Markets and Models, *Risk Magazine*, 14(7), pp. 48-50.

Derman, E., and Kani, I. (1994), Riding on a Smile, *Risk Magazine*, 7(2), pp. 139-145.

Derman, E., Kani, I., and Kamal, M. (1996), Trading and Hedging Local Volatility, *Quantitative Strategies Research Notes*, Goldman Sachs.

Devenow, A., and Welch, I. (1996), Rational Herding in Financial Economics, *European Economic Review*, 40(3), pp. 603-615.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag.

Devroye, L. (2010), On Exact Simulation Algorithms for Some Distributions Related to Brownian Motion and Brownian Meanders, Chapter 1 in Devroye, L., Karasözen, B., Kohler, M. and Korn, R. (Eds), *Recent Developments in Applied Probability and Statistics: Dedicated to the Memory of Jürgen Lehn*, Physica-Verlag Heidelberg, pp. 1-35.

Diamond, D.W., and Dybvig, P.H. (1983), Bank Runs, Deposit Insurance, and Liquidity, *Journal of Political Economy*, 91(3), pp. 401-419.

Dickey, D.A., and Fuller, W.A. (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American statistical association*, 74(366), pp. 427-431.

Dickey, D.A., and Fuller, W.A. (1981), Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 49(4), pp. 1057-1072.

Diebold, F.X., Hickman, A., Inoue, A., and Schuermann, T. (1998), Scale Models, *Risk Magazine*, 11(1), pp. 104-107.

Doucet, A., De Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, Springer.

Doucet, A., and Johansen, A.M. (2009), A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later, Chapter 24 in Crisan, D, and Rozovskii, B. (Eds), *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, pp. 656-704.

Drehmann, M., and Nikolaou, K. (2013), Funding Liquidity Risk: Definition and Measurement, *Journal of Banking & Finance*, 37(7), pp. 2173-2182.

Duffie, D., and Huang, M. (1996), Swap Rates and Credit Quality, *Journal of Finance*, 51(3), pp. 921-949.

Duffie, D., and Pan, J. (1997), An Overview of Value at Risk, *Journal of Derivatives*, 4(3), pp. 7-49.

Duffie, D., and Rahi, R. (1995), Financial Innovation and Security Design: An Introduction, *Journal of Economic Theory*, 65(1), pp. 1-42.

Duffie, D. and Singleton, K.J. (1993), Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61(4), pp. 929-952.

Dunbar, N., and Irving, R. (1998), This is the Way the World Ends, *Risk Magazine*, 11, December.

Dunn, K.B., and McConnell, J.J. (1981), Valuation of GNMA Mortgage-Backed Securities, *Journal of Finance*, 36(3), pp. 599-616.

Dupire, B. (1994), Pricing with a Smile, *Risk Magazine*, 7(1), pp. 18-20.

Dupire, B. (1998), *Monte Carlo: Methodologies and Applications for Pricing and Risk Management*, Risk Books.

Durand, D. (1941), Risk Elements in Consumer Instalment Financing, *Financial Research Program: Studies in Consumer Instalment Financing*, 8, National Bureau of Economic Research.

Durrleman, V. (2010), From Implied to Spot Volatilities, *Finance and Stochastics*, 14(2), pp 157-177.

Dwyer, D., and Korablev, I. (2009), Moody's KMV LossCalc V3.0, *Moody's KMV*, April 2009.

Eberlein, E., Keller, U., and Prause, K. (1998), New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model, *Journal of Business*, 71(3), pp. 371-405.

El Karoui, N., Jeanblanc, M., and Shreve, S.E. (1998), Robustness of the Black and Scholes Formula, *Mathematical Finance*, 8(2), pp. 93-126.

El Karoui, N., Myneni, R., and Viswanathan, R. (1992a), Arbitrage Pricing and Hedging of Interest Rate Claims with State Variables: I Theory, *Working Paper*, University of Paris VI.

El Karoui, N., Myneni, R., and Viswanathan, R. (1992b), Arbitrage Pricing and Hedging of Interest Rate Claims with State Variables: II Applications, *Working Paper*, University of Paris VI.

Elderfield, M. (2013), The Regulatory Agenda Facing the Insurance Industry, *Address to the European Insurance Forum 2013*, Dublin, May 9.

Elie, R., Frachot, A., Georges, P., and Neuvial, P. (2002), A Model of Prepayment for the French Residential Loan Market, *Working Paper*.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer.

Emmer, S., and Tasche, D. (2005), Calculating Credit Risk Capital Charges with the One-factor Model, *Journal of Risk*, 7(2), pp. 85-101.

Engle, R.F. (1976), Interpreting Spectral Analyses in terms of Time-domain Models, *Annals of Economic and Social Measurement*, 5(1), pp. 89-109.

Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50(4), pp. 987-1007.

Engle, R.F., and Bollerslev, T. (1986), Modelling the Persistence of Conditional Variances, *Econometric Reviews*, 5(1), pp. 1-50.

Engle, R.F., and Granger, C.W.J. (1987), Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55(2), pp. 251-276.

Engle, R.F., Jondeau, E., and Rockinger, M. (2015), Systemic Risk in Europe, *Review of Finance*, 19(1), pp. 145-190.

Epperlein, E., and Smillie, A. (2006), Cracking VaR with Kernels, *Risk Magazine*, 19(8), pp. 70-74.

European Banking Authority (2015a), *EBA Report on CVA*, February 2015.

European Banking Authority (2015b), *Draft EBA Guidelines on Limits on Exposures to Shadow Banking Entities which Carry out Banking Activities Outside a Regulated Framework under Article 395 para. 2 Regulation (EU) No. 575/2013*, Consultation Document, March 2015.

European Banking Authority (2017), *EBA Report on IRB Modelling Practices*, November 2017.

European Banking Authority (2018a), *Guidelines on the Management of Interest Rate Risk Arising From Non-Trading Book Activities*, July 2018.

European Banking Authority (2018b), *Guidelines on Institutions' Stress Testing*, July 2018.

Falkenstein, E., Boral, A., and Carty, L. (2000), RiskCalc for Private Companies: Moody's Default Model, *SSRN*, www.ssrn.com/abstract=236011.

Fama, E.F., and French, K.R. (1993), Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33(1), pp. 3-56.

Fawcett, T. (2006), An Introduction to ROC Analysis, *Pattern Recognition Letters*, 27(8), pp. 861-874.

Federal Deposit Insurance Corporation (2019), FDIC Quarterly Banking Profile: Fourth Quarter 2018, 13(1), www.fdic.gov/bank/analytical/quarterly.

Federal Housing Finance Agency (2018), *Prepayment Monitoring Report*, Second Quarter 2018, July.

Fengler, M.R. (2009), Arbitrage-free Smoothing of the Implied Volatility Surface, *Quantitative Finance*, 9(4), pp. 417-428.

Filipović, D. (2002), Separable Term Structures and the Maximal Degree Problem, *Mathematical Finance*, 12(4), pp. 341-349.

Financial Stability Board (2009), *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations*, Report to the G-20 Finance Ministers and Central Bank Governors, October 2009.

Financial Stability Board (2010), *Reducing the Moral Hazard Posed by Systemically Important Financial Institutions*, Consultation Document, October 2010.

Financial Stability Board (2011), *Shadow Banking: Strengthening Oversight and Regulation*, October 2011.

Financial Stability Board (2013), *Strengthening Oversight and Regulation of Shadow Banking: An Overview of Policy Recommendations*, August 2013.

Financial Stability Board (2015a), *Assessment Methodologies for Identifying Non-bank Non-insurer Global Systemically Important Financial Institutions*, Second Consultation Document, March 2015.

Financial Stability Board (2015b), *2015 Update of List of Global Systemically Important Banks (G-SIBs)*, November 2015.

Financial Stability Board (2015c), *2015 Update of List of Global Systemically Important Insurers (G-SIIs)*, November 2015.

Financial Stability Board (2015d), *Total Loss-Absorbing Capacity (TLAC) Principles and Term Sheet*, November 2015.

Financial Stability Board (2015e), *Global Shadow Banking Monitoring Report 2015*, November 2015.

Financial Stability Board (2016), *2016 List of Global Systemically Important Insurers (G-SIIs)*, November 2016.

Financial Stability Board (2017), *Policy Recommendations to Address Structural Vulnerabilities from Asset Management Activities*, January 2017.

Financial Stability Board (2018a), *Global Shadow Banking Monitoring Report 2017*, March 2018.

Financial Stability Board (2018b), *2018 List of Global Systemically Important Banks (G-SIBs)*, November 2018.

Finkelstein, V., Lardy, J.P., Pan, G., Ta, T., and Tierney, J. (2002), *CreditGrades Technical Document*, RiskMetrics Group.

Fisher, I. (1935), *100% Money*, Adelphi Company.

Fisher, R.A. (1929), Tests of Significance in Harmonic Analysis, *Proceedings of the Royal Society of London A*, 125(796), pp. 54-59.

Fisher, R.A. (1935), The Logic of Inductive Inference, *Journal of the Royal Statistical Society*, 98(1), pp. 39-82.

Fisher, R.A. (1936), The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*, 7(2), pp. 179-188.

Föllmer, H., and Schied, A. (2002), Convex Measures of Risk and Trading Constraints, *Finance and Stochastics*, 6(4), pp. 429-447.

Forde, M., and Jacquier, A. (2009), Small-time Asymptotics for Implied Volatility under the Heston Model, *International Journal of Theoretical and Applied Finance*, 12(06), pp. 861-876.

Frachot, A. (2001), A Note on Behavioral Models for Managing Optionality in Banking Books, *Working Paper*.

Frachot, A., Georges, P., and Roncalli, T. (2001), Loss Distribution Approach for Operational Risk, *SSRN*, www.ssrn.com/abstract=1032523.

Frachot, A., Moudoulaud, O., and Roncalli, T. (2006) Loss Distribution Approach in Practice, Chapter 23 in Ong, M.K. (Ed.), *The Basel Handbook*, Second edition, Risk Books.

Frachot, A., Roncalli, T., and Salomon, E. (2004), Correlation and Diversification Effects in Operational Risk Modelling, *Operational Risk*, 5(5), pp. 34-38 (May 2004).

Freund, Y. (1995), Boosting a Weak Learning Algorithm by Majority, *Information and Computation*, 121(2), pp. 256-285.

Freund, Y., and Schapire, R.E. (1997), A Decision-theoretic Generalization of On-line Learning and an Application to Boosting, *Journal of Computer and System Sciences*, 55(1), pp. 119-139.

Friedman, A. (1975), Stochastic Differential Equations and Applications, Academic Press (reprinted in 2006 by Dover).

Friedman, J.H. (2002), Stochastic Gradient Boosting, *Computational Statistics & Data Analysis*, 38(4), pp. 367-378.

Friedman, J.H., Hastie, T., and Tibshirani, R. (2000). Additive Logistic Regression: A Statistical View of Boosting, *Annals of Statistics*, 28(2), pp. 337-407.

Frishling, V. (2002), A Discrete Question, *Risk Magazine*, 15(1), pp. 115-116.

Galambos, J. (1982), The Role of Functional Equations in Stochastic Model Building, *Aequationes Mathematicae*, 25(1), pp. 21-41.

Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics*, Second edition, Krieger Publishing.

Galambos, J., and Kotz, S. (1978), *Characterizations of Probability Distributions*, Lecture Notes in Mathematics, 675, Springer.

Gallagher, R.B. (1956), Risk Management: New Phase of Cost Control, *Harvard Business Review*, 34(5), pp. 75-86.

Gao, F., He, A.X., and He, P. (2018), A Theory of Intermediated Investment with Hyperbolic Discounting Investors, *Journal of Economic Theory*, 177, pp. 70-100.

Gasull, A., Jolis, M., and Utzet, F. (2015), On the Norming Constants for Normal Maxima, *Journal of Mathematical Analysis and Applications*, 422(1), pp. 376-396.

Gatheral, J. (2004), A Parsimonious Arbitrage-free Implied Volatility Parameterization with Application to the Valuation of Volatility Derivatives, *Global Derivatives and Risk Management 2004*, Madrid.

Gatheral, J., and Jacquier, A. (2011), Convergence of Heston to SVI, *Quantitative Finance*, 11(8), pp. 1129-1132.

Gatheral, J., and Jacquier, A. (2014), Arbitrage-free SVI Volatility Surfaces, *Quantitative Finance*, 14(1), pp. 59-71.

Geanakoplos, J. (2010), The Leverage Cycle, Chapter 1 in Acemoglu, D., Rogoff K.S., and Woodford, M. (Eds), *NBER Macroeconomics Annual 2009*, Volume 24, pp. 1-65.

Gelfand, A.E., and Smith, A.F.M. (1990), Sampling-based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, 85(410), pp. 398-409.

Geman, H., El Karoui, N., and Rochet, J.C. (1995), Changes of Numéraire, Changes of Probability Measure and Option Pricing, *Journal of Applied probability*, 32(2), pp. 443-458.

Geman, S., and Geman, D. (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 721-741.

Genest, C., Ghoudi, K., and Rivest, L.P. (1995), A Semiparametric Estimation Procedure for Dependence Parameters in Multivariate Families of Distributions, *Biometrika*, 82(3), pp. 543-552.

Genest, C., and MacKay, J. (1986a), Copules Archimédiennes et Familles de Lois Bidimensionnelles dont les Marges sont Données, *Canadian Journal of Statistics*, 14(2), pp. 145-159.

Genest, C., and MacKay, J. (1986b), The Joy of Copulas: Bivariate Distributions with Uniform Marginals, *American Statistician*, 40(4), pp. 280-283.

Gennotte, G., and Leland, H.E. (1990), Market Liquidity, Hedging, and Crashes, *American Economic Review*, 80(5), pp. 999-1021.

Georges, P., Lamy, A.G., Nicolas, E., Quibel, G., and Roncalli, T. (2001), Multivariate Survival Modelling: A Unified Approach with Copulas, *SSRN*, www.ssrn.com/abstract=1032559.

Getter, D.E. (2014), U.S. Implementation of the Basel Capital Regulatory Framework, *Congressional Research Service*, R42744, www.crs.gov.

Geweke, J., and Porter-Hudak, S. (1983), The Estimation and Application of Long Memory Time Series Models, *Journal of Time Series Analysis*, 4(4), pp. 221-238.

Ghoudi, K., Khoudraji, A., and Rivest, L.P. (1998), Propriétés Statistiques des Copules de Valeurs Extrêmes Bidimensionnelles, *Canadian Journal of Statistics*, 26(1), pp. 187-197.

Gikhman, I.I., and Skorokhod, A.V. (1972), Stochastic Differential Equations, Springer.

Gilkeson, J.H., List, J.A., and Ruff, C.K. (1999), Evidence of Early Withdrawal in Time Deposit Portfolios, *Journal of Financial Services Research*, 15(2), pp. 103-122.

Gilkeson, J.H., Porter, G.E., and Smith, S.D. (2000), The Impact of the Early Withdrawal Option on Time Deposit Pricing, *Quarterly Review of Economics and Finance*, 40(1), pp. 107-120.

Glasserman, P. (2003), *Monte Carlo Methods in Financial Engineering*, Springer.

Glasserman, P. (2005), Measuring Marginal Risk Contributions in Credit Portfolios, *Journal of Computational Finance*, 9(2), pp. 1-41.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (1999), Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options, *Mathematical Finance*, 9(2), pp. 117-152.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2002), Portfolio Value-at-Risk with Heavy-tailed Risk Factors, *Mathematical Finance*, 12(3), pp. 239-269.

Glasserman, P., Kang, C., and Kang, W. (2015), Stress Scenario Selection by Empirical Likelihood, *Quantitative Finance*, 15(1), pp. 25-41.

Glasserman, P., and Li, J. (2005), Importance Sampling for Portfolio Credit Risk, *Management Science*, 51(11), pp. 1643-1656.

Golub, G.H., and Van Loan, C.F. (2013), *Matrix Computations*, Fourth edition, Johns Hopkins University Press.

Golub, G.H., and Welsch, J.H. (1969), Calculation of Gauss Quadrature Rules, *Mathematics of Computation*, 23(106), pp. 221-230.

Goodhart, C., Hofmann, B., and Segoviano, M.A. (2004), Bank Regulation and Macroeconomic Fluctuations, *Oxford Review of Economic Policy*, 20(4), pp. 591-615.

Gordy, M.B. (2000), A Comparative Anatomy of Credit Risk Models, *Journal of Banking & Finance*, 24(1-2), pp. 119-149.

Gordy, M.B. (2003), A Risk-factor Model Foundation for Ratings-based Bank Capital Rules, *Journal of Financial Intermediation*, 12(3), pp. 199-232.

Gordy, M.B. (2004), Granularity Adjustment in Portfolio Credit Risk Measurement, Chapter 8 in Szegö, G. (Ed.), *Risk Measures for the 21st Century*, John Wiley & Sons.

Gordy, M.B., and Heitfield, E. (2002), Estimating Default Correlations from Short Panels of Credit Rating Performance Data, *Working Paper*.

Gordy, M.B., and Jones, D. (2003), Random Tranches, *Risk Magazine*, 16(3), pp. 78-83.

Gordy, M.B., and Lütkebohmert, E. (2013), Granularity Adjustment for Regulatory Capital Assessment, *International Journal of Central Banking*, 9(3), pp. 33-70.

Gordy, M.B., and Marrone, J. (2012), Granularity Adjustment for Mark-to-market Credit Risk Models, *Journal of Banking & Finance*, 36(7), pp. 1896-1910.

Gouriéroux, C. (1992), Courbes de Performance, de Sélection et de Discrimination, *Annales d'Économie et de Statistique*, 28(4), 1992, pp. 107-123.

Gouriéroux, C., and Jasiak, J. (2007), *The Econometrics of Individual Risk: Credit, Insurance, and Marketing*, Princeton University Press.

Gouriéroux, C., Laurent, J-P., and Scaillet, O. (2000), Sensitivity Analysis of Values at Risk, *Journal of Empirical Finance*, 7(3-4), pp. 225-245.

Goyenko, R.Y., Holden, C.W., and Trzcinka, C.A. (2009), Do Liquidity Measures Measure Liquidity?, *Journal of Financial Economics*, 92(2), pp. 153-181.

Granger, C.W.J., and Joyeux, R. (1980), An Introduction to Long-Memory Time Series Models and Fractional Differencing, *Journal of Time Series Analysis*, 1(1), pp. 15-29.

Granger, C.W.J., and Newbold, P. (1974), Spurious Regressions in Econometrics, *Journal of Econometrics*, 2(2), pp. 111-120.

Gray, R.M. (2006), Toeplitz and Circulant Matrices: A Review, *Foundations and Trends in Communications and Information Theory*, 2(3), pp. 155-239.

Greene, W.H. (2017), *Econometric Analysis*, Eighth edition, Pearson.

Grinblatt, M., Titman, S., and Wermers, R. (1995), Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, *American Economic Review*, 85(5), pp. 1088-1105.

Grossman, S.J., and Miller, M.H. (1988), Liquidity and Market Structure, *Journal of Finance*, 43(3), pp. 617-633.

Grundke, P. (2011), Reverse Stress Tests with Bottom-up Approaches, *Journal of Risk Model Validation*, 5(1), pp. 71-90.

Grundke, P., and Pliszka, K. (2018), A Macroeconomic Reverse Stress Test, *Review of Quantitative Finance and Accounting*, 50(4), pp. 1093-1130.

Guill, G.D. (2009), Bankers Trust and the Birth of Modern Risk Management, Wharton School, *Financial Institutions Center's Case Studies*, April.

Gupton G.M., Finger C.C. and Bhatia M. (1997), *CreditMetrics — Technical Document*, J.P. Morgan.

Hagan, P.S., Kumar, D., Lesniewski, A.S., and Woodward, D.E. (2002), Managing Smile Risk, *Wilmott Magazine*, September, pp. 84-108.

Hallerbach, W.G. (2003), Decomposing Portfolio Value-at-Risk: A General Analysis, *Journal of Risk*, 5(2), pp. 1-18.

Hand, D.J. (2006), Classifier Technology and the Illusion of Progress, *Statistical Science*, 21(1), pp. 1-14.

Hand, D.J., and Henley, W.E. (1997), Statistical Classification Methods in Consumer Credit Scoring: A Review, *Journal of the Royal Statistical Society: Series A*, 160(3), pp. 523-541.

Hansen, L.P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50(4), pp. 1029-1054.

Hansen, L.P. (2012), Challenges in Identifying and Measuring Systemic Risk, *National Bureau of Economic Research*, 18505.

Harvey, A.C. (1990), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A.C., Ruiz, E., and Shephard, N. (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies*, 61(2), pp. 247-264.

Harvey, A.C., and Shephard, N. (1996), Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns, *Journal of Business and Economic Statistics*, 14(4), pp. 429-434.

Hasbrouck, J. (2009), Trading Costs and Returns for US Equities: Estimating Effective Costs from Daily Data, *Journal of Finance*, 64(3), pp. 1445-1477.

Hasbrouck, J., and Schwartz, R.A. (1988), Liquidity and Execution Costs in Equity Markets, *Journal of Portfolio Management*, 14(3), pp. 10-16.

Hastie, T., Tibshirani, R., and Friedman, J.H. (2009), *The Elements of Statistical Learning*, Second edition, Springer.

Hastings, W.K. (1970), Monte Carlo Sampling Methods using Markov Chains and Their Applications, *Biometrika*, 57(1), pp. 97-109.

Haug, E.G., Haug, J., and Lewis, A. (2003), Back to Basics: A New Approach to the Discrete Dividend Problem, *Wilmott Magazine*, September, pp. 37-47.

Hayre, L.S., Chaudhary, S., and Young, R.A. (2000), Anatomy of Prepayments, *Journal of Fixed Income*, 10(1), pp. 19-49.

Heath, D., Jarrow, R.A., and Morton, A. (1992), Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, *Econometrica*, 60(1), pp. 77-105.

Heston, S.L. (1993), A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6(2), pp. 327-343.

Higham, N.J. (2002), Computing the Nearest Correlation Matrix — A Problem from Finance, *IMA Journal of Numerical Analysis*, 22(3), pp. 329-343.

Hillegeist, S.A., Keating, E.K., Cram, D.P., and Lundstedt, K.G. (2004), Assessing the Probability of Bankruptcy, *Review of Accounting Studies*, 9(1), pp. 5-34.

Ho, T.S.Y., and Lee, S.B. (1986), Term Structure Movements and Pricing Interest Rate Contingent Claims, *Journal of Finance*, 41(5), pp. 1011-1029.

Hodrick, R.J., and Prescott, E.C. (1997), Postwar US Business Cycles: An Empirical Investigation, *Journal of Money, credit, and Banking*, 29(1) pp. 1-16.

Hoerl, A.E. and Kennard, R.W. (1970), Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12(1), pp. 55-67.

Horn, R.A., and Johnson, C.R. (2012), *Matrix Analysis*, Second edition, Cambridge University Press.

Hosking, J.R.M. (1981), Fractional Differencing, *Biometrika*, 68(1), pp. 165-176.

Hsu, C.W., and Lin, C.J. (2002), A Comparison of Methods for Multiclass Support Vector Machines, *IEEE Transactions on Neural Networks*, 13(2), pp. 415-425.

Huang, R.D., and Stoll, H.R. (1996), Dealer versus Auction Markets: A Paired Comparison of Execution Costs on NASDAQ and the NYSE, *Journal of Financial Economics*, 41(3), pp. 313-357.

Huang, R.D., and Stoll, H.R. (1997), The Components of the Bid-ask Spread: A General Approach, *Review of Financial Studies*, 10(4), pp. 995-1034.

Huber, P.J. (1964), Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics*, 35(1), pp. 73-101.

Hull, J., and White, A. (1990), Pricing Interest-rate-derivative Securities, *Review of Financial Studies*, 3(4), pp. 573-592.

Hull, J., and White, A. (1994), Numerical Procedures for Implementing Term Structure Models I: Single-factor Models, *Journal of Derivatives*, 2(1), pp. 7-16.

Hull, J., and White, A. (2012), CVA and Wrong-way Risk, *Financial Analysts Journal*, 68(5), pp. 58-69.

Hutchison, D.E., and Pennacchi, G.G. (1996), Measuring Rents and Interest Rate Risk in Imperfect Financial Markets: The Case of Retail Bank Deposits, *Journal of Financial and Quantitative Analysis*, 31(3), pp. 399-417.

Ieda, A., Marumo, K., and Yoshiba, T. (2000), A Simplified Method for Calculating the Credit Risk of Lending Portfolios, *Institute for Monetary and Economic Studies, Bank of Japan*, 2000-E-10.

International Association of Insurance Supervisors (2013a), *Global Systemically Important Insurers: Initial Assessment Methodology*, July 2013.

International Association of Insurance Supervisors (2013b), *Global Systemically Important Insurers: Policy Measures*, July 2013.

International Association of Insurance Supervisors (2013c), *Insurance Core Principles, Standards, Guidance and Assessment Methodology*, October 2013.

International Monetary Fund, (2014a), *Review of the Financial Sector Assessment Program: Further Adaptation to the Post Crisis Era*, September 2014.

International Monetary Fund (2014b), *Global Financial Stability Report — Risk Taking, Liquidity, and Shadow Banking: Curbing Excess While Promoting Growth*, October 2014.

International Organization of Securities Commissions (2012a), *Policy Recommendations for Money Market Funds*, October 2012.

International Organization of Securities Commissions (2012b), *Global Developments in Securitisation Regulation*, November 2012.

International Organization of Securities Commissions (2015a), *Peer Review of Regulation of Money Market Funds*, September 2015.

International Organization of Securities Commissions (2015b), *Peer Review of Implementation of Incentive Alignment Recommendations for Securitisation*, September 2015.

International Swaps and Derivatives Association (2003), *2003 ISDA Credit Derivatives Definitions*, https://www.isda.org.

International Swaps and Derivatives Association (2014a), *2014 ISDA Credit Derivatives Definitions*, https://www.isda.org.

International Swaps and Derivatives Association (2014b), *Exposure Draft: Credit and Debit Valuation Adjustments*, Letter to International Valuation Standard Council, https://www.isda.org.

Israel, R.B., Rosenthal, J.S., and Wei, J.Z. (2001), Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings, *Mathematical Finance*, 11(2), pp. 245-265.

Jacklin, C.J., Kleidon, A.W., and Pfeiderer, P. (1992), Underestimation of Portfolio Insurance and the Crash of October 1987, *Review of Financial Studies*, 5(1), pp. 35-63.

Jacobs, Jr., M. (2010), An Empirical Study of Exposure at Default, *Journal of Advanced Studies in Finance*, 1(1), pp. 31-59.

Jafry, Y., and Schuermann, T. (2004), Measurement, Estimation and Comparison of Credit Migration Matrices, *Journal of Banking & Finance*, 28(11), pp. 2603-2639.

Jamshidian, F. (1989), An Exact Bond Option Formula, *Journal of Finance*, 44(1), pp. 205-209.

Jamshidian, F. (1997), LIBOR and Swap Market Models and Measures, *Finance and Stochastics*, 1(4), pp. 293-330.

Janosi, T., Jarrow, R.A., and Zullo, F. (1999), An Empirical Analysis of the Jarrow-van Deventer Model for Valuing Non-maturity Demand Deposits, *Journal of Derivatives*, 7(1), pp. 8-31.

Jarrow, R.A., Lando, D., and Turnbull, S.M. (1997), A Markov Model for the Term Structure of Credit Risk Spreads, *Review of Financial Studies*, 10(2), pp. 481-523.

Jarrow, R.A., and Protter, P. (2007), Liquidity Risk and Option Pricing Theory, Chapter 17 in Birge, J.R. and Linetsky, V. (Eds), *Handbooks in Operations Research and Management Science — Financial Engineering*, Volume 15, Elsevier, pp. 727-762.

Jarrow, R.A., and van Deventer, D.R. (1998), The Arbitrage-free Valuation and Hedging of Demand Deposits and Credit Card Loans, *Journal of Banking & Finance*, 22(3), pp. 249-272.

Jickling, M., and Murphy, E.V. (2010), Who Regulates Whom? An Overview of U.S. Financial Supervision, *Congressional Research Service*, R40249, www.crs.gov.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, 73, Chapmann and Hall.

Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12(2-3), pp. 231-254.

Johansen, S. (1991), Cointegration and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59(6), pp. 1551-1580.

Jolliffe, I.T. (2002), *Principal Component Analysis*, Springer Series in Statistics, Second edition, Springer.

Jones, E.P., Mason, S.P., and Rosenfeld, E. (1984), Contingent Claims Analysis of Corporate Capital Structures: An Empirical Investigation, *Journal of Finance*, 39(3), pp. 611-625.

Jones, M.C., Marron, J.S., and Sheather, S.J. (1996), A Brief Survey of Bandwidth Selection for Density Estimation, *Journal of the American Statistical Association*, 91(433), pp. 401-407.

Jorion, P. (2000), Risk Management Lessons from Long-Term Capital Management, *European Financial Management*, 6(3), pp. 277-300.

Jorion, P. (2007), *Value at Risk: The New Benchmark for Managing Financial Risk*, Third edition, McGraw-Hill.

Jouanin, J.-F., Riboulet, G. and Roncalli, T. (2004), Financial Applications of Copula Functions, Chapter 14 in Szegö, G. (Ed.), *Risk Measures for the 21st Century*, John Wiley & Sons.

J.P. Morgan, (1996), *RiskMetrics — Technical Document*, Fourth edition.

Kacperczyk, M., and Schnabl, P. (2013), How Safe are Money Market Funds?, *Quarterly Journal of Economics*, 128(3), pp. 1073-1122.

Kahl, C., and Jäckel, P. (2005), Not-so-complex Logarithms in the Heston Model, *Wilmott Magazine*, September, pp. 94-103.

Kalkbrener, M. (2005), An Axiomatic Approach to Capital Allocation, *Mathematical Finance*, 15(3), pp. 425-437.

Kalkbrener, M., and Willing, J. (2004), Risk Management of Non-Maturing Liabilities, *Journal of Banking & Finance*, 28(7), pp. 1547-1568.

Karatzas, I., and Shreve, S.E. (1991), Brownian Motion and Stochastic Calculus, Second edition, Springer.

Kashyap, A.K., and Stein, J.C. (2004), Cyclical Implications of the Basel II Capital Standards, *Federal Reserve Bank Of Chicago, Economic Perspectives*, 28(1), pp. 18-33.

Kavvathas, D. (2001), Estimating Credit Rating Transition Probabilities for Corporate Bonds, *SSRN*, www.ssrn.com/abstract=248421.

Keys, B.J., Pope, D.G., and Pope, J.C. (2016), Failure to Refinance, *Journal of Financial Economics*, 122(3), pp. 482-499.

Kim, S., Shephard, N., and Chib, S. (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65(3), pp. 361-393.

Kitagawa, G. (1996), Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *Journal of Computational and Graphical Statistics*, 5(1), pp. 1-25.

Kloeden, P.E., and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, 23, Springer.

Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2012), *Loss Models: From Data to Decisions*, Wiley Series in Probability and Statistics, 715, Fourth edition, John Wiley & Sons.

Koenker, R. (2005), *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

Kopeliovich, Y., Novosyolov, A., Satchkov, D., and Schachter, B. (2015), Robust Risk Estimation and Hedging: A Reverse Stress Testing Approach, *Journal of Derivatives*, 22(4), pp. 10-25.

Kupiec, P.H. (1998), Stress Testing in a Value at Risk Framework, *Journal of Derivatives*, 6(1), pp. 7-24.

Kurpiel, A., and Roncalli, T. (2000), Hopscotch Methods for Two-state Financial Models, *Journal of Computational Finance*, 3(2), pp. 53-90.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and, Shin, Y. (1992), Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root, *Journal of Econometrics*, 54(1-3), pp. 159-178.

L'Ecuyer, P. (1999), Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators, *Operations Research*, 47(1), pp. 159-164.

Lamberton, D., and Lapeyre, B. (2007), *Introduction to Stochastic Calculus Applied to Finance*, Second edition, CRC Press.

Laurent, J.P., and Gregory, J. (2005), Basket Default Swaps, CDOs and Factor Copulas, *Journal of Risk*, 7(4), pp. 103-122.

Lee, D.D. and Seung, H.S. (1999), Learning the Parts of Objects by Non-negative Matrix Factorization, *Nature*, 401, pp. 788-791.

Lee, D.D. and Seung, H.S. (2001), Algorithms for Non-negative Matrix Factorization, in Leen, T.K., Dietterich, T.G. and Tresp, V. (Eds), *Advances in Neural Information Processing Systems*, 13, pp. 556-562.

Lee, R. (2004), The Moment Formula for Implied Volatility at Extreme Strikes, *Mathematical Finance*, 14(3), pp. 469-480.

Lee, R., and Wang, D. (2012), Displaced Lognormal Volatility Skews: Analysis and Applications to Stochastic Volatility Simulations, *Annals of Finance*, 8(2-3), pp. 159-181.

Lee, S.X., and McLachlan, G.J. (2013), On Mixtures of Skew Normal and Skew t-distributions, *Advances in Data Analysis and Classification*, 7(3), pp. 241-266.

Lehmann, E.L. (1999), *Elements of Large-sample Theory*, Springer Texts in Statistics, Springer.

Leland, H.E., and Rubinstein, M. (1988), Comments on the Market Crash: Six Months After, *Journal of Economic Perspectives*, 2(3), pp. 45-50.

Lennox, C. (1999), Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches, *Journal of Economics and Business*, 51(4), pp. 347-364.

Levy-Carciente, S., Kenett, D.Y., Avakian, A., Stanley, H.E., and Havlin, S. (2015), Dynamical Macroprudential Stress Testing Using Network Theory, *Journal of Banking & Finance*, 59, pp. 164-181.

Lewis, P.A.W., Goodman, A.S., and Miller, J.M. (1969), A Pseudo-random Number Generator for the System/360, *IBM Systems Journal*, 8(2), pp. 136-146.

Li, D.X. (2000), On Default Correlation: A Copula Function Approach, *Journal of Fixed Income*, 9(4),pp. 43-54.

Lin, C-J. (2007), Projected Gradient Methods for Non-negative Matrix Factorization, *Neural Computation*, 19(10), pp. 2756-2779.

Liptser, R., and Shiryaev, A.N. (1974), *Statistics of Random Processes, I. General Theory*, Springer.

Litterman, R.B. (1996), Hot Spots and Hedges, *Risk Management Series*, Goldman Sachs, October 1996.

Litterman, R.B., and Scheinkman, J.A. (1991), Common Factors Affecting Bond Returns, *Journal of Fixed Income*, 1(1), pp. 54-61.

Little, R.J.A., and Rubin, D.B. (2014), *Statistical Analysis with Missing Data*, Second edition, Wiley Series in Probability and Statistics, 333, John Wiley & Sons.

Litzenberger, R.H. (1992), Swaps: Plain and Fanciful, *Journal of Finance*, 47(3), pp. 831-851.

Lo, A.W. (1991), Long-Term Memory in Stock Market Prices, *Econometrica*, 59(5), pp. 1279-1313.

Longstaff, F.A., Mithal, S., and Neis, E. (2005), Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market, *Journal of Finance*, 60(5), pp. 2213-2253.

Longstaff, F.A., and Schwartz, E.S. (1992), Interest Rate Volatility and the Term Structure: A Two-factor General Equilibrium Model, *Journal of Finance*, 47(4), pp. 1259-1282.

Lopez, J.A. (2005), Stress Tests: Useful Complements to Financial Risk Models, *FRBSF Economic Letter*, 2005-14.

Loterman, G., Brown, I., Martens, D., Mues, C., and Baesens, B. (2012), Benchmarking Regression Algorithms for Loss Given Default Modeling, *International Journal of Forecasting*, 28(1), pp. 161-170.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.

MacKinnon, J. (1996), Numerical Distribution Functions for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, 11, pp. 601-618.

Maillard, D. (2018), A User's Guide to the Cornish Fisher Expansion, *SSRN*, www.ssrn.com/abstract=1997178.

Mark, N.C. (1988), Time-varying Betas and Risk Premia in the Pricing of Forward Foreign Exchange Contracts, *Journal of Financial Economics*, 22(2), pp. 335-354.

Markit (2014), *Markit Credit Indices: A Primer*, January 2014, www.markit.com.

Markowitz, H. (1952), Portfolio Selection, *Journal of Finance*, 7(1), pp. 77-91.

Marsaglia, G., and Olkin, I. (1984), Generating Correlation Matrices, *SIAM Journal on Scientific and Statistical Computing*, 5(2), pp. 470-475.

Marshall, A.W., and Olkin, I (1988), Families of Multivariate Distributions, *Journal of the American Statistical Association*, 83(403), pp. 834-841.

McFadden, D. (1984), Econometric Analysis of Qualitative Response Models, Chapter 24 in Griliches, Z. and Intriligator, M.D. (Eds), *Handbook of Econometrics*, 2, pp. 1395-1457.

McFadden, D. (1989), A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica*, 57(5), pp. 995-1026.

McNeil, A.J. (2008), Sampling Nested Archimedean Copulas, *Journal of Statistical Computation and Simulation*, 78(6), pp. 567-581.

Mehr, R.I., and Hedges, B.A. (1963), *Risk Management in the Business Enterprise*, Richard D. Irwin, Inc., babel.hathitrust.org/cgi/pt?id=mdp.39076005845305.

Merton, R.C. (1973), Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science*, 4(1), pp. 141-183.

Merton, R.C. (1974), On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance*, 29(2), pp. 449-470.

Merton, R.C. (1976), Option Pricing when Underlying Stock Returns are Discontinuous, *Journal of Financial Economics*, 3(1-2), pp. 125-144.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), Equation of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 21(6), pp. 1087-1092.

Metropolis, N. and Ulam, S. (1949), The Monte Carlo Method, *Journal of the American Statistical Association*, 44(247), p. 335-341.

Meucci, A. (2005), *Risk and Asset Allocation*, Springer.

Modigliani, F., and Miller, M.H. (1958), The Cost of Capital, Corporation Finance and the Theory of Investment, *American Economic Review*, 48(3), pp. 261-297.

Moler, C., and Van Loan, C.F. (2003), Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later, *SIAM Review*, 45(1), pp. 3-49.

Morini, M. (2001), *Understanding and Managing Model Risk: A Practical Guide for Quants, Traders and Validators*, Risk Books.

Morokoff, W.J., and Caflisch, R.E. (1994), Quasi-Random Sequences and Their Discrepancies, *SIAM Journal on Scientific Computing*, 15(6), pp. 1251-1279.

Murphy, E.V. (2015), Who Regulates Whom and How? An Overview of U.S. Financial Regulatory Policy for Banking and Securities Markets, *Congressional Research Service*, R43087, www.crs.gov.

Nagpal, K., and Bahar, R. (2001), Measuring Default Correlation, *Risk Magazine*, 14(3), pp. 129-132.

Nelsen, R.B. (2006), *An Introduction to Copulas*, Second edition, Springer.

Nelson, C.R., and Siegel, A.F. (1987), Parsimonious Modeling of Yield Curves, *Journal of Business*, 60(4), pp. 473-489.

Newey, W.K., and West, K.D. (1987), A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55(3), pp. 703-708.

Nikolaou, K. (2009), Liquidity (Risk) Concepts: Definitions and Interactions, *ECB Working Paper*, 1009.

Nocedal, J. and Wright, S.J. (2006), *Numerical Optimization*, Second edition, Springer.

O'Brien, J.M. (2001), Estimating the Value and Interest Rate Risk of Interest-Bearing Transactions Deposits, Federal Reserve Board, *FEDS Working Paper*, 2000-53.

O'Kane, D. (2008), *Modelling Single-name and Multi-name Credit Derivatives*, John Wiley & Sons.

Oakes, D. (1989), Bivariate Survival Models Induced by Frailties, *Journal of the American Statistical Association*, 84 (406), pp. 487-493.

Office of the Comptroller of the Currency (2011), *Selected Asset and Liability Price Tables As of June 30, 2011*, US Department of the Treasury.

Office of Thrift Supervision (2001), *The OTS Net Portfolio Value Model*, US Department of the Treasury.

Ohlson, J.A. (1980), Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18(1), pp. 109-131.

ØKSENDAL, B. (2010), *Stochastic Differential Equations: An Introduction with Applications*, Sixth edition, Springer.

ONG, L.L. (2014), *A Guide to IMF Stress Testing: Methods and Models*, International Monetary Fund.

PAKES, A., and POLLARD, D. (1989), Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, 57(5), pp. 1027-1057.

PANJER, H.H. (1981), Recursive Evaluation of a Family of Compound Distributions, *ASTIN Bulletin: The Journal of the IAA*, 12(1), pp. 22-26.

PAWITAN Y., and O'SULLIVAN, F. (1994), Nonparametric Spectral Density Estimation using Penalized Whittle Likelihood, *Journal of the American Statistical Association*, 89(426), pp. 600-610.

PERSAUD, A.D. (2003), *Liquidity Black Holes: Understanding, Quantifying and Managing Financial Liquidity Risk*, Risk Books.

PHILLIPS, P.C.B., and PERRON, P. (1988), Testing for Unit Roots in Time Series Regression, *Biometrika*, 75(2), pp. 335-346.

PICKANDS, J. (1975), Statistical Inference using Extreme Order Statistics, *Annals of Statistics*, 3(1), pp. 119-131.

PLATEN, E., and WAGNER, W. (1982), On a Taylor Formula for a Class of Ito Processes *Probability and Mathematical Statistics*, 3(1), pp. 37-51.

POLLOCK, D.S.G. (1999), *Handbook of Time Series Analysis, Signal Processing, and Dynamics*, Academic Press.

POWERS, D.M. (2011), Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation, *Journal of Machine Learning Technologies*, 2(1), pp 37-63.

POZSAR, Z., ADRIAN, T., ASHCRAFT, A.B., and BOESKY, H. (2013), Shadow Banking, *Federal Reserve Bank of New York, Economic Policy Review*, 19(2), pp. 1-16.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T., and FLANNERY, B.P. (2007), *Numerical Recipes: The Art of Scientific Computing*, Third edition, Cambridge University Press.

PRIESTLEY, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press.

PYKHTIN, M. (2012), Model Foundations of the Basel III Standardised CVA Charge, *Risk Magazine*, 25(7), pp. 60-66.

PYKHTIN, M., and ROSEN, D. (2010), Pricing Counterparty Risk at the Trade Level and CVA Allocations, *Journal of Credit Risk*, 6(4), pp. 3-38.

PYKHTIN, M., and ZHU, S.H. (2006), Measuring Counterparty Credit Risk for Trading Products Under Basel II, Chapter 6 in Ong, M.K. (Ed.), *The Basel Handbook*, Second edition, Risk Books.

PYKHTIN, M., and ZHU, S.H. (2007), A Guide to Modeling Counterparty Credit Risk, *GARP Risk Review*, 37(7), pp. 16-22.

Qi, M. (2009), Exposure at Default of Unsecured Credit Cards, *OCC Economics Working Paper*, Washington, DC: Office of the Comptroller of the Currency, 2009-2.

Quesada Molina, J.J., and Rodríguez Lallena, J.A. (1994), Some Advances in the Study of the Compatibility of Three Bivariate Copulas, *Journal of Italian Statistical Society*, 3(3), pp. 397-417.

Rapuch, G., and Roncalli, T. (2004), Technical Note: Dependence and Two-asset Options Pricing, *Journal of Computational Finance*, 7(4), pp. 23-33.

Rebonato, R. (2001), Model Risk: New Challenges, New Solutions, *Risk Magazine*, 14(3), pp. 87-90.

Rebonato, R. (2002), *Modern Pricing of Interest-rate Derivatives: The LIBOR Market Model and Beyond*, Princeton University Press.

Redon, C. (2006), Wrong Way Risk Modelling, *Risk Magazine*, 19(4), pp. 90-95.

Reinhart, C.M. and Rogoff, K.S. (2009), *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press.

Resnick, S.I. (1987), *Extreme Values, Point Processes and Regular Variation*, Springer.

Revuz, D., and Yor, M. (1999), *Continuous Martingales and Brownian Motion*, Fundamental Principles of Mathematical Sciences, 293, Third edition, Springer.

Ritchken, P., and Sankarasubramanian, L. (1995), Volatility Structures of Forward Rates and the Dynamics of the Term Structure, *Mathematical Finance*, 5(1), pp. 55-72.

Roll, R. (1984), A Simple Implicit Measure of the Effective Bid-ask Spread in an Efficient Market, *Journal of Finance*, 39(4), pp. 1127-1139.

Roncalli, T. (1996), *TSM — Gauss Library for Time Series and Wavelets Modeling*, Global Design.

Roncalli, T. (2009), *La Gestion des Risques Financiers*, Second edition, Economica.

Roncalli, T. (2013), *Introduction to Risk Parity and Budgeting*, Chapman and Hall/CRC Financial Mathematics Series.

Roncalli, T., and Weisang, G. (2009), Tracking Problems, Hedge Fund Replication and Alternative Beta, *SSRN*, www.ssrn.com/abstract=1325190.

Roncalli, T., and Weisang, G. (2015), Asset Management and Systemic Risk, *SSRN*, www.ssrn.com/abstract=2610174.

Roncalli, T., and Zheng, B. (2015), Measuring the Liquidity of ETFs: An Application to the European Market, *Journal of Trading*, 9(3), pp. 79-108.

Rosen, D., and Saunders, D. (2012), CVA The Wrong Way, *Journal of Risk Management in Financial Institutions*, 5(3), pp. 252-272.

Rosenblatt, F. (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, 65(6), pp. 386-408.

Ross, S.A. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, 13(3), pp. 341-360.

Ross, S.M. (2012), *Simulation*, Fifth edition, Academic Press.

Rubinstein, M. (1983), Displaced Diffusion Option Pricing, *Journal of Finance*, 38(1), pp. 213-217.

Rubinstein, M., and Reiner, E. (1991), Breaking Down the Barriers, *Risk Magazine*, 4(8), pp. 28-35.

Ruder, S. (2016), An Overview of Gradient Descent Optimization Algorithms, *arXiv*, 1609.04747.

Salmon, M. (1982), Error Correction Mechanisms, *Economic Journal*, 92(367), pp. 615-629.

Sancetta, A., and Satchell, S.E. (2004), Bernstein Copula and its Applications to Modeling and Approximations of Multivariate Distributions, *Econometric Theory*, 20(3), pp. 535-562.

Sarr, A., and Lybek, T. (2002), Measuring Liquidity in Financial Markets, *IMF Working Paper*, 02/232.

Scaillet, O. (2004), Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall, *Mathematical Finance*, 14(1), pp. 115-129.

Schaede, U. (1989), Forwards and Futures in Tokugawa-Period Japan: A New Perspective on the Dojima Rice Market, *Journal of Banking & Finance*, 13(4), pp. 487-513.

Schapire, R.E. (1990), The Strength of Weak Learnability, *Machine Learning*, 5(2), pp. 197-227.

Schmee, J., and Hahn, G.J. (1979), A Simple Method for Regression Analysis with Censored Data, *Technometrics*, 21(4), pp. 417-432.

Schmidt W., and Ward I. (2002), Pricing Default Baskets, *Risk Magazine*, 15(1), pp. 111-114.

Schönbucher, P.J. (1999), A Market Model for Stochastic Implied Volatility, *Philosophical Transactions of the Royal Society of London, Series A*, 357(1758), pp. 2071-2092.

Schuermann, T. (2004), What Do We Know About Loss Given Default?, in Shimko, D. (Ed.), *Credit Risk Models and Management*, Risk Books.

Schwartz, E.S., and Torous, W.N. (1989), Prepayment and the Valuation of Mortgage-Backed Securities, *Journal of Finance*, 44(2), pp. 375-392.

Schwartz, E.S., and Torous, W.N. (1992), Prepayment, Default, and the Valuation of Mortgage Pass-through Securities, *Journal of Business*, 65(2), pp. 221-239.

Schweizer, B., and Wolff, E.F. (1981), On Nonparametric Measures of Dependence for Random Variables, *Annals of Statistics*, 9(4), pp. 879-885.

Securities Industry and Financial Markets Association (2019a), US Bond Market Issuance and Outstanding, February 2019, www.sifma.org/resources/archive/research.

Securities Industry and Financial Markets Association (2019b), US Mortgage-Related Issuance and Outstanding, Februray 2019, www.sifma.org/resources/archive/research.

Securities Industry and Financial Markets Association (2019c), US ABS Issuance and Outstanding, February 2019, www.sifma.org/resources/archive/research.

Segoviano, M.A., Jones, B., Lindner P., and Blankenheim, J. (2013), Securitization: Lessons Learned and the Road Ahead, *IMF Working Paper*, 13/255.

Segoviano, M.A., and Padilla, P. (2014), Portfolio Credit Risk and Macroeconomic Shocks: Applications to Stress Testing under Data-restricted Environments, Chapter 31 in Ong, L.L. (Ed.), *A Guide to IMF Stress Testing: Methods and Models*, International Monetary Fund, pp. 485-511.

Selvaggio, R. (1996), Using the OAS Methodology to Value and Hedge Commercial Bank Retail Demand Deposit Premiums, in Fabozzi, F.J. and Konishi, A. (Eds), *The Handbook of Asset/Liability Management*, Irwin/McGraw-Hill, pp. 363-373.

Sharpe, W.F. (1964), Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, 19(3), pp. 425-442.

Shih, J.H., and Louis, T.A. (1995), Inferences on the Association Parameter in Copula Models for Bivariate Survival Data, *Biometrics*, 51(4), pp. 1384-1399.

Shiller, R.J. (1987), Investor Behavior in the October 1987 Stock Market Crash: Survey Evidence, *National Bureau of Economic Research*, 2446.

Shumway, T. (2001), Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *Journal of Business*, 74(1), pp. 101-124.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (26), Chapman and Hall.

Sims, C.A. (1980), Macroeconomics and Reality, *Econometrica*, 48(1), pp. 1-48.

Sklar, A. (1959), Fonctions de Répartition à n Dimensions et leurs Marges, *Publications de l'Institut de Statistique de l'Université de Paris*, 8(1), pp. 229-231.

Skintzi, V.D., and Refenes, A.N. (2003), Implied Correlation Index: A New Measure of Diversification, *Working Paper*, Athens University.

Smith, M. (1993), *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold.

Snider, H.W. (1956), The Risk Manager, *Insurance Law Journal*, pp. 119-125.

Snider, H.W. (1991), Risk Management: A Retrospective View, *Risk Management*, 38(4), pp. 47-54.

Sowell, F. (1992), Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models, *Journal of Econometrics*, 53(1-3), pp. 165-188.

Spanos, A. (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press.

Spector, L.C., and Mazzeo, M. (1980), Probit Analysis and Economic Education, *Journal of Economic Education*, 11(2), pp. 37-44.

Stahl, G. (1997), Three Cheers, *Risk Magazine*, 10(5), pp. 67-69.

Stanhouse, B., and Stock, D. (2004), The Impact of Loan Prepayment Risk and Deposit Withdrawal Risk on the Optimal Intermediation Margin, *Journal of Banking & Finance*, 28(8), pp. 1825-1843.

Stewart, G.W. (1980), The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators, *SIAM Journal on Numerical Analysis*, 17(3), pp. 403-409.

Stoer, J., and Bulirsch, R. (1993), *Introduction to Numerical Analysis*, Text in Applied Mathematics, 12, Second edition, Springer.

Stone, M. (1977), An Asymptotic Equivalence of Choice of Model by Cross-validation and Akaike's Criterion, *Journal of the Royal Statistical Society B*, 39(1), pp. 44-47.

Stulz, R.M. (1996), Rethinking Risk Management, *Journal of Applied Corporate Finance*, 9(3), pp. 8-25.

Syrkin, M., and Shirazi, A. (2015), Potential Future Exposure (PFE), Credit Value Adjustment (CVA) and Wrong Way Risk (WWR) Analytic Solutions, *SSRN*, www.ssrn.com/abstract=2588685.

Tamari, M. (1966), Financial Ratios as a Means of Forecasting Bankruptcy, *Management International Review*, 6(4), pp. 15-21.

Tankov, P. (2011), Improved Fréchet Bounds and Model-free Pricing of Multi-asset Options, *Journal of Applied Probability*, 48(2), pp. 389-403.

Tanner, M.A. (1993), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer.

Taplin, R., To, H.M., and Hee, J. (2007), Modeling Exposure at Default, Credit Conversion Factors and the Basel II Accord, *Journal of Credit Risk*, 3(2), pp. 75-84.

Tarullo, D.K. (2008), *Banking on Basel: The Future of International Financial Regulation*, Peterson Institute for International Economics.

Tarullo, D.K. (2016), *Next Steps in the Evolution of Stress Testing*, Speech given at the Yale University School of Management Leaders Forum, September 26, 2016.

Tasche, D. (2002), Expected Shortfall and Beyond, *Journal of Banking & Finance*, 26(7), pp. 1519-1533.

Tasche, D. (2008), Capital Allocation to Business Units and Sub-Portfolios: The Euler Principle, Chapter 17 in Resti, A. (Ed.), *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, Risk Books, pp. 423-453.

Tawn, J.A. (1988), Bivariate Extreme Value Theory: Models and Estimation, *Biometrika*, 75(3), pp. 397-415.

Tawn, J.A. (1990), Modelling Multivariate Extreme Value Distributions, *Biometrika*, 77(2), pp. 245-253.

Tchen, A.H. (1980), Inequalities for Distributions with Given Marginals, *Annals of Probability*, 8(4), pp. 814-827.

Thomas, L.C. (2000), A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers, *International Journal of Forecasting*, 16(2), pp. 149-172.

Thomas, L.C., Crook, J.N., and Edelman, D.B. (2017), Credit Scoring and Its Applications, Second edition, SIAM.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society B*, 58(1), pp. 267-288.

Tong, E.N.C, Mues, C., Brown, I., and Thomas, L.C. (2016), Exposure at Default models With and Without the Credit Conversion Factor, *European Journal of Operational Research*, 252(3), pp. 910-920.

Tong, E.N.C., Mues, C., and Thomas, L.C. (2013), A Zero-adjusted Gamma Model for Mortgage Loan Loss Given Default, *International Journal of Forecasting*, 29(4), pp. 548-562.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press.

Tufano, P., and Kyrillos, B.B. (1995), Leland O'Brien Rubinstein Associates, Inc.: Portfolio Insurance, *Harvard Business School Case*, 294061.

Vasicek, O. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5(2), pp. 177-188.

Vasicek, O. (1987), Probability of Loss on Loan Portfolio, *KMV Working Paper*.

Vasicek, O. (1991), Limiting Loan Loss Probability Distribution, *KMV Working Paper*.

Vasicek, O. (2002), The Distribution of Loan Portfolio Value, *Risk Magazine*, 15(12), pp. 160-162.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley & Sons.

Wang, J.N., Yeh, J.H., and Cheng, N.Y.P. (2011), How Accurate is the Square-root-of-time Rule in Scaling Tail Risk: A Global Study, *Journal of Banking & Finance*, 35(5), pp. 1158-1169.

Wermers, R. (1999), Mutual Fund Herding and the Impact on Stock Prices, *Journal of Finance*, 54(2), pp. 581-622.

West, D. (2000), Neural Network Credit Scoring Models, *Computers & Operations Research*, 27(11-12), pp. 1131-1152.

Whittle, P. (1953), The Analysis of Multiple Stationary Time Series, *Journal of the Royal Statistical Society B*, 15(1), pp. 125-139.

Wilde, T. (2001a), IRB Approach Explained, *Risk Magazine*, 14(5), pp. 87-90.

Wilde, T. (2001b), Probing Granularity, *Risk Magazine*, 14(8), pp. 103-106.

Wilmott, P. (2002), Cliquet Options and Volatility Models, *Wilmott Magazine*, December, pp. 78-83.

Witzany, J. (2011), Exposure at Default Modeling with Default Intensities, *SSRN*, www.ssrn.com/abstract=2489672.

Yang, B.H., and Tkachenko, M. (2012), Modeling Exposure at Default and Loss Given Default: Empirical Approaches and Technical Implementation, *Journal of Credit Risk*, 8(2), pp. 81-102.

Zangari, P. (1996), A VaR Methodology for Portfolios That Include Options, *RiskMetrics Monitor*, First quarter, pp. 4-12.

Zigrand, J.P. (2014), Systems and Systemic Risk in Finance and Economics, *Systemic Risk Centre, Special Paper Series*, 1.

Zou, H., and Hastie, T. (2005), Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society: Series B*, 67(2), pp. 301-320.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Subject Index

A

ABS, 5, **137–140**, 154–155
Accounting valuation, 23, 278, 373–374
ADV, 129, 350
Advanced internal ratings-based approach, *see* IRB
Advanced Measurement Approaches, *see* AMA
AIFMD, 487
ALCO, **375–376**, 383, 393
ALM, 2, 24, 357, **369–452**
Alternative investment fund managers directive, *see* AIFMD
AM-CVA, 284
AMA, 19, **310–311**
Analytical VaR, *see* VaR
Ancillary own funds, 24
Antithetic variates, 844–850
APT, 2, 76, 455
AR process, *see* ARMA process
Arbitrage pricing theory, *see* APT
ARCH process, 81, 634, **664–668**
Archimedean copula, *see* Copula
ARFIMA process, *see* Fractional process
ARMA process, **644–647**, 650, 653, 677–679, 697, 710, 818, 1078
Asian option, 853–856
Asset liability management, *see* ALM
Asset-backed security, *see* ABS
Attachment point (CDO), **155**, 158, 186, 234–235
Augmented Dickey-Fuller test (ADF), 661–662
Available-for-sale (AFS), 373
Average daily volume, *see* ADV

B

BA-CVA, **285–288**, 289
Backtesting, 46, **49–52**, 61, 102
Bagging, 1001–1003
Balance sheet, 23, 158, 345, 367, **370–373**, 376, 377, 381, 385–393, 395, 401–403, 409–413, 420–421, 458, 477, 484, 894
Banking book, 22, **38**, 311, 369, 373, 393, 402, 486
Barrier option, **531–532**, 541–546, 559, 589, 813, 833
Base correlation (CDO), 238
Basel Accords
- · Basel I, **13–17**, 38–52, 160–162
- · Basel II, **17–19**, 38–52, 162–181, 266–270, 307–311, 898
- · Basel II.5 (or 2.5), **20**, 48–49, 69, 898
- · Basel III, **20–22**, 52–61, 64, 181–189, 270–274, 283–289, 311–312, 360–367, 376, 393–404, 462

Basel Committee on Banking Supervision, *see* BCBS
Basic indicator approach, *see* BIA
Basic own funds, 24
Basis risk, 40, 41, 394
Basket default swap, **151–155**, 157
Basket option, **583–592**, 602, 811–813
BCBS, 10, 12, 460
Best estimate, 24
Best execution, 25
Best linear unbiased estimator (BLUE), 607
Best unbiased estimator (BUE), 617
Beta, 49, 77, 454, 467
BIA, 18, **308**
Bid-ask spread, **348–349**, 352
Binary option, 541–545
Black model, 521, 570, **594**
Black-Scholes model, 44, 73, 94, 102, 121, **491–513**, 530, 561, 583, 593, 751, 811, 842, 867
Board of Governors of the Federal Reserve System, *see* FRB
Bond
- · Clean price, 132

· Defaultable, 63, 91, 134, 147
· Dirty price, 132
· Market size, 128–129
· Portfolio, 77
· Pricing, **131–136**, 514–516, 522, 593, 596, 601–602
· Zero-coupon, 63–64, 77, 394, 514–516, 575, 593, 594, 597, 601
Boosting, **1004–1008**, 1029
Bootstrap method (calibration), 205, 293
Box-Muller algorithm, 793, 886
Brownian bridge, **829–833**, 843, 891
Brownian motion, 219, 275, 301, 818, 833, **1065**, 1073–1074, 1078, 1079, 1082
Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS), 981, **1048**
Bubble risk, 37, 126, 456

C
CAD, 13
Calendar spread arbitrage, 556
Calibration
· Default correlation, 227–234, 238
· Implied volatility, 526–528
· Loss given default, 193–198, 200–201
· Probability of default, 151, **204–205**, 249
· Scenario analysis, 340
· Stochastic volatility, 554–557, 568
· Yield curve, 516–518
Call, *see* Option
Capital adequacy directive, *see* CAD
Capital allocation, **104–115**, 294–300
Capital asset pricing model, *see* CAPM
Capital charge, 46, 177, 266, 314, 327, 890
Capital requirements directive, *see* CRD
Capital requirements regulation, *see* CRR
Capital-at-risk, 314, 890
Caplet, **518–519**, 521, 526–528, 577–580, 601
CAPM, 49, **76**, 119, 355–357, 454–455, 610
Cash flow hedge (CFH), 374
CCF, 160–163, 179, 183, **190–191**, 270
CCP, 5, 25, 258, 459, 893
CDO, 137, **155–158**, 185–189, 234–235
CDS, 5, **141–151**, 205–206, 247, 249, 286
Central counterparty clearing house, *see* CCP
CET1, 20, 26, 464–466
CFTC, 12, 26, 461, 476, 893
Change of numéraire, **519–522**, 529, 594–598
Chapman-Kolmogorov equation
· Backward equation, 1071
· Forward equation, 207, 546, 878, 1072, 1081, 1083
Characteristic function, 331
Chebyshev's inequality, 89
Chicago Board of Options, 3
Chicago Board of Trade, 3
Chief financial officer (CFO), 45
Chief risk officer (CRO), 45
Cholesky decomposition, 223–224, 602, 646, 650, 736, 802–803, 833, 848, 890
Classification tree, 1002
Cliquet option, *see* Ratchet option
Closure property under affine transformation, 87, 1059
Clustering, 944–950
CMBS, *see* MBS
Coefficient of determination, 606–607, 610, **611–612**, 691
Coherent risk measure, *see* Risk measure
Cointegration, 420–421, **655–663**, 712
Collateral, 125, 137–141, 165–167, 183, 185, 186, 271, **293–300**, 303–304, 359, 365, 457, 487, 925
Collateralized bond obligation (CBO), *see* CDO
Collateralized debt obligation, *see* CDO
Collateralized loan obligation (CLO), *see* CDO
Collective risk model, 1
Commercial mortgage-backed securities, *see* MBS
Commodity Futures Trading Commission, *see* CFTC
Commodity risk, 9, 37, 41–42, 55, 117, 272, 899
Common equity tier 1, *see* CET1
Comonotonicity, 471, **722–724**, 784
Compound distribution, **313–315**, 327, 331–336, 825
Comprehensive risk measure (CRM), 20, 49
Concentration risk, 20, 105, 171, 181, **241–246**, 253–255, 1016
Concordance measure, **724–728**, 743
Concordance ordering, 590, **719–722**, 754
Conditional expectation, 109, 111, 171, 192, 276–277, 469, 473, 603, 608, 641,

691, 708, 852, 860, 863, 909–912, 920, **1062–1063**
Conditional prepayment rate (CPR), 400, 446–447
Conditional probability distribution, 277, 318, 608, 737, 738, 799–802, 885, 889–891, 904, 918, 921, 1062
Conditional value-at-risk, *see* CoVaR
Confusion matrix, 1019
Conjugate gradient method, 981, **1048**
Conservation buffer (CB), 20, 465
Constant balance sheet, **386**, 393, 396, 898
Constant correlation matrix, 221, 588–589, 812
Consumer credit, 126, 138, 923, 926
Control variates, 850–856
Convolution, 314, 332, 677
Cooke ratio, 13, 16, 18, **160**
Copula, **715–752**
· Associativity, 733
· Bounds, 718
· Correlation model, 113, 220–225, 275, 290–291, 338, 589–591, **715–752**, 907, 912
· Density function, **716**, 733, 735, 738
· Dependogram, 236, 237, **741–743**, 746, 788
· Empirical process, 741, 806
· Extreme value, **778–779**, 781, 783, 784, 907
· Frailty, 735, 803
· Function
Ali-Mikhail-Haq, 889
Archimedean, **732–735**, 801, 889
Clayton, 113, 731, 733, 735, 801, 803, 877
Cubic, 720
Farlie-Gumbel-Morgenstern, 725, 748, 778
Frank, 720, 733, 746, 889
Fréchet, 590, **718**, 728, 733, 762, 846, 847, 907
Galambos, 781
Gumbel logistic, 716, 732, 748
Gumbel-Barnett, 750
Gumbel-Hougaard, 725, 731, 733–735, 744, 762, 778, 779, 781, 784, 785, 889, 907
Hüsler-Reiss, 781
Joe, 733
Marshall-Olkin, 781, 785
Normal, 220–225, 247, 249, 275, 291, 590, 717, **735–737**, 747, 785, 803, 890, 913
Pareto, 751
Product, **716**, 720, 722, 733, 741, 761, 762, 802, 907
Student's t, **737–741**, 747, 803, 914
· Multivariate copula, 719, 734–735
· Negative quadrant dependence, **720**, 724
· Option pricing, 589–591
· Parametric family, 720, **731–741**
· Positive quadrant dependence, **720**, 724, 780, 781
· Scale invariance property, 723
· Simulation, 223, 290–291, 800–808, 889–891
· Sklar's theorem, 715–716
· Survival copula, 220, 749, 761
Cornish-Fisher expansion, **85–86**, 100, 121–122
Correlation matrix, 117, 222, 296, 303, 525, 609, 735, 747, 809–813, 833, 952, 1037, 1065
Correlation risk, 94, 394, 456, **583–591**, 811
Corridor option, 545
Cost-of-carry, 41, 94, 260, 492, **497**, 537, 583
Countercyclical capital buffer (CCB), 20, 462
Countermonotonicity, 722–724, 846–847
Counterparty credit risk, 9, 20, **257–278**, 458, 893
CoVaR, 469–471
Covariance matrix, 75, 106, 117, 119, 605, 618–619, 621, 635, 645, 649, 686–687, 698, 807–809, 948, 950, 1021, 1024, 1036, 1060
Cox-Ingersoll-Ross process, 562, 708, 823, **1076–1077**, 1083
CRD, 19, 20, 22
Credit conversion factor, *see* CCF
Credit default index (CDX), **153–155**, 158, 287
Credit default swap, *see* CDS
Credit default tranche (CDT), 158
Credit derivative, **137–158**, 162, 167, 271, 272
Credit event, 141, 144, 151
Credit portfolio management, 2

Credit risk, 9, 14, 18, 79, **125–255**, 257, 357, 360, 403, 478, 893, 902
Credit risk mitigation, **165–167**, 183
Credit scoring, 2, 125, **923–1030**
Credit spread, 53, **135**, 142, 147, 154, 226, 286, 292, 403
Credit triangle, 144, 283
Credit valuation adjustment, *see* CVA
Cross-validation, 938–941, 972, 1022
Crossover (XO), 155, 158
CRR, 22
Currency risk, 3, 9, 37, 40–41, 55, 58, 402
Current exposure (CE), 264
Current exposure method (CEM), 269
Curvature risk, 54, 288
CVA, 21, **278–293**, 294, 302, 894

D
Davidon-Fletcher-Powell algorithm (DFP), 981, **1048**
Debit valuation adjustment, *see* DVA
Deep learning, 978
Default correlation, 152, 158, 173, 178, **220–238**, 242, 247, 462
Default risk, 2, 91, 134, 141, **159**, 252, 283
Default time, 64, 92, 134–136, 142, 150, 152, 169, **201–220**, 247, 249, 251, 258, 275, 280, 302, 303, 750, 762, 792, 798, 921
Delta, *see* Option
Delta hedging, 98, **498–502**, 513, 575
Delta method, 43, 53, 120
Dendrogram, 946–950
Density estimation, *see* Kernel method
Dependogram, 236, 237, **741–743**, 746, 788
Detachment point (CDO), **156**, 158, 186, 234–235
Dickey-Fuller test, 661
Diffusion bridge, *see* Brownian bridge
Diffusion process, 439, 495, 598, 707, **1068**, 1070, 1074–1077
Digital option, *see* Binary option
Discrete Fourier transform (DFT), 683
Discriminant analysis, 958–968
Diversification, 2, 46, 62, 105, 106, 242, 297, 337, 456
Dividend, 497, 581
Dodd-Frank, 12, **26**, 181, 461, 893, 896, 899
Downgrading risk, 159
Dual problem, 992, 997, 1026–1028, **1046**
Dupire model, *see* Local volatility model
Duration, 40, 132, 271, 389, **405–407**, 427, 450–452
Duration gap analysis, 405–410
DVA, 279, 302
Dynamic balance sheet, **386**, 393, 396, 412, 898
Dynamic hedging, 495–506
Dynamic programming (DP), 929, **1049–1051**

E
EAD, 14, 57, 160, 168, 179, **190–191**, 255, 258, 260, 266, 270, 275, 293, 300, 302, 371
EBA, 12, 461, 893, 899
ECB, 12, 461, 893, 899
Economic value, 24, 393, **394–395**, 402, 403, 406, 408
Economic value of equity, *see* EVE
EE, **264**, 265, 301, 302
EEE, **265**, 266, 301
EEPE, **265**, 266, 268, 270, 276, 301, 302
Effective expected exposure, *see* EEE
Effective expected positive exposure, *see* EEPE
Eigendecomposition, 79, 224, 253, 686, 802, 809, 951, 967, **1033–1034**
EIOPA, 12, 461
Elastic net regression, 1021–1022
EMIR, 25, 462
Empirical VaR, *see* VaR
EnE, 281
Engle-Granger cointegration method, 660
EPE, **264**, 265, 268, 301
EpE, **281**, 290, 294–300, 302–304
Equity risk, 9, 37, 40, 55, 60, 272, 479, 899
Equity tranche, **155–158**, 185, 187, 188, 238
Equivalent martingale measure, **519–522**, 593, 597–598
Error correction model, 420–421, **659–660**
ES
- Analytical (or parametric) ES, 66, **74**, 118
- Credit portfolio, 249
- Definition, 61, 63, **65**, 118, 467
- Gaussian ES, **74**, 123, 467
- Historical (or non-parametric) ES, 66, **70–71**, 96, 114, 123, 124
- Kernel method, 71, 124
- Monte Carlo ES, 66, **90–92**

· Peak over threshold method, 775–777
· Regulatory capital, 22, **57**, 66
· Risk decomposition, 107, 109, 123, 249
ESFS, 12, 461
ESMA, 12, 25, 461, 893
ESRB, 12, 461, 899
Euler allocation principle, **105–108**, 123, 172, 246, 294–300, 304, 470
Euler-Maruyama scheme, 708, **820–824**, 835–838
European Banking Authority, *see* EBA
European Central Bank, *see* ECB
European Insurance and Occupational Pensions Authority, *see* EIOPA
European market infrastructure regulation, *see* EMIR
European Securities and Markets Authority, *see* ESMA
European System of Financial Supervision, *see* ESFS
European Systemic Risk Board, *see* ESRB
EVE, **395–397**, 403, 408–410
EVT, **753–785**, 803, 905
Excess kurtosis, *see* Kurtosis
Expectation-Maximization (EM) algorithm, 621–627
Expected exposure, *see* EE
Expected loss (EL), **62**, 168, 175, 178, 185, 243, 245–246, 251, 253, 310, 333
Expected positive exposure, *see* EPE
Expected shortfall, *see* ES
Exponential default time, 92, 135, 144, 150, 151, 197, **202**, 230, 247, 302, 303, 750, 921
Exponentially weighted moving average (EWMA), 80–82
Exposure at default, *see* EAD
Extreme order statistic, **755–757**, 784, 885, 906
Extreme value theory, *see* EVT

F

Factor model, 75–81, 117, 173, 225–229, 291, 454, 525, 575, 610
Fair value hedge (FVH), 374
Fair value through other comprehensive income (FVOCI), 374
Fair value through profit and loss (FVTPL), 373
Fannie Mae, 139
Fast Fourier transform (FFT), 684
Faure sequence, 882
Feature extraction, 950
Federal Insurance Office, *see* FIO
Feynman-Kac formula, 492, 514, 584, 594, **1070–1071**, 1081
FICO score, 923, 925
Financial sector assessment program (FSAP), 893, 895, 898
Financial Stability Board, *see* FSB
Financial Stability Oversight Council, *see* FSOC
Finite difference method, 531, 557, 579, 582, 1038, **1041–1045**
FIO, 12
First-to-default swap, **151–153**, 157, 247, 762
Fisher-Tippet theorem, **763–765**, 778, 785, 919
Fixed-strike parametrization, 537
Flash crash, 459, 474
Fletcher-Reeves algorithm, 1048
Floating-strike parametrization, 538
Floorlet, **519**, 526
Fokker-Planck equation, *see* Chapman-Kolmogorov equation
Foreign exchange risk, *see* Currency risk
Forward probability measure, **520**, 523, 526, 577–578, 594, 596–598, 601
Forward rate agreement (FRA), 419, 420, 516
Foundation internal ratings-based approach, *see* IRB
Fractional process, **694–700**, 714
Fractional white noise, *see* Fractional process
FRB, 12, 461, 476
Fréchet bounds, 718
Fréchet class, 717–722
Freddie Mac, 139
Frequency domain maximum likelihood (FDML), 687
FSB, 12, 26, 27, **453**, 460, 463–466, 478–485, 893
FSOC, 12, 26, 461
Full pricing, **94–96**, 101
Funding gap, 376
Funding liquidity, 9, 347, **357–359**, 375, 447, 459, 480
Funding ratio, 376
Funding risk, 360, 376, 447, 484, 893

Funds transfer pricing, 421–427
Futures contract, 3, 12, 26, 94, 474

G
G-SIB, 26, 460, **463–466**
G-SII, 26, 460, **466**
Gamma function, 694, 1052, 1060
Gap risk, 40, **393**, 410
GARCH process, *see* ARCH process
Gauss-Markov theorem, 607
Gaussian quadrature, 135, 150, 578, 584, 841, 910, **1038–1041**
General market risk, 38–40, 43, 44, 53, 116
Generalized eigendecomposition, 967, 1034
Generalized method of moments (GMM), 316, 317, 319, 343, 429, **631–634**, 708
Geometric Brownian motion, 73, 214, 216, 261, 282, 303, 491, 495, 560, 582, 593, 594, 707, 708, 819, 822, 823, 830, 832–835, 848, **1074–1075**, 1077, 1081
Gibbs sampler, 354, 871–875
Gini coefficient, 1016
Ginnie Mae, 139, 437
Girsanov theorem, 492, 520, 560, 594–596, **1072**, 1081, 1083
Glivenko-Cantelli theorem, 806
Global Financial Crisis (GFC), 26, 37, 234, 359, 453, 588, 898
Granger representation theorem, 660, 712
Granularity, 171, 180, 187, **241–246**, 253–255

H
Halton sequence, 882
Hammersley sequence, 884
Hazard function, **201–205**, 209, 219, 226, 249–251, 292, 386, 428, 442–446, 449, 452, 762
Heavy-tailed distribution, 84, 91, 328, 334, 666
Hedging, 2, 103–104, 120, 147, 284–288, 374, 392, 403, 419, 495–502, 557
Hedging portfolio, 3, 103, 261, 288, 434, 491, 542, 561
Hedging set, 270–274
Held-for-trading (HFT), 374
Held-to-maturity (HTM), 159, 373
Herfindahl index, 187, 389
Heston model, **562–568**, 835–837, 1077, 1083
Hierarchical clustering, **946–950**
High yield (HY), 53, 138, 154, 158, 273
Higher loss absorbency (HLA), 26, 464
Historical probability measure, 261, 283
Historical scenario, 48, 66–70, 95, 123, 396, 894–900
Historical VaR, *see* VaR
Hitting time, *see* Stopping time
HJM model, **522–525**, 596
Hurst exponent, 699
Hypothetical scenario, 48, 396, **894–896**, 903–904

I
IAIS, 12, 22, 27, 460, 466
IAS, 373–374
ICAAP, 181, 241
Idiosyncratic risk, 9–10, 60, 76, 117, 119, 170, 248, 286, 291, 304, 404, 454, 455, 462
IFM estimator, 746–748
IFRS, 278–279, 373–375
IGARCH process, 82, 665
IMA, 15, 38, **45–52**, 52, **57–61**, 70, 104, 116
IMM, 22, **268–269**, 270, 286
Implied correlation, 234, **586**, 588–589, 591
Implied volatility, 43, 55, 94–101, 120–121, 261, **508–513**, 526–528, 530, 540, 541, 543–558, 569, 598–600, 902
Importance sampling, 111, **856–861**, 879, 891
Income statement, **370–373**
Incomplete gamma function, 302, 1052
Incremental risk charge (IRC), 20, 49
Infinitely fine-grained portfolio, *see* Granularity
Insurance, 1, 22–25, 329, 460, 466, 473, 480, 481
Integral transform, 290, **790**, 913
Interest rate risk, 2, 9, 37, 39–40, 79–80, 116, 126, 404–427, 513–529
interest rate risk in the banking book, *see* IRRBB
Interest rate swap, 4, 266, 303, 425, 434
Internal capital adequacy assessment process, *see* ICAAP
Internal model method, *see* IMM
Internal model-based approach, *see* IMA

Internal rating system, 12, 169, 179
Internal ratings-based approach, *see* IRB
International accounting standards, *see* IAS
International Association of Insurance Supervisors, *see* IAIS
International financial reporting standards, *see* IFRS
International Organization of Securities Commissions, *see* IOSCO
Investment grade (IG), 39, 53, 146, 154, 158, 182, 273
IOSCO, 12, 27, 186, 460, 486
IRB
· AIRB, 18, 169, 177, **179**, 184, 185
· Counterparty credit risk, 268–269, 302
· Definition, 18, **168–181**, 184–185
· FIRB, 18, 169, 177, **179**, 184, 185
· Securitization, 186–188
IRRBB, 22, **393–402**
Itô calculus
· Chasles decomposition, 1066, 1079
· Integration by parts, 1066, 1070
· Isometry, 1066, 1079
· Linearity property, 1066, 1079
· Martingale, 1067
· Stochastic integral, **1066–1067**, 1079
· Stochastic process, 1067
· Stochastic Taylor expansion, 495, 560, 598, 599, 821, 823, 836, **1067–1070**, 1074–1076, 1080

J
Johansen cointegration method, 663
Jump-diffusion process, 827–828
Jump-to-default (JTD), 53, 56, **226–229**, 288, 404

K
K-means clustering, **944–946**, 1023
k-nearest neighbor (k-NN), **974–975**
Kalman filter, 647, **647–648**, 649–656, 667–670, 709–711
Kendall's tau, 717, **724–728**, 733, 736, 738, 743–747, 749, 779
Kernel method
· Classification, 997–1000
· Density estimation, 71–72, 84, **637–640**
· Non-parametric regression, 112, **641–645**
· Quantile estimation, 71–72, 124
Kolmogorov-Smirnov test, 321, 692, **1015**
KPSS test, 662
k^{th}-to-default swap, 151–153, 157, 247
Kurtosis, **84**, 86, 87, 122, 236, 600, 666, 1054, 1055, 1061, 1083

L
L-CAPM, 355–357
Lack of memory property, **202**, 1054
LAD regression, **613**, 614, 997
Lagrange function, 626, 706, 864, 934, 967, 992, 997, 1027, 1046, 1049
Lamperti transform, 824
Laplace transform, 332, 734, 735, 803, 805
Lasso regularization, **934–938**, 972, 1021–1022
LCR, 21, **360–365**, 376, 462
LDA, **312–342**, 788
Least absolute deviation, *see* LAD regression
Least squares, 203, 541, **604–608**, 612, 642, 650, 705, 957, 996, 1021, 1026, 1030
Legal risk, 305, 307
Lehman Brothers collapse, 10, 19, 21, 48, 146, 257, 453, 458, 899
Leverage ratio, 20–22, 214, 360, **367**, 408, 457, 459, 472–474, 926
Leverage risk, 52, 360, 367, 456–458
LGD, 56, 168, 172, 177, 179, 185, **191–201**, 242, 248, 250, 268, 284, 618, 629, 749, 750, 805, 894, 910
Libor market model, 526, 597–598
Libor scandal, 307
Linear congruential generator (LCG), 788
Linear correlation, *see* Correlation matrix or Pearson's correlation
Linear discriminant analysis (LDA), 961–966
Linear regression, 111, 232, 572, **603–614**, 621, 632, 654, 657, 662, 691, 699, 704, 705, 707, 852–855, 909–910, 921, 936–941, 951, 974, 997, 1022, 1025, 1063
Liquidation ratio, 350–352
Liquidity coverage ratio, *see* LCR
Liquidity gap, **376–393**, 412–415
Liquidity risk, 2, 9, 20, 21, 40, 49, 58, 103, 129, 141, 146, **347–367**, 376–393, 459, 478, 591–592, 893
Lloyd's algorithm, 945, 1023

Loan, 125–127, 449
Loan equivalent exposure (LEE), 268, 276
Loan-to-value ratio (LTV), 22, 181–183
Local correlation model, 236–238
Local polynomial regression, 642–643
Local volatility model, 103, **546–559**, 599
Logistic regression, *see* Logit model
Logit model, 200, 449, 910, 925, 941, **971–973**, 987–989, 1003, 1005, 1006, 1008, 1029
Look-back option, 842
Lorenz curve, 1016–1018
Loss data collection exercise, 307, 310
Loss distribution approach, *see* LDA
Loss frequency distribution, 313, **321–327**, 343–344
Loss given default, *see* LGD
Loss severity distribution, 313, **315–321**, 342, 345
Loss threshold, **317–321**, 326, 327, 343, 346
Low discrepancy sequence, 882
LTCM, 257, 457, 458, 474

M
MA process, *see* ARMA process
Macro-prudential regulation, 11, 20, 21, 461, 898
Mahalanobis distance, 947
Margin call, 457, 459
Margin period of risk, *see* MPOR
Marginal expected shortfall, *see* MES
Marginal risk, **106–107**, 248, 467, 468
Mark-to-market, 38, 42, 50, 64, 102, 142–151, 159, 257–265, 271, 275–282, 293–300, 302, 303, 425, 451
Mark-to-model, 38, **102–104**, 260, 491
Market discipline, *see* Pillars
Market liquidity, 9, 147, **347–357**, 360, 404, 459, 895
Market risk, 7, 13–22, 24, **37–124**, 136, 257, 359, 360, 455, 893–895
Markets in financial instruments directive, *see* MiFID
Markets in financial instruments regulation, *see* MiFIR
Markit
· CDT, 158
· CDX, 154, 158
· iTraxx, 154, 158
Markov chain
· Continuous-time, 210–215, 251–252, 870
· Discrete-time, 206–210, 1009
· Simulation, 814–817
· Stationary distribution, 208, 871, 874, 1009
Markov chain Monte Carlo (MCMC), 670, **870–882**
Markov generator, 210–213, 226, 251–253
Markov property, 206, 878, **1064**, 1073, 1078
Martingale, 520, 521, 523, 524, 526, 529, 537, 593–598, **1064**, 1065, 1067, 1072, 1078, 1080
Martingale representation theorem, 1067
Matrix exponential, 211, 1035, 1084
Matrix logarithm, 212, 1035, 1084
Maximum domain of attraction (MDA), 765–769, 781–782, 785
Maximum likelihood estimation, 193, 232, 250, 316, 319, 323, 342–346, 429, **614–627**, 649, 653, 654, 662–663, 666, 686–688, 706–708, 745–748, 758, 907, 933, 969–973, 980, 1025, 1029
Maximum of Brownian motion, 218, 1074, 1082
Maximum peak exposure (MPE), 264, 301
MBS, 5, 129, **137–140**, 154, 361, 437
MCR, 24
Median regression, *see* LAD regression
MES, 467–468, 472–474
Method of moments, 85, 193, 195, 250, 316, 323, 324, 341, 344, **628–631**, 707, 743, 749, 750
Metropolis-Hastings algorithm, 873–877
Mezzanine tranche, **155–157**, 185, 187, 188
Micro-prudential regulation, 11, 20, 461, 462
MiFID, 25, 462
Milstein scheme, **821–824**, 835–838
Min-stable multivariate exponential distribution, 779
Minimum capital requirement, *see* Pillars
Minimum capital requirement (insurance), *see* MCR
Minkowski distance, 948
Mixture distribution, 197, 236, 540, **624–627**, 707, 798, 803, 877
MMF, 10, 27, 458, 478, 486

Model risk, 88, 102–104, 120, **491–602**
Money market fund, *see* MMF
Monte Carlo simulation, **787–891**
· ALM risk, 420–421
· CCR, 260–263, 265
· Credit risk, 223–225, 245
· CVA, 282, 290–291
· Inversion method, 790–792, 885
· Market risk, 90–92
· Method of conditional distributions, 799–802
· Method of mixtures, 798
· Numerical integration, 787, 838–844
· Operational risk, 327–331
· Random matrix, 807–813
· Rejection sampling, **794–797**, 875, 887–888
· Scenario generation, 261
· Stochastic process, 813–838
· Transformation method, 793–794, 802–808, 886
· Variance reduction, 844–870
Moore-Penrose pseudo-inverse, 711, 1033, 1048
Moral hazard, 454, 478
Mortgage, 125, 200, 377–384, 391, 437–441, 443, 445, 446, 450–452
Mortgage-backed security, *see* MBS
MPOR, 272, **293**, 294
Multiple recursive generator, 789
Multiplication factor, **46**, 50, 60, 88–90, 119
Multivariate probability distribution
· Normal distribution, 75–78, 108, 221, 224, 249, 284, 467, 469, 608, 620, 624, 735, 802, 960, **1054**, 1062, 1065
· Skew normal distribution, 87–88, 90, 91, **1057–1058**
· Skew t distribution, 88, **1059**
· Student's t distribution, 737, 803, **1055**
· Wishart distribution, 813, **1060**

N

Nadaraya-Watson regression, 112, **641–644**
NAIC, 23
National Association of Insurance Commissioners, *see* NAIC
NBNI SIFI, 26, 27, 460, **466**, 478
Nelson-Siegel model, **131**, 133, 150, 205, 398, 402, 517–518
Net interest income, *see* NII
Net interest margin, 414, 422–424
Net interest spread, 423, 424
Net investment hedge (NIH), 374
Net stable funding ratio, *see* NSFR
Netting agreement, 259, 265, 269, 300
Network risk, 455, 458–459, 462–463, **474–477**
Neural network, 973, **975–989**, 998, 1025–1026, 1029
Newton-Raphson algorithm, 790, 805, 981, 1025, **1047**
NII, 373, 393, 395, 411, **412–418**, 419, 422
Non-linear optimization, 1047–1049
Non-maturity deposit (NMD), 366, 371, 394, 398, 401, **427–437**, 451
Non-negative matrix factorization (NMF), 954–959
Non-parametric estimation, 66, 71, 553, **637–645**, 724, 741–743, 746, 776, 974–1008
Non-parametric VaR, *see* VaR
NSFR, 21, **365–367**, 376, 462
Numerical integration, 86, 135, 578, 584, 787, 825, 838, 910, **1037–1041**

O

Off-balance sheet item, 14, 21, 22, 37, 160, 161, 163, 165, 179, 183, **190–191**, 358, 365, 367, 371, 393, 397
Omnibus estimator, 746–748
On-balance sheet item, 21, 37, 161, **190–191**, 367, 384, 388, 421
Operational risk, 9, 17–18, 22, 24, 103, 104, **305–346**, 636, 788, 805, 885, 890, 898, 909
Optimization, **1046–1051**
Option
· American, 437, 439
· Asian, 853–856
· At-the-money (ATM), **493**, 499, 506, 510, 513, 551, 569–574, 591, 600
· Barrier, **531–532**, 541–546, 559, 589, 813, 833
· Best-of, 585
· Binary, 541–545
· Call, 93–101, 119, 121, 214, 260, **492–494**, 497, 502, 562, 581–583, 594, 598–600, 854, 856, 867, 869
· Cega, 585–586

- · Delta, 43, 53, **96–100**, 121, 498–506, 531, 557, 561, 573–575, 582, 592, 898
- · Gamma, 43–44, **98–100**, 121, 504–506, 512, 530–531, 558
- · In-the-money (ITM), **493**, 501, 504, 508
- · Intrinsic value, 493
- · Market risk, 42–45, **92–104**
- · Out-of-the-money (OTM), **493**, 501, 504, 508
- · Pricing, **491–602**
- · Put, 103, 120, **492–494**, 497, 502, 562, 582, 583, 856, 859, 1082
- · Put-call parity, **493**, 856
- · Ratchet, 532–536
- · Theta, **99–100**, 121, 504–506
- · Time value, 493
- · Vanna, 99, 104
- · Vega, 43–44, 53, 99–100, 121, 543, 544, **558**, 561, 574–575, 898
- · Volga, 99
- · Worst-of, 585

Order statistic, 67–71, 91, 101, 109, 111, 113, 153, 265, 321, 328, 640, **753–763**, 783, 784, 885, 906
Ordinary differential equation (ODE), 600
Ordinary least squares (OLS), 230, **604–605**, 613, 652, 657, 704–706, 708, 709, 937, 997, 1023
Ornstein-Uhlenbeck process, 429–431, 514, 562, 575, 593, 601, 707, 708, 819, 823, **1075–1076**, 1081
ORSA, 25
OTC, **5–9**, 21, 25, 92, 102, 141, 257–260, 278, 300, 462, 485, 491
Over-the-counter market, *see* OTC
Own funds, 24
Own risk and solvency assessment, *see* ORSA

P

Packaged retail and insurance-based investment products, *see* PRIIPS
Panjer recursion, 332–334
Parametric VaR, *see* VaR
Partial differential equation (PDE), 492, 496–497, 504, 514, 530–531, 534–535, 548–549, 558–560, 562, 576, 578, 579, 584–587, 593–594, 601, 1041–1045, 1070–1072, 1081
Particle filter, 878–882
Passporting, 25
PD, 125, 126, 136, 144, 162, 168–170, 173–181, **201–220**, 235, 253, 268, 276, 283, 284, 302, 357, 910, 912, 921, 924, 926, 942
Peak exposure (PE), 264–266, 301, 302
Peak over threshold, 340, 773–777
Pearson's correlation, 727–729
Penalized maximum likelihood, 341, 972
Periodogram, **683–686**, 687, 689, 692, 693, 712, 714
PFE, 264–266, 269–271, 274, 280, 301, 302
Piecewise exponential model, **202–205**, 209, 219, 249, 251, 253, 293, 792
Pillars
- · Market discipline (Pillar 3), 18, 20, 25, **181**
- · Minimum capital requirement (Pillar 1), **17–18**, 20, 25, 159, 241, 246
- · Supervisory review process (Pillar 2), 17, 20, 22, 25, **181**, 190, 241, 246, 249, 342, 393, 396, 893

Poisson process, 321–324, 337, 339, 340, 346, 793, 824–827
Polak-Ribiere algorithm, 1048
Potential future exposure, *see* PFE
Predicted residual error sum of squares (PRESS), 939–941, 1022
Prepayment risk, 383–384, 400–401, **437–447**, 452
Pricing function, 66
PRIIPS, 25, 462
Principal component analysis (PCA), 79, **950–955**, 958, 1024
Probability distribution
- · Bernoulli distribution, 50, 169, 193–197, 223, 236, 323, 389, 615–617, 791, 873, 970, 1051, **1051**, 1052, 1078
- · Beta distribution, 193–198, 250–251, 618, 652, 717, 754, 805, 873, 886, 887, **1053**
- · Binomial distribution, 50, 232, 793, **1051**
- · Chi-squared distribution, 620, 684, 692, 793, 929–931, 970, **1053–1054**
- · Erlang distribution, 339
- · Exponential distribution, 135, 151, 202, 219, 247, 322, 339, 345, 384,

387, 391, 417, 450, 707, 764, 767, 774, 782, 784–785, 790, 793, 805, 824, 877, 886, **1054**
· Fréchet distribution, **763**, 764–768, 770, 778
· Gamma distribution, 253, 315–317, 324, **1052**
· Generalized extreme value distribution, 316–317, **770–773**, 774, 775, 779, 885, 904–906, 919, 1056
· Generalized Pareto distribution, 773–777, **1057**
· Geometric distribution, 323, 346, 795, **1052**
· Gompertz distribution, 150, 202, 204
· Gumbel distribution, **763**, 764, 765, 767, 770, 778
· Gumbel logistic distribution, 716
· Log-gamma distribution, 315–317, 336
· Log-logistic distribution, 202, 316–317, 336, 345, 885
· Log-normal distribution, 202, 314, 316–317, 319–320, 327, 333, 336, 338, 340–343, 345, 522, 529, 535–545, 568–571, 573, 598, 636, 722–724, 728–729, 835, 854, 885, 920, **1055**, 1075, 1077
· Negative binomial distribution, 323–327, 346, 798, **1052**
· Noncentral chi-squared distribution, **1053–1054**, 1076, 1083
· Normal distribution, 67, 72–74, 84–85, 89–90, 118, 173, 232, 235, 248, 514, 608, 640, 666, 669, 754, 756, 767, 770, 790, 793, 795, 796, 887, 972, **1054**, 1062, 1067, 1076
· Pareto distribution, 118, 316–317, 336, 342, 344, 751–752, 764, 767, 785, 790, 919, **1056**
· Poisson distribution, 314, 321–328, 333, 340, 343–344, 706, 788, 793–794, 798, 826, **1052**
· Skew normal distribution, 87–88, **1057–1058**
· Skew t distribution, 87–88, **1059–1060**
· Student's t distribution, 84–85, 89, 113, 640, 661, 756, 758, 793, 887–888, 914, **1055**
· Uniform distribution, 118, 193, 195, 250, 251, 290, 435, 638, 716, 743, 747, 754, 765, 768, 774, 779, 782–783, 785, 787–790, 793, 839, 1009
· Weibull distribution, 202, 762, **763**, 765, 768, 770, 778
Probability of default, *see* PD
Probit model, 200, 925, **972**, 1025
Procyclicality, 20, 21, 177
Pseudorandom numbers, *see* Random number generation
Put, *see* Option

Q
QQ plot, 321, 692
Quadratic discriminant analysis (QDA), 961–966
Quadratic Gaussian model, **575–580**, 601–602
Quadratic programming (QP), 705–706, 934, 990–997, 999, 1022, 1026–1028, **1046–1047**, 1049
Quantile, 16, 63, 65, 67, 71, 86, 242, 264–265, 300, 321, 328, 334, 431, 467, 590, 716, 731, 737–740, 771, 773, 784, 805–807, 912–916, 920, 924, 1062
Quantile regression, **613–614**, 643–645, 912–916, 920
Quantitative impact study (QIS), 176–177
Quasi-Monte Carlo simulation, 880–884
Quasi-Newton algorithm, **1048**, 1049

R
Random correlation matrix, 809–813
Random covariance matrix, 809
Random forests, **1003–1004**, 1008
Random number generation, **787–798**
Random orthogonal matrix, 807–809
Rank statistic, 69, 109, 465, 725, 741–746
Rare event simulation, 859
RAROC, 2, 105
Ratchet option, 532–536
Rating, 18, 39, 56, 158, 162, **163**, 168, 182, 188, 206–215, 251–253, 284, 361, 814–817, 923, 942
Recovery rate, 63, 91–93, 134–136, 142, 155, 168, **191–192**, 197, 200, 216–219, 247, 249, 250, 280, 284, 289, 302
Recursive least squares, 650–652

Redemption risk, 394, 400–401, **447–449**
Reflection principle, 218, 1073–1074
Regular variation, 767–768
Regularization technique, 111–113, 972
Regulatory arbitrage, 17, 27, 162
Rejection sampling, **794–797**, 875, 887–888
Repo, 9, 38, 54, 183, 459, 478, 480, 487
Reputational risk, 305
Residential mortgage-backed securities, *see* MBS
Residual risk, 53, 57, 288, 506
Resolution plan, 27, 462, 465
Retail exposure, 180–185
Return period, 208, 771, 784, 905–909, 920
Reverse stress testing, **916–919**
Revolving credit, 126, 165, 180, 190
Ridge regression, **939–941**, 972, 995, 997, 1021–1023
Right way risk, 289
Risk allocation, *see* Capital allocation
Risk contribution, **104–115**, 122–123, 168, 172, 175–177, 241, 246, 248, 295–297, 303–304, 468–469, 472
Risk factor, 2, 45, 46, 53–60, 65, 66, 75–81, 90, 92–101, 103, 136, 169, 172, 192, 222, 225, 232, 241, 266, 278, 288, 291, 455, 513, 561, 577, 894–896, 899–903, 921, 925
Risk management unit, 45, 305
Risk measure
· Coherence property, 61, 118, 470
· Convexity, 62, 107
· Definition, **61–64**, 118
· Homogeneity, 62, 107
· Monotonicity, 62
· Subadditivity, 62–64, 106, 118
· Translation invariance, 62–63
Risk weight, *see* RW/RWA
Risk-based capital, *see* RBC
Risk-neutral probability measure, 60, 261, 265, 280–284, 439, 492, 499, **508–511**, 514, 516, 519, 522–523, 529, 537, 546, 560, 562, 575, 581, 583, 589, 596, 602
Risk-weighted asset, *see* RW/RWA
RiskMetrics, 37, 80–81
Risky PV01, 143, 148–151, 247
RMBS, *see* MBS
Robust regression, 612–614
ROC curve, 1017–1020
Rogue trading, 9, 305, 307, 318, 898
Ruin theory, *see* Collective risk model
Run-off balance sheet, **385**, 387, 388, 393, 396, 412
RW/RWA, **14–15**, 18, 40, 53–57, 160–163, 166–167, 177–183, 186–189, 254, 268, 285–289, 361, 366, 486, 894

S
SA-CCR, **270–274**
SA-CR, **162–167**, **181–184**
SA-CVA, **288–289**
SA-TB, **53–57**
SABR model, 554, **568–575**, 590–592, 599, 1077, 1083
Scenario analysis, **339–341**, 894–897
Schur decomposition, **1034–1036**, 1037, 1084
SCR, 23–25
SEC, 12, 26, 461, 476
Securities and Exchange Commission, *see* SEC
Securitization, 5, 17, 20, 45, 57, **137–141**, 183, 185–189, 481, 485, 487
Senior tranche, **155–158**, 185–189
Sequential Monte Carlo, *see* Particle filter
Sequential quadratic programming (SQP), **1049**
Shadow banking, 27, 460, **478–487**
Shannon entropy, 1009–1011
Sherman-Morrison-Woodbury formula, 651
Shifted log-normal model, **535–546**, 598
SIFI, 12, 26–27, 460, **463–477**
Simulated method of moments (SMM), 635–637
Simulated VaR, *see* VaR Monte Carlo
Single loss approximation (SLA), 333–336
Single supervisory mechanism, *see* SSM
Skewness, 84–87, 122, 193, 236, 600, 707, 1054, 1061, 1083
Sklar's theorem, 589, **715–716**, 723, 778
Slowly varying function, *see* Regular variation
SM-CCR, 270
SM-CVA, 284–285
SMA, 22, 311–312
SMM, 15, **38–45**, 70, 116
Sobol sequence, 882–884
Softmax function, **973**, 981, 983, 1008, 1026
Solvency capital requirement (insurance), *see* SCR

Solvency I, 23
Solvency II, 22–24, 456
Solvency ratio (Insurance), 25
Spearman's rho, **724–728**, 736, 738, 744, 747–749
Special purpose vehicle (SPV), 138
Specific risk, 38–40, 48, 53, 116
Spectral analysis, **670–704**
Spectral density function, **672–673**, 677–682, 688–691, 696–698, 703, 713–714
Spread
· Bond, 135
· CDO, 157
· CDS, **143–151**, 205–206, 247, 249, 250, 281, 283, 292
· CDX, 154
· Credit curve, 146
· FtD, 152
· Risk, 37, 42, 53, 58, 60, 396, 403
Spread option, **585–587**, 590, 849, 884
Square-root random process, *see* Cox-Ingersoll-Ross process
Square-root-of-time rule, **46**, 73, 266
SREP, 181, 461
SRISK, 471–475
SSM, 461
Standard deviation-based risk measure, 62, 74, 106
Standardized approach (counterparty credit risk), *see* SA-CCR
Standardized approach (credit risk), *see* SA-CR
Standardized approach (credit valuation adjustment), *see* SA-CVA
Standardized approach (trading book), *see* SA-TB
Standardized measurement approach (operational risk), *see* SMA
Standardized measurement method (market risk), *see* SMM
Standardized method (counterparty credit risk), *see* SM-CCR
Standardized method (credit valuation adjustment), *see* SM-CVA
State space model, **647–656**, 661, 669, 679–681, 688, 709–711, 713, 815, 817, 818, 880
Static hedging, 542
Stochastic correlation model, 235
Stochastic differential equation (SDE), 492, 499, 522, 525, 530, 539, 546, 562, 596–598, 708, 819–824, 827, 835, 842, 849, **1067–1070**, 1074–1076, 1080–1083
Stochastic integral, *see* Itô integral
Stochastic volatility, 104, 554, **560–575**, 599–601, 667–670, 1077
Stopping time, 1073–1074
Strategic risk, 305
Stratified sampling, 862–870
Stress scenario, 60, 394–402, 455, 456, 472–473, 894, 897, **899–904**
Stress testing, 45, 47–48, 395, 462, **893–921**
Stressed value-at-risk (SVaR), *see* VaR
Strong Markov property, 1073, 1082
Subexponential distribution, 333–336
Super senior tranche, 155–158
Supervised learning, **958–1008**
Supervisory review and evaluation process, *see* SREP
Supervisory review process, *see* Pillars
Support vector machine, **989–1001**, 1026–1028, 1030
Survival function, 92, 135, 143, 151–152, **201–205**, 209, 212, 215, 218, 220, 225, 247, 249, 251–253, 275, 281–283, 304, 383, 386–392, 442–446, 448, 450, 452, 730, 749, 762, 779, 792
SVI model, 554–558
SVM regression, 996–997, 1028
Swap market model (SMM), 529
Swaption, 274, 519, 522, 529, 578
Systematic risk factor, 220, 253, 291, **454–455**, 468, 909
Systemic risk, 7–11, 26–27, 347, 359, **453–477**, 899, 1031
Systemic risk contribution, *see* SRISK
Systemically important financial institution, *see* SIFI

T

Tail dependence, **730–731**, 735, 737–740, 748, 749, 751, 783–785
Tail risk, 76, 456–458, 753
Technical provision, 23–25
Term deposit redemption ratio (TDRR), 400
The Federal Reserve Board of Governors, *see* FRB

The standardized approach (operational risk), *see* TSA
Time domain maximum likelihood (TDML), 687
TLAC, 462, 465–466
Tobit model, 708
Total loss absorbing capacity, *see* TLAC
Trade repository, 25, 462, 485
Trading book, 19, 22, **38**, 52, 288, 311, 373, 385, 894
Trading volume, 130, **349–350**, 352
Transition probability matrix, **206–212**, 252, 253
TSA, 18, **308–310**
Turnover, 7, 350, 353

U
UCITS, 25, 487
Uncertain volatility model (UVM), 530–536
Undertakings for collective investment in transferable securities, *see* UCITS
Underwriting risk, 24
Unexpected loss (UL), **178**, 185, 251, 310, 333, 470
Unit root, 82, 572, 655, 660, **661–662**, 665–666, 693
Unsupervised learning, **944–959**

V
Value-at-risk, *see* VaR
VaR
· Analytical (or parametric) VaR, 66, **72–90**, 90, 121
· Cornish-Fisher VaR, *see* Cornish-Fisher expansion
· Credit portfolio, 92–93, 169–176, 241–246, 248–249, 253–255
· Definition, 15, 46, 61–63, **64–65**, 118
· Gaussian VaR, 73, 123, 1037
· GEV method, 771–773
· Historical (or non-parametric) VaR, 66, **66–72**, 90, 110, 119–120, 124, 894
· Hybrid method, 101
· Kernel method, 71–72
· Monte Carlo VaR, 66, **90–92**
· Option portfolio, 94–102
· Quadratic VaR, 100, 121
· Risk decomposition, 107–114, 122
· Stressed VaR, 20, **48**, 69, 462
VAR process, *see* ARMA process
VARMA process, *see* ARMA process
Vasicek model, **514–518**, 524–525, 593
VECM, *see* Error correction model
VIX, 700, 894, 902
Volatility, 62, 80–82, 94–580, 664–670, 692, 760
Volatility index, *see* VIX
Volatility skew, **508**, 537–538, 543–545, 550, 558, 565, 570, 598
Volatility smile, **508**, 542–545, 554–556, 558, 561, 565–568, 570, 573, 579, 586, 601
Volcker rule, 26, 462

W
Weighted average life (WAL), 389
Weighted least squares (WLS), 541, 612–614, 1006
Whittle estimator, **686–689**, 698–700, 714
Wiener process, *see* Brownian motion
Wrong way risk, **275–278**, 289–293

X
Yield to maturity, **132–133**, 133, 135, 136, 147, 405–407, 409

Y
Zero-coupon rate, 77–79, 92, 131–133, 150, 406, **515–517**, 527

Author Index

A
Abramowitz, Milton, 669, 1041
Acemoglu, Daron, 476
Acerbi, Carlo, 63
Acharya, Viral V., 355, 356, 456, 458, 467, 471–473
Admati, Anat, 3
Adrian, Tobias, 469, 478
AFME, 137, 138
Agarwal, Sumit, 440
Ahn, Dong-Hyun, 576
Aldasoro, Iñaki, 403
Allen, David M., 939
Allen, Franklin, 458
Altman, Edward I., 197, 924, 925
Amemiya, Takeshi, 708, 969
Ames, Mark, 342
AMF, 893
Amihud, Yakov, 354
Andersen, Leif, 235
Anderson, Theodore Wilbur, 809
Arellano-Valle, Reinaldo B., 1057
Arrow, Kenneth J., 1
Artzner, Philippe, 61
Arulampalam, M. Sanjeev, 879
Ashcraft, Adam, 478
Augustin, Patrick, 144
Avellaneda, Marco, 103, 530
Azzalini, Adelchi, 1057–1060

B
Baesens, Bart, 197
BaFin, 893
Bahar, Reza, 234
Bailey, Ralph W., 887
Baillie, Richard T., 695
Banasik, John, 943
Bank of Japan, 428
Bartlett, Bruce, 574
Bastos, João A., 200
Baud, Nicolas, 318
Baxter, Marianne, 704
BCBS, 10, 13, 15, 17, 19–22, 26, 37–61, 64, 77, 104, 116, 117, 159–190, 241, 246, 253–255, 258, 266–274, 276, 278, 283–289, 305–312, 318, 336, 360–367, 385, 386, 393–404, 427, 431, 463–466, 486, 893–896
Beaver, William H., 924
Bellman, Richard, 1050
Bellotti, Tony, 197, 200
Bendel, Robert B., 810
Benoit, Sylvain, 456
Berestycki, Henri, 551
Bernstein, Peter L., 3
Bertsekas, Dimitri P., 1050
Bessis, Joël, 375, 408, 417, 421, 426
Bhatia, Mickey, 222
Billio, Monica, 476
BIS, 5–8, 126–128, 141
Bishop, Christopher M., 959, 968, 978, 979, 981, 984, 998
Black, Fisher, 2, 3, 73, 216, 491, 594
Blankenheim, Johannes, 141
Blöchlinger, Andreas, 428
Böcker, Klaus, 333
Boesky, Hayley, 478
Boggs, Paul T., 1049
Bohn, Jeff, 215, 216
Bollerslev, Tim, 81, 665
Boral, Andrew, 926
Borio, Claudio, 901
Boutsidis, Christos, 957
Bouveret, Antoine, 893
Bouyé, Eric, 747, 909
Box, George E.P., 793, 886
Brace, Alan, 526
Brady, Nicholas F., 4
Breeden, Douglas T., 508, 557
Brei, Michael, 21
Breiman, Leo, 1001–1003
Brennan, Michael J., 437, 575

Brigo, Damiano, 513, 518, 520, 521, 524, 529, 537, 540, 598
Britten-Jones, Mark, 100
Brooks, Robert, 448
Broto, Carmen, 670
Brown, Iain, 191, 197
Brown, Robert L., 652
Brownlees, Christian T., 471
Brunnermeier, Markus K., 21, 357, 358, 456, 459, 469
Bulirsch, Roland, 1041
Burtschell, Xavier, 236
Busca, Jérôme, 551
Busch, Ramona, 898

C

Caflisch, Russel E., 882
Calomiris, Charles W., 11
Canabarro, Eduardo, 258, 264, 268
Capitanio, Antonella, 1057–1060
Carlin, Bradley P., 880
Carlton, Dennis W., 3
Carroll, R.B., 111
Carty, Lea, 926
Casella, George, 871, 872
Cazalet, Zélia, 77
Cespedes, Juan Carlos Garcia, 278, 290
Çetin, Umut, 591
CGFS, 898
Chan, Kalok C., 708, 821
Chaudhary, Sharad, 438
Chen, Song Xi, 72
Cheng, Nick Ying-Pin, 46
Chernov, Mikhail, 437
Cherubini, Umberto, 590
Chervonenkis, Alexey, 990
Chib, Siddhartha, 670, 874, 875
Chou, Ray Yeutien, 81
Christoffersen, Peter, 234
Clapp, Tim, 879
Cline, Brandon N., 448
Coles, Stuart, 731, 770, 775
Colliard, Jean-Edouard, 456
Cont, Rama, 84, 293, 476
Cooley, Thomas, 458
Cortes, Corinna, 990
Cox, David R., 201, 969
Cox, John C., 216, 522
Cram, Donald P., 925
Crammer, Koby, 1001
Crépey, Stéphane, 554
Crockford, Neil G., 1
Crook, Jonathan, 197, 200, 925, 943
Crosbie, Peter, 215, 216
Crouhy, Michel, 70, 408

D

Dall'Aglio, Giorgio, 719
Dalla Valle, Alessandra, 1057
Daníelsson, Jón, 46
Davidson, Russell, 618
Davies, Philip I., 811
Davis, Philip J., 1039
De Bandt, Olivier, 455, 456, 458, 459, 895, 903
De Jong, Frank, 451
de Juan Herrero, Juan Antonio, 278, 290
Dees, Stéphane, 910
Deheuvels, Paul, 722, 741, 743, 777
Delbaen, Freddy, 61
DeMarzo, Peter, 141
Dembiermont, Christian, 126
Demey, Paul, 234, 384–387, 415, 417, 418, 425, 426, 428, 431
Dempster, Arthur P., 622
Denault, Michel, 105
Deng, Yongheng, 437
Dent, Kieran, 898
Derman, Emanuel, 102, 546, 599
Devenow, Andrea, 456
Devroye, Luc, 794, 829, 888
Diamond, Douglas W., 2, 458
Dickey, David A., 661
Diebold, Francis X., 46
Dittmar, Robert F., 576
Doucet, Arnaud, 879
Drehmann, Mathias, 126, 347, 357, 358, 365, 901
Drescher, Christian, 898
Driscoll, John C., 440
Duffie, Darrell, 92, 97, 141, 258, 264, 278, 579, 635
Dunbar, Nick, 896
Dunn, Brett R., 437
Dunn, Kenneth B., 437
Dupire, Bruno, 91, 546, 550
Durand, David, 923, 924
Durbin, James, 652
Durrleman, Valdo, 565, 747, 909
Dwyer, Douglas, 200
Dybvig, Philip H., 2, 458

E
EBA, 200, 278, 279, 385, 899, 916
Eber, Jean-Marc, 61
Eberlein, Ernst, 91
Edelman, David B., 925
Ehlers, Torsten, 403
El Karoui, Nicole, 511, 512, 520, 530, 575, 577, 601
Elderfield, Matthew, 23
Elie, Romuald, 438, 444
Embrechts, Paul, 763, 769
Emmer, Susanne, 246
Engle, Robert F., 81, 471–473, 634, 660, 664, 665, 690, 691
Epperlein, Eduardo, 112
Evans, James M., 652

F
Falkenstein, Eric G., 926
Fama, Eugene F., 77
Fawcett, Tom, 1019
FDIC, 370
Fengler, Matthias R., 554
FHFA, 446, 447
Filipović, Damir, 575
Finger, Christopher C., 222
Finkelstein, Vladimir, 216, 217
Fisher, Irving, 2
Fisher, Ronald Aylmer, 693, 958, 966, 969
Flannery, Brian P., 882, 1047, 1048
Florent, Igor, 551
Föllmer, Hans, 62
Forde, Martin, 565
Frachot, Antoine, 312, 318, 326, 337, 384–387, 415, 417, 418, 425, 426, 428, 431, 435, 438, 444
FRB, 127, 896, 899
French, Kenneth R., 77
Freund, Yoav, 1004
Friedman, Avner, 1064, 1065, 1067
Friedman, Jerome H., 933, 939, 943, 944, 968, 1003, 1006, 1008, 1022
Frishling, Volf, 582
FSB, 26, 27, 453, 460, 461, 463–466, 478–485, 893
Fuller, Wayne A., 661

G
Galai, Dan, 70, 408
Galambos, Janos, 202, 777
Gale, Douglas, 458
Gallagher, Russell B., 1
Gallant, A. Ronald, 576
Gallopoulos, Efstratios, 957
Gambacorta, Leonardo, 21
Gao, Feng, 448
Gasull, Armengol, 770
Gatarek, Dariusz, 526
Gatheral, Jim, 556, 565
Geanakoplos, John, 457
Gelfand, Alan E., 871
Geman, Donald, 871
Geman, Hélyette, 520
Geman, Stuart, 871
Genest, Christian, 715, 732, 733, 746
Gennotte, Gerard, 4
Genton, Marc G., 1057
George, Edward I., 871, 872
Georges, Pierre, 312, 438, 444, 761
Getmansky, Mila, 476
Getter, Darryl E., 19
Geweke, John, 698, 699
Ghoudi, Kilani, 746, 781
Gikhman, Iosif I., 1064, 1068
Gilkeson, James H., 447–449
Glasserman, Paul, 91, 111, 529, 830, 833, 882, 919
Golub, Gene H., 810, 1035, 1036, 1038
Goodhart, Charles, 177
Goodman, Allan S., 788
Gordon, Neil, 879
Gordy, Michael B., 171, 187, 232, 245, 246, 253
Gouriéroux, Christian, 71, 108, 246, 1011, 1013
Goyenko, Ruslan Y., 349, 354
Granger, Clive W.J., 657, 660, 694, 696
Gray, Robert M., 686
Greenberg, Edward, 874, 875
Greene, William H., 987
Gregory, Jon, 152, 236
Grinblatt, Mark, 456
Grossman, Sanford J., 347
Grundke, Peter, 919
Guill, Gene D., 2
Gupton, Greg M., 222

H
Hagan, Patrick S., 559, 568, 600
Hahn, Gerald J., 623
Hallerbach, Winfried G., 108
Hand, David J., 925, 942

Hansen, Lars Peter, 455, 631, 632
Hartmann, Philipp, 455, 456, 458, 459
Harvey, Andrew C., 647, 648, 654, 669, 670, 679, 681, 817
Hasbrouck, Joel, 353, 354
Hastie, Trevor, 933, 939, 943, 944, 968, 1003, 1006, 1022
Hastings, Keith W., 873
Haug, Espen G., 582
Haug, Jorgen, 582
Hayre, Lakhbir S., 438
He, Alex Xi, 448
He, Ping, 448
Heath, David, 61, 522, 525
Hedges, Bob A., 1
Hee, Jarrad, 190
Heffernan, Janet, 731
Heidelberger, Philip, 91
Heitfield, Erik, 232
Hellwig, Martin, 3
Henley, William E., 942
Henry, Jérôme, 910
Heston, Steven L., 562
Hickman, Andrew, 46
Higham, Nicholas J., 811, 1037
Hillegeist, Stephen A., 925
Ho, A., 111
Ho, Thomas S.Y., 524
Hodrick, Robert J., 704
Hoerl, Arthur E, 939
Hofmann, Boris, 177
Holden, Craig W., 349, 354
Horn, Roger A., 1033, 1036
Hosking, Jonathan R.M., 694, 695
Hsu, Chih-Wei, 1001
Huang, Ming, 278
Huang, Roger D., 354
Huber, Peter J., 612, 613
Hull, John, 278, 290, 292, 516, 524
Hurlin, Christophe, 456
Hutchison, David E., 432, 433

I

IAIS, 27, 466
Ieda, Akira, 170
IMF, 478, 487, 898
Ingersoll, Jonathan E., 522
Inoue, Atsushi, 46
IOSCO, 486, 487
Irving, Richard, 896
ISDA, 144, 274, 275
Israel, Robert B., 212

J

Jäckel, Peter, 563
Jacklin, Charles J., 4
Jacobs, Jr., Michael, 191
Jacobs, Kris, 234
Jacquier, Antoine, 556, 565
Jafry, Yusuf, 208
Jamshidian, Farshid, 520, 529
Janosi, Tibor, 451
Jarrow, Robert A., 210, 434, 451, 522, 525, 591
Jasiak, Joann, 1011
Jeanblanc, Monique, 511, 512, 530
Jickling, Mark, 12
Jin, Xisong, 234
Joe, Harry, 715, 730, 735, 746, 777
Johansen, Adam M., 879
Johansen, Søren, 663
Johnson, Charles R., 1033, 1036
Jolis, Maria, 770
Jolliffe, Ian T., 950, 951, 1024
Jondeau, Eric, 473
Jones, Bradley, 141
Jones, David, 187
Jones, E. Philip, 215
Jones, Mary Cover, 639
Jorion, Philippe, 1, 4, 9, 37, 457, 894
Jouanin, Jean-Frédéric, 224, 234
Joyeux, Roselyne, 694, 696
J.P. Morgan, 81

K

Kacperczyk, Marcin, 459
Kahl, Christian, 563
Kalkbrener, Michael, 106, 451
Kalotay, Egon A., 197
Kamal, Michael, 599
Kan, Rui, 579
Kang, Chulmin, 919
Kang, Wanmo, 919
Kani, Iraj, 546, 599
Karatzas, Ioannis, 1064, 1067
Karolyi, G. Andrew, 708, 821
Kashyap, Anil K., 177
Kavvathas, Dimitrios, 207, 208
Keating, Elizabeth K., 925
Keller, Ulrich, 91
Kennard, Robert W., 939
Keys, Benjamin J., 437, 438

Khoudraji, Abdelhaq, 781
Kim, Sangjoon, 670
King, Robert G., 704
Kitagawa, Genshiro, 880
Klüppelberg, Claudia, 333, 763, 769
Kleidon, Allan W., 4
Kloeden, Peter E., 821
Klugman, Stuart A., 312
Koenker, Roger, 614
Kopeliovich, Yaacov, 919
Korablev, Irina, 200
Kotz, Samuel, 202
Kumar, Deep, 559, 568, 600
Kupiec, Paul H., 894
Kurpiel, Adam, 1041
Kwiatkowski, Denis, 662
Kyrillos, Barbara B., 3

L
Laibson, David I., 440
Laird, Nan M., 622
Lamberton, Damien, 856
Lamy, Arnaud-Guilhem, 761
Lando, David, 210
Langlois, Hugues, 234
Lapeyre, Bernard, 856
Lardy, Jean-Pierre, 216, 217
Laurent, Jean-Paul, 71, 108, 152, 236, 246
L'Ecuyer, Pierre, 789
Lee, Daniel D., 956
Lee, Roger, 540, 557, 598
Lee, Sand-Bin, 524
Lee, Sharon X., 1057
Lehmann, Erich Leo, 67
Leland, Hayne E., 3, 4
Lennox, Clive, 925
Lesniewski, Andrew S., 559, 568, 600
Levy, Arnon, 103, 530
Lewis, Alan, 582
Lewis, Peter A.W., 788
Li, David X., 222, 230
Lin, Chih-Jen, 957, 1001
Lindner, Peter, 141
Liptser, Robert, 1064
List, John A., 447–449
Litterman, Robert B., 79, 105
Little, Roderick J.A, 943
Litzenberger, Robert H., 278, 508, 557
Lo, Andrew W., 476, 699, 700
Longstaff, Francis A., 146, 437, 575, 708, 821
Lopez, Jose A., 894
Loterman, Gert, 197
Louis, Thomas A., 746
Luciano, Elisa, 590
Lundstedt, Kyle G., 925
Lütkebohmert, Eva, 246
Lütkepohl, Helmut, 645–647
Lybek, Tonny, 347, 353

M
MacKay, Jock, 715, 732, 733
MacKinnon, James G., 618, 661
Maillard, Didier, 86
Mark, Nelson C., 634
Mark, Robert, 70, 408
Markit, 154, 155, 158
Markowitz, Harry M., 2
Marron, J. Steve, 639
Marrone, James, 246
Marsaglia, George, 813
Marshall, Albert W., 734, 803
Martens, David, 197
Martin, Reiner, 910
Marumo, Kohei, 170
Maskell, Simon, 879
Mason, Scott P., 215
Mazzeo, Michael, 987
McConnell, John J., 437
McFadden, Daniel, 635, 969
McLachlan, Geoffrey J., 1057
McNeil, Alexander J., 805
Mehr, Robert Irwin, 1
Memmel, Christoph, 898
Mercurio, Fabio, 513, 518, 520, 521, 524, 529, 537, 540, 598
Merton, Robert C., 2, 3, 169, 173, 214, 827
Metropolis, Nicholas, 838, 875
Meucci, Attilio, 1033
Mickey, M. Ray, 810
Mikosch, Thomas, 763, 769
Miller, James M., 788
Miller, Merton H., 3, 347
Mithal, Sanjay, 146
Modigliani, Franco, 3
Moler, Cleve, 1035
Morini, Massimo, 102
Morokoff, William J., 882
Morton, Andrew, 522, 525
Moudoulaud, Olivier, 326
Moussa, Amal, 476
Mues, Christophe, 191, 197, 200

Muksakunratana, Siriporn, 126
Muller, Mervin E., 793, 886
Murphy, Edward V., 12, 461
Musiela, Marek, 526
Myneni, Ravi, 575, 577, 601

N
Nagpal, Krishan, 234
Neis, Eric, 146
Nelsen, Roger B., 220, 715, 719, 722, 724, 725, 727, 734, 742
Nelson, Charles R., 79, 131
Neuvial, Pierre, 438, 444
Newbold, Paul, 657
Newey, Whitney K., 634
Nicolas, Emeric, 761
Nikeghbali, Ashkan, 747, 909
Nikolaou, Kleopatra, 347, 357–359, 365
Nocedal, Jorge, 1049
Novosyolov, Arcady, 919

O
Oakes, David, 734
O'Brien, James M., 435
OCC, 436
Oehmke, Martin, 456
Ohlson, James A., 925
O'Kane, Dominic, 144, 148
Øksendal, Bernt, 1064
Olkin, Ingram, 734, 803, 809, 813
O'Sullivan, Finbarr, 692
OTS, 436, 444
Oung, Vichett, 895, 903
Ozdaglar, Asuman, 476

P
Pakes, Ariel, 635
Pan, George, 216, 217
Pan, Jun, 92, 97
Panjer, Harry H., 312, 332
Parás, Antonio, 103, 530
Pawitan, Yudi, 692
Pedersen, Lasse Heje, 21, 355–358, 459, 467, 471
Pelizzon, Loriana, 476
Pennacchi, George G., 432, 433
Pérignon, Christophe, 456
Perron, Pierre, 661
Perry, T., 111
Persaud, Avinash D., 355
Pfeiderer, Paul, 4
Philippon, Thomas, 467, 471
Phillips, Peter C.B., 661, 662
Pickands, James, 774
Picoult, Evan, 268
Platen, Eckhard, 821
Pliszka, Kamil, 919
Pollard, David, 635
Pollock, David Stephen G., 671, 672
Polson, Nicholas G., 880
Pope, Devin G., 437, 438
Pope, Jaren C., 437, 438
Porter, Gary E., 448
Porter-Hudak, Susan, 698, 699
Powers, David Martin, 1017
Pozsar, Zoltan, 478
Prause, Karsten, 91
Prescott, Edward C., 704
Press, William H., 882, 1047, 1048
Priestley, Maurice Bertram, 693
Protter, Philip, 591
Pykhtin, Michael, 258, 266, 284, 295–297, 303

Q
Qi, Min, 190
Quesada-Molina, José Juan, 719
Quibel, Guillaume, 761
Quigley, John M., 437

R
Rabinowitz, Philip, 1039
Rahi, Rohit, 141
Rapuch, Grégory, 585, 588, 590
Rebonato, Riccardo, 102, 529
Redon, Christian, 275
Refenes, Apostolos N., 588
Reiner, Eric, 532, 546, 559
Reinhart, Carmen M., 7
Resnick, Sidney I., 765, 780
Revuz, Daniel, 829
Riboulet, Gaël, 224, 384–387, 415, 417, 418, 425, 426, 428, 431, 747, 909
Richardson, Matthew, 458, 467, 471–473
Ritchken, Peter, 525
Rivest, Louis-Paul, 746, 781
Rochet, Jean-Charles, 520
Rockinger, Michael, 473
Rodríguez-Lallena, José Antonio, 719
Roget, Céline, 234
Rogoff, Kenneth S., 7
Roll, Richard, 354

Roncalli, Thierry, 77, 105, 224, 234, 312, 318, 326, 337, 351–353, 465, 466, 585, 588, 590, 654, 689, 747, 761, 909, 1041
Rosen, Dan, 278, 290, 291, 295–297, 303
Rosenblatt, Frank, 976
Rosenbluth, Arianna W., 875
Rosenbluth, Marshall N., 875
Rosenfeld, Eric, 215
Rosenthal, Jeffrey S., 212
Ross, Sheldon M., 847, 848
Ross, Stephen A., 2, 76, 522
Rubin, Donald B., 622, 943
Rubinstein, Mark, 3, 4, 532, 537, 546, 559
Ruder, Sebastian, 982
Ruff, Craig K., 447–449
Ruiz, Esther, 669, 670

S
S&P, 164, 165
Sabato, Gabriele, 925
Salmon, Mark, 660
Salomon, Eric, 337
Sancetta, Alessio, 743
Sanders, Anthony B., 708, 821
Sankarasubramanian, Lakshminarayanan, 525
Santos, Edson B., 476
Sarr, Abdourahmane, 347, 353
Satchell, Stephen, 743
Satchkov, Daniel, 919
Saunders, David, 278, 290, 291
Scaillet, Olivier, 71, 108, 111, 116, 246
Schachter, Barry, 919
Schaede, Ulrike, 3
Schaefer, Stephen M., 100
Schapire, Robert E., 1004
Scheinkman, José A., 79
Schied, Alexander, 62
Schmee, Josef, 623
Schmidt, Peter, 662
Schmidt, Wolfgang, 226
Schnabl, Philipp, 459
Scholes, Myron, 2, 3, 73, 491
Schönbucher, Philipp J., 565
Schuermann, Til, 46, 192, 208, 342
Schwartz, Eduardo S., 437, 575
Schwartz, Robert A., 353
Schweizer, Berthold, 722, 725
Scott, Hal S., 342
Segoviano, Miguel A., 141, 177, 898
Selvaggio, Robert, 432, 433
Seung, H. Sebastian, 956
Shahabuddin, Perwez, 91
Sharpe, William F., 2, 49, 76, 355, 610
Sheather, Simon J., 639
Shephard, Neil, 669, 670
Shih, Joanna H., 746
Shiller, Robert J., 4
Shin, Yongcheol, 662
Shirazi, Ali, 282, 303
Shiryaev, Albert N., 1064
Shreve, Steven E., 511, 512, 530, 1064, 1067
Shumway, Tyler, 925
Sidenius, Jakob, 235
Siegel, Andrew F., 79, 131
SIFMA, 129, 130, 140
Silverman, Bernard W., 638
Sims, Christopher A., 902
Singer, Yoram, 1001
Singleton, Kenneth J., 635
Skintzi, Vasiliki D., 588
Sklar, Abe, 715, 716
Skorokhod, Anatolij V., 1064, 1068
Smillie, Alan, 112
Smith, Adrian FM, 871
Smith, Murray, 981
Smith, Stanley D., 448
Snider, H. Wayne, 1
Soner, H. Mete, 591
Sowell, Fallaw, 697, 698
Spanos, Aris, 651
Spector, Lee C., 987
Sprittulla, Jacob, 333
Stahl, Gerhard, 88, 89
Stanhouse, Bryan, 448
Stegun, Irene A., 669, 1041
Stein, Jeremy C., 177
Stewart, G.W., 809
Stock, Duane, 448
Stoer, Josef, 1041
Stoffer, David S., 880
Stoll, Hans R., 354
Stone, Mervyn, 939
Stulz, René M., 1
Subrahmanyam, Marti G., 144
Syrkin, Mark, 282, 303

T
Ta, Thomas, 216, 217
Tahbaz-Salehi, Alireza, 476
Tamari, Meir, 924

Tang, Dragon Yongjun, 144
Tankov, Peter, 590
Tanner, Martin A., 624
Taplin, Ross, 190
Tarullo, Daniel K., 160, 168, 893
Tasche, Dirk, 63, 105, 106, 109, 246
Tawn, Jonathan A., 731, 780
Tchen, André H., 727
Teller, Augusta H., 875
Teller, Edward, 875
Teukolsky, Saul A., 882, 1047, 1048
Thomas, Lyn C., 191, 200, 924, 925
Tibshirani, Robert, 933, 934, 939, 943, 944, 968, 1003, 1006, 1022
Tierney, John, 216, 217
Titman, Sheridan, 456
Tkachenko, Mykola, 191
To, Huong Minh, 190
Tolle, Jon W., 1049
Tong, Edward N.C., 191, 200
Tong, Howell, 693
Torous, Walter N., 437
Touzi, Nizar, 591
Trzcinka, Charles A., 349, 354
Tsatsaronis, Kostas, 901
Tufano, Peter, 3
Turnbull, Stuart M., 210

U
Ulam, Stanislaw, 838
Underhill, Les G., 809
Utzet, Frederic, 770

V
van Deventer, Donald R., 434, 451
Van Loan, Charles F., 810, 1035, 1036
Van Order, Robert, 437
Vapnik, Vladimir, 990, 996
Vasicek, Oldrich, 2, 169, 173, 514, 515, 522, 593
Vetterling, William T., 882, 1047, 1048
Viswanathan, Ravi, 575, 577, 601

W
Wagner, Wolfgang, 821
Walter, Ingo, 458
Wang, Dan, 540, 598
Wang, Jying-Nan, 46
Wang, Sarah Qian, 144
Warachka, Mitch, 591
Ward, Ian, 226
Wei, Jason Z., 212
Weisang, Guillaume, 351–353, 465, 466, 654
Welch, Ivo, 456
Welsch, John H., 1038
Wermers, Russ, 456
West, David, 925
West, Kenneth D., 634
Westwood, Ben, 898
White, Alan, 278, 290, 292, 516, 524
Whittle, Peter, 686
Wielhouwer, Jacco, 451
Wilde, Tom, 171, 245, 246, 253, 268
Willing, Jan, 451
Willmot, Gordon E., 312
Wilmott, Paul, 534, 535
Wilson, Nicholas, 925
Witzany, Jiří, 191
Wolff, Edward F., 722, 725
Woodward, Diana E., 559, 568, 600
Wright, Stephen J., 1049

Y
Yang, Bill Huajian, 191
Yang, H., 111
Yeh, Jin-Huei, 46
Yor, Marc, 829
Yorulmazer, Tanju, 456
Yoshiba, Toshinao, 170
Young, Robert A., 438

Z
Zangari, Peter, 85, 100
Zheng, Ban, 353
Zhu, Steven H., 258, 266
Zigrand, Jean-Pierre, 46, 454
Zou, Hui, 1022
Zullo, Ferdinando, 451