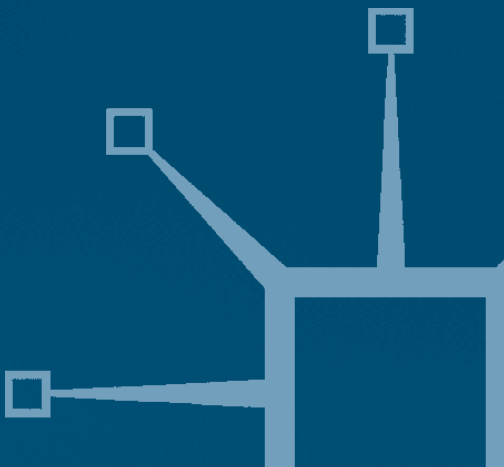
palgrave
macmillan

Derivatives and Internal Models

Second Edition

Hans-Peter Deutsch



Derivatives and Internal Models

Disclaimer

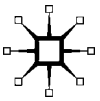
The information presented in this book has been derived from sources believed to be accurate and reliable, but it has not been independently verified in all cases. Accordingly, neither the author nor the publisher gives any representation or warranty of the accuracy, completeness or currentness of such information. The models, methods and examples in this book are only intended to demonstrate the relevant concepts in general terms. They may have to be adjusted or modified to be applied in real transactions. Moreover, the information presented in this book is not intended as financial advice or as a recommendation of any financial transactions. Neither the author nor the publisher is liable for any actions prompted or caused by the information presented in this book. In particular, neither the author nor the publisher gives any representation or warranty of the completeness, accuracy or functionality of the spread sheets and programs (macros) made available in this book. The spread sheets and programs (macros) are alterable. The cells of the spread sheets are not protected, nor are the source codes of the macros. The reader shall be fully responsible for correct handling and usage of these.

Derivatives and Internal Models

Second Edition

Hans-Peter Deutsch

palgrave



© Hans-Peter Deutsch 2002

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London W1P 0LP.

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First edition 1999
Reprinted three times
Second edition 2002

Published by
PALGRAVE
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N. Y. 10010
Companies and representatives throughout the world

PALGRAVE is the new global academic imprint of
St. Martin's Press LLC Scholarly and Reference Division and
Palgrave Publishers Ltd (formerly Macmillan Press Ltd).

ISBN 0–333–97706–8

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication data has been applied for.

10 9 8 7 6 5 4 3 2 1
11 10 09 08 07 06 05 04 03 02

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham, Wiltshire

Preface to the Second Edition

The great success of the first edition of this book (appearing under the co-authorship of Roland Eller and Hans-Peter Deutsch) quickly made it clear just how necessary a book of this type was. The philosophy of the first and even more so of the second edition is to provide an introduction to the valuation and risk management of modern financial instruments formulated in precise (and mathematically correct) expressions, covering all pertinent topics with a consistent and exact notation and with a depth of detail sufficient to give the reader a truly sound understanding of the material. An understanding which even places the reader in a position to independently *develop* pricing and risk management algorithms (including actually writing computer programs), should this be necessary. The reader is not expected to have previous knowledge of finance, but rather a sound mathematical and analytical background typical of scientists, mathematicians, computer scientists, engineers, etc.

In the second edition, the *breadth* of subjects covered has increased as has the *depth* of detail of their treatment. In addition, entirely new topics have been introduced. The second edition is also accompanied by a CD-ROM with Microsoft Excel$^{\text{TM}}$ workbooks presenting concrete realizations of the concepts discussed in the book in the form of executable algorithms. Of course, the reader has full access to all source codes of the Visual Basic$^{\text{TM}}$ modules as well as to all calculations done in the spread sheet cells. The CD-ROM thus contains a collection of literally thousands of examples providing the reader with valuable assistance in understanding the complex material and serving as the potential basis for the further development of the reader's own particular pricing and risk management procedures.

The second edition should equip the reader with a wide array of tools needed for all essential topics in the field of modern market risk management. The novice is not even required to be familiar with ideas as fundamental as compounding interest. The book, however, is certainly also of interest to the experienced risk manager or financial engineer, since the concepts introduced are widely elaborated upon and analyzed down to the foundations on which they are based, making a comprehension of the material possible which goes significantly beyond the level held to be "common knowledge" in this field.

Since the beauty of a room behind a closed door is of little use if the door cannot be found, emphasis has been placed on providing an easy *entry* into the analysis of each of the various topics. As the author does not wish to lose the reader at the outset, or expect the reader to first engage in the study of quoted literature before proceeding,

the book is practically self-contained. An explanation of almost every expression or notion needed can be found in the book itself, ranging from compounding interest to term structure models, from expectation to Value at Risk, from time series analysis to GARCH models, from arbitrage to differential equations and exotic options, from the normal distribution to martingales, and so on.

The selection of the topics and the nature of their presentation result to a great extent from my personal experience in *Financial and Commodity Risk Consulting* with *Andersen, Germany*, which it has been my pleasure to establish and direct in recent years. In this function, I have been in a position to observe and identify exactly what knowledge and methods are required in the financial world as well as to see what tools are indispensable for a newcomer to this world. The entire field of *credit risk* has been excluded from the range of topics covered, it being so extensive and complex that it deserves a book of its own. I am currently engaged in writing such a book on credit risk whose concepts will build upon the methods, thought processes and ideas presented here and which, as such, is conceived as Volume II of this book. The philosophy and presentation of material will follow that of the current volume here.

I would like to take this opportunity to thank several of the (in part former) members of the *Financial and Commodity Risk Consulting* team, namely *Mark Beinker*, *Matthias Heurich*, *Christian Hille* and *Michael Röhl* for their valuable input and many fruitful discussions, in particular on the topics, "Finite Differences and Trinomial Trees", "Principal Component Analysis", "Delta-Gamma Value at Risk" and "Autoregressive Models". I also wish to thank *Roland Eller*, some of whose contributions to the first edition (Sections 2, 4, 7.8, 8.12, 9.1.2 and the Glossary in that first edition) also appear (although in part strongly modified) in this edition (Sections 2, 3.1.1, 3.1.2, 4, 20 and 27.1.3 in the second edition). Last but not least, I would like to express my thanks to *Arthur Andersen, Germany GmbH* for the considerable funds made available for this project.

Wiesbaden, October 2001

Hans-Peter Deutsch

Contents

Preface to the Second Edition **v**

Part I Fundamentals **1**

1 Introduction **3**

2 Legal Environment **7**

3 Fundamental Risk Factors of Financial Markets **11**
3.1 Interest Rates . . . 11
3.1.1 Day Count Conventions . . . 12
3.1.2 Business Day Conventions . . . 14
3.1.3 Discount Factors . . . 15
3.1.4 Compounding Methods . . . 16
3.1.5 Spot Rates . . . 21
3.1.6 Forward Rates . . . 21
3.2 Market Prices . . . 23
3.3 An Intuitive Model for Financial Risk Factors . . . 24
3.3.1 Random Walks as the Basis for Pricing and Risk Models . . . 24
3.3.2 Risk Factors as Random Walks . . . 27
3.4 Ito Processes and Stochastic Analysis . . . 34
3.4.1 General Diffusion Processes . . . 34
3.4.2 Ito's Lemma . . . 34
3.4.3 Transition Probabilities, Forward and Backward Equation . . . 38
3.4.4 Forward and Backward Equation in the Black-Scholes World . . . 44

4 Financial Instruments: A System of Derivatives and Underlyings **47**
4.1 Spot Transactions . . . 48
4.1.1 Money Market Securities . . . 48
4.1.2 Capital Market Securities . . . 52
4.1.3 Swaps . . . 54
4.2 Forward Transactions . . . 55
4.3 Options . . . 58

Part II Methods **61**

5 Overview of the Assumptions **63**

6 Present Value Methods, Yields and Traditional Risk Measures **67**
6.1 Present Value and Yield to Maturity 67
6.2 Internal Rate of Return and Net Present Value 69
6.3 Accrued Interest, Residual Debt and Par Rates 71
6.4 Traditional Sensitivities of Interest Rate Instruments 74
6.4.1 Average Lifetime and Macaulay Duration 74
6.4.2 Modified Duration and Convexity 75
6.4.3 Summation of Traditional Sensitivities 78

7 Arbitrage **81**
7.1 Forward Contracts 81
7.1.1 Forward Price and Cash & Carry Arbitrage 81
7.1.2 The Stochastic Process for the Forward Price 83
7.1.3 Forward Positions 84
7.1.4 Future Positions and Basis Risk 84
7.2 Options 85
7.2.1 Upper and Lower Bounds for Option Prices 85
7.2.2 Early Exercise of American Options 86
7.2.3 Relationships between Puts and Calls 87

8 The Black-Scholes Differential Equation **89**
8.1 The Black-Scholes Equation from Arbitrage Arguments 90
8.1.1 The Black-Scholes Equation for European Options 90
8.1.2 The Black-Scholes *In*equality for American Options 92
8.1.3 A First Contact with the Risk-Neutral World 95
8.2 The Black-Scholes Equation and the Backward Equation 95
8.2.1 A Second Contact with the Risk-Neutral World 98
8.3 The Relationship to the Heat Equation 98

9 Integral Forms and Analytic Solutions in the Black-Scholes World **103**
9.1 Option Prices as Solutions of the Heat Equation 103
9.2 Option Prices and Transition Probabilities 105
9.3 Compilation of Black-Scholes Option Prices for Different Underlyings . 108
9.3.1 Options on the Spot Price 108
9.3.2 Options on the Forward Price 108
9.3.3 Options on Interest Rates 110

10 Numerical Solutions Using Finite Differences **113**
10.1 Discretizing the Black-Scholes Equation 114
10.1.1 The Explicit Method 115
10.1.2 The Implicit Method 115
10.1.3 Combinations of Explicit and Implicit Methods (Crank-Nicolson) 115

10.1.4 Symmetric Finite Differences of the Underlying Price 117
10.2 Difference Schemes 119
10.2.1 Initial Conditions 122
10.2.2 Dirichlet Boundary Conditions 123
10.2.3 Neumann Boundary Condition 127
10.2.4 Unspecified Boundary Conditions 132
10.2.5 Free Boundary Conditions for American Options 135
10.3 Convergence Criteria 139
10.3.1 Improving the Convergence Properties 141
10.4 Discrete Dividends 143
10.5 Example 144

11 Binomial and Trinomial Trees **147**
11.1 General Trees 147
11.1.1 Evolution of the Underlying and the Replicating Portfolio . . . 147
11.1.2 Evolution of the Derivative 148
11.1.3 Forward Contracts 150
11.2 Recombinant Trees 150
11.2.1 The Underlying 150
11.2.2 The Binomial Distribution for European Derivatives 151
11.2.3 A Third Contact with the Risk-Neutral World 155
11.3 The Relationship between Random Walk and Binomial Parameters . . 157
11.4 The Binomial Model with Infinitesimal Steps 160
11.4.1 Components of the Black-Scholes Option Pricing Formula . . . 161
11.5 Trinomial Trees 162
11.5.1 The Trinomial Tree as an Improved Binomial Tree 165
11.5.2 Relationship to the Explicit Finite Difference Method 165

12 Monte Carlo Simulations **167**
12.1 A Simple Example: The Area of a Disk 169
12.2 The General Approach to Monte Carlo Simulations 172
12.3 Monte Carlo Simulation of Risk Factors 173
12.3.1 Simulation of the Evolution of a *Single* Risk Factor 173
12.3.2 Simulation of *Several* Correlated Risk Factors 176
12.4 Pricing 178

13 Hedging **181**
13.1 Replicating Portfolios as Synthetic Derivatives 181
13.2 Hedging Derivatives with Spot Transactions 181
13.2.1 Forwards and Futures as Derivatives 183
13.3 Hedging Derivatives with Forward Contracts 184
13.3.1 Hedging with Forwards 185
13.3.2 Hedging with Futures 187
13.3.3 The Differential Equation for Derivatives on Futures 188
13.4 Hedge-Ratios for Arbitrary Combinations of Financial Instruments . . 189
13.5 “Greek” Risk Management with Sensitivities 191

13.5.1 Sensitivities and a Portfolio's Change in Value . . . 191
13.5.2 Omega and Beta . . . 194
13.5.3 Summation of Sensitivities of Different Underlyings . . . 196
13.6 Computation of the Greek Risk Variables . . . 197
13.6.1 Sensitivities in the Binomial Model . . . 197
13.6.2 Sensitivities in the Black-Scholes Model . . . 199
13.6.3 Sensitivities by Means of Finite Difference Methods . . . 200
13.6.4 Sensitivities by Means of Monte Carlo Simulations . . . 201

14 Martingale and Numeraire 203
14.1 The Martingale Property . . . 203
14.2 The Numeraire . . . 205
14.3 Self-Financing Portfolio Strategies . . . 208
14.4 Generalization to Continuous Time . . . 211
14.5 The Drift . . . 219
14.6 The Market Price of Risk . . . 222
14.7 Tradable Underlyings . . . 223
14.8 Applications in the Black-Scholes World . . . 224

15 Interest Rates and Term Structure Models 229
15.1 Instantaneous Spot Rates and Instantaneous Forward Rates . . . 230
15.2 Important Numeraire Instruments . . . 231
15.2.1 The Risk-Neutral Measure . . . 232
15.2.2 The Forward-Neutral Measure . . . 234
15.3 The Special Case of Deterministic Interest Rates . . . 235
15.4 Tradable and Non-Tradable Variables . . . 236
15.5 Convexity Adjustments . . . 238
15.5.1 LIBOR-In-Arrears Swaps . . . 241
15.5.2 Money Market Futures . . . 242
15.6 Arbitrage-Free Interest Rate Trees . . . 244
15.6.1 Backward Induction . . . 245
15.6.2 Forward Induction and Green's Functions . . . 248
15.7 Market Rates vs. Instantaneous Rates . . . 252
15.7.1 Arrow-Debreu Prices . . . 253
15.7.2 Pricing Caplets Using Arrow-Debreu Prices . . . 256
15.8 Explicit Specification of Short Rate Models . . . 258
15.8.1 The Effect of Volatility . . . 259
15.8.2 Normal Models . . . 260
15.8.3 Lognormal Models . . . 263
15.9 The Example Program TermStructureModels.xls . . . 266
15.9.1 Construction of Interest Rate Trees and Option Pricing . . . 266
15.9.2 Absolute and Relative Volatilities . . . 268
15.9.3 Calibration of Volatilities . . . 269
15.10 Monte Carlo on the Tree . . . 271
15.11 The Drift in Term Structure Models . . . 272
15.11.1 Heath-Jarrow-Morton Models . . . 273

15.11.2 Short Rate Models . . . 274
15.12 Short Rate Models with Discrete Compounding . . . 277
15.12.1 Normal Models . . . 278
15.12.2 Lognormal Models . . . 279

Part III Instruments 281

16 Spot Transactions on Interest Rates 283
16.1 Zero Bonds . . . 283
16.1.1 Cash Flows and Present Value . . . 283
16.1.2 Yield To Maturity and Par Rate . . . 284
16.1.3 Sensitivities . . . 284
16.2 Floaters . . . 285
16.2.1 Cash Flows and Present Value . . . 285
16.2.2 Yield to Maturity, Par Rate and Sensitivities . . . 286
16.3 Coupon Bonds . . . 287
16.3.1 Cash Flows and Present Value . . . 287
16.3.2 Yield To Maturity . . . 287
16.3.3 Par Rates . . . 289
16.3.4 Sensitivities . . . 291
16.4 Swaps . . . 292
16.4.1 Cash Flows and Present Value . . . 294
16.4.2 Yield to Maturity and Swap Rate . . . 294
16.4.3 Sensitivities . . . 296
16.5 Annuity Loans . . . 297
16.5.1 Cash Flows and Residual Debt . . . 297
16.5.2 Present Value . . . 299
16.5.3 Yield To Maturity and Par Rates . . . 302
16.5.4 Sensitivities . . . 305
16.5.5 Remarks on Credit Risk . . . 305

17 Forward Transactions on Interest Rates 307
17.1 Forward Rate Agreements . . . 307
17.2 Interest Rate Futures . . . 308
17.2.1 Futures on Zero Bonds . . . 308
17.2.2 Futures on Coupon Bonds . . . 308
17.3 Forward Swaps . . . 311
17.3.1 Present Value . . . 312
17.3.2 Forward Swap Rates and Forward Yield to Maturity . . . 313
17.4 Forward Bonds . . . 315
17.4.1 Present Value . . . 315
17.4.2 Forward Par Rate and Forward Yield to Maturity . . . 317

18 Plain Vanilla Options **319**
18.1 Options on Spot and Forward Prices . . . 320
18.1.1 European Options . . . 320
18.1.2 American Options . . . 321
18.2 Index Options and Index Futures . . . 322
18.3 Foreign Exchange Options and Futures . . . 324
18.3.1 Put-Call-Equivalence for FX Options . . . 324
18.3.2 FX Forward Contracts and FX Swap Rates . . . 324
18.4 Interest Rate Options . . . 325
18.4.1 Options on Bonds . . . 326
18.4.2 Options on Bond Futures . . . 326
18.4.3 Caps and Floors . . . 327
18.4.4 Swaptions . . . 333

19 Exotic Options **339**
19.1 Traditional and General Definition of an Option . . . 339
19.2 Payoff Profiles for Selected Exotics . . . 339
19.2.1 Power Options . . . 340
19.2.2 Cliquet and Coupe Options . . . 340
19.2.3 Look-Back Options . . . 340
19.2.4 Asian Options . . . 341
19.2.5 Rainbow and Exchange Options . . . 341
19.2.6 Compound and Bermuda Options . . . 342
19.3 Black-Scholes for Exotics . . . 343
19.3.1 Pay-Later Options . . . 343
19.3.2 Digital Options . . . 344
19.3.3 Barrier Options . . . 346
19.3.4 Ladder Options . . . 350
19.4 Numerical Pricing Methods for Exotics . . . 352
19.4.1 Monte Carlo for European Exotics . . . 353
19.4.2 The Binomial Model for American Exotics . . . 357

20 Structured Products and Stripping **363**
20.1 Coupon and LIBOR Strips vs. Bonds . . . 364
20.2 Structured Bonds . . . 365

Part IV Risk **371**

21 Fundamentals **373**
21.1 Risk, Confidence, Percentile and Value at Risk . . . 374
21.2 The Value at Risk of a Single Risk Factor . . . 376
21.3 Approximations in the Distribution of Risk Factors . . . 381
21.4 The Covariance Matrix . . . 383
21.4.1 Covariance Matrices of Data Providers . . . 385
21.4.2 Cholesky Decomposition of the Covariance Matrix . . . 386

22 The Variance-Covariance Method **391**
22.1 Portfolios vs. Financial Instruments . . . 393
22.2 The Delta-Normal Method . . . 395
22.2.1 The Value at Risk with Respect to a Single Risk Factor . . . 395
22.2.2 The Value at Risk with Respect to Several Risk Factors . . . 397
22.3 The Delta-Gamma Method . . . 399
22.3.1 Decoupling of the Risk Factors . . . 400
22.3.2 Diagonalization of the Gamma Matrix . . . 401
22.3.3 The Distribution of the Portfolio Value Changes . . . 406
22.3.4 Moments of the Portfolio Value Distribution . . . 407
22.3.5 Fourier-Transformation of the Portfolio Value Distribution . . . 415
22.3.6 Monte Carlo Simulation of the Portfolio Value Distribution . . . 418

23 Simulation Methods **419**
23.1 Monte Carlo Simulation . . . 419
23.1.1 The Risk Factors as Correlated Random Walks . . . 420
23.1.2 Structured Monte Carlo . . . 421
23.2 Historical Simulation . . . 422
23.3 Crash and Stress Testing: Worst Case Scenarios . . . 424

24 Interest Rate Risk and Cash Flows **427**
24.1 Cash Flow Structures of Financial Instruments . . . 427
24.1.1 Spot Transactions . . . 428
24.1.2 Forward Transactions . . . 430
24.1.3 Options . . . 431
24.2 Interpolation and Cash Flow Mapping . . . 440
24.2.1 Interpolation . . . 440
24.2.2 Risk-Based Cash Flow Mapping . . . 441
24.2.3 Duration-Based Cash Flow Mapping . . . 443

25 Example of a VaR Computation **445**
25.1 The Portfolio . . . 445
25.2 Market Data . . . 446
25.3 Cash Flow Mapping . . . 447
25.4 Calculation of Risk . . . 448

26 Backtesting: Checking the Applied Methods **451**
26.1 Profit-Loss Computations . . . 451
26.2 The Traffic Light Approach of the Supervising Authorities . . . 452
26.2.1 Adjusting the Value at Risk (Yellow zone) . . . 452
26.2.2 Criteria for Rejecting a Model (Red Zone) . . . 454
26.2.3 The Green Zone . . . 456
26.2.4 Multiplication Factor and Add-On . . . 456

Part V Market Data **459**

27 Interest Rate Term Structures **461**
27.1 Bootstrapping . . . 462
27.1.1 The General Bootstrapping Equation . . . 462
27.1.2 Bootstrapping for Coupon Periods of Equal Length . . . 463
27.1.3 Classical Bootstrapping . . . 467
27.2 Interpolations . . . 472

28 Volatility **475**
28.1 Implied Volatilities . . . 475
28.1.1 Smiles and Volatility Indices . . . 475
28.2 Local Volatility Surfaces . . . 477
28.2.1 Implicit Transition Probabilities . . . 477
28.2.2 Implicit Local Volatility Surfaces . . . 479
28.3 Volatility Transformations . . . 483
28.3.1 Transformation between Relative and Absolute Volatility . . . 483
28.3.2 Summation of Volatilities . . . 484
28.3.3 Transformation between Yield and Price Volatility . . . 486
28.3.4 Currency Transformation of Volatilities and Correlations . . . 487

29 Market Parameter from Historical Time Series **495**
29.1 Historical Yields, Volatility and Correlation . . . 495
29.2 Autocorrelation and Autocovariance . . . 497
29.3 Error Estimates . . . 498
29.3.1 Treatment of Autocorrelations . . . 503

30 Time Series Modeling **505**
30.1 Stationary Time Series and Autoregressive Models . . . 508
30.1.1 AR(p) Processes . . . 508
30.1.2 Univariate GARCH(p, q) Processes . . . 513
30.1.3 Simulation of GARCH Processes . . . 516
30.2 Calibration of Time Series Models . . . 516
30.2.1 Parameter Estimation for AR(p) Processes . . . 518
30.2.2 Parameter Estimation for GARCH(p, q) Processes . . . 520
30.2.3 Simulated Annealing . . . 521

31 Forecasting with Time Series Models **525**
31.1 Forecasting with Autoregressive Models . . . 526
31.2 Volatility Forecasts with GARCH(p, q) Processes . . . 528
31.2.1 Forecast Over Several Time Steps . . . 528
31.2.2 Forecast for the Total Variance . . . 532
31.2.3 Volatility Term Structure . . . 533
31.3 Volatility Forecasts with GARCH(1, 1) Processes . . . 533
31.4 Volatility Forecasts with Moving Averages . . . 535

32 Principle Component Analysis **539**
32.1 The General Procedure . . . 539
32.2 Principle Component Analysis of the German Term Structure . . . 545

33 Pre-Treatment of Time Series and Assessment of Models **549**
33.1 Pre-Treatment of Time Series . . . 549
33.1.1 Differencing . . . 549
33.1.2 Filters . . . 550
33.1.3 Scaling . . . 552
33.2 Measuring the Goodness of Time Series Models . . . 552
33.2.1 Hypothesis Tests . . . 553
33.2.2 Goodness of Fit vs. Goodness of Forecast . . . 557
33.2.3 Examples: Goodness of GARCH Models . . . 558

Appendix A: Probability and Statistics **565**
A.1 Probability, Expectation and Variance . . . 565
A.2 Multivariate Distributions, Covariance, Correlation and Beta . . . 567
A.3 Moments and Characteristic Functions . . . 570
A.3.1 Moment Generating Functions . . . 571
A.3.2 Characteristic Functions . . . 573
A.4 A Collection of Important Distributions . . . 574
A.4.1 The Uniform Distribution . . . 574
A.4.2 The Binomial Distribution and the Bernoulli Experiment . . . 576
A.4.3 The Normal Distribution and the Central Limit Theorem . . . 578
A.4.4 The Lognormal Distribution . . . 584
A.4.5 The Gamma Distribution . . . 585
A.4.6 The χ^2-Distribution . . . 587
A.5 Transformations Between Distributions . . . 593
A.5.1 Summations . . . 593
A.5.2 Box-Muller Transformations . . . 594
A.5.3 Inversion of Cumulative Distribution Functions . . . 595

Bibliography **599**

Index **609**

Part I

Fundamentals

1

Introduction

The explosive development of derivative financial instruments continues to provide new possibilities and increasing flexibility to manage finance and risk in a way specifically tailored to the needs of individual investors or firms. This holds in particular for banks and financial services companies who deal primarily with financial products, but is also becoming increasingly important in other sectors as well. Active financial and risk management in corporate treasury can make a significant contribution to the stability and profitability of a company. For example, the terms (price, interest rate, etc.) of transactions to be concluded at a future date can be fixed today, if desired even giving the company the option of declining to go ahead with the transaction later on. These types of transactions obviously have some very attractive uses such as arranging a long-term fixed-rate credit agreement at a specified interest rate a year in advance of the actual transaction with the option to forgo the agreement if the anticipated need for money proves to have been unwarranted (this scenario is realized using what is known as a "payer swaption") or providing a safeguard against fluctuations in a foreign currency exchange rate by establishing a minimum rate of exchange today for changing foreign currency into euros at a future date (using a foreign currency option).

However, the complexity of today's financial instruments and markets and the growing pace of technological progress have led to an almost uncontrollable increase in the risks involved in trading and treasury while simultaneously reducing drastically the time available for decision making. Thus the inappropriate use of financial instruments may quickly result in losses wiping out gains achieved over years in a company's primary business (for example, the production and sale of automobiles or computer chips).

In recent years we have seen an increase in the number of sizable losses incurred in consequence of derivative transactions which, in some cases, have resulted in bankruptcy. This phenomenon has not been restricted to banks but has involved companies in various other sectors as well. Spectacular examples include Metallgesellschaft (oil futures), Procter & Gamble (speculation with exotic "power options"), Orange County (interest rate derivatives, highly leveraged portfolio), Barings (very large, open index futures positions), Daiwa Bank (short-term US Bonds), NatWest Markets (incorrect valuation of options in consequence of incorrect volatility assumptions) and – of course – the

hedge fund LTCM.

Not only have *financial instruments* become increasingly risky ("more volatile"), so have the *markets* themselves. Since the beginning of the 1980s we have seen a fundamental change to the economic framework in the financial world. In today's investment environment, yields, foreign currency exchange rates, commodity and stock prices can shift daily to an extent which would have been inconceivable in previous years. Increased market fluctuation (volatility) is the financial markets' reaction to developments such as the accumulation of capital, the globalization of financial markets, an increase in the budget deficits of the leading industrialized nations and the dismantling of government regulations, to name just a few.

The main prerequisite for the continued success of a bank or treasury department in an environment of highly volatile markets and extremely complex (and thus very sensitive) financial instruments is a sound understanding of the products being traded and efficient management of the risk involved in these transactions. Derivatives are therefore the main *reason for* and the most effective *means of* conducting risk management, and thus can be viewed as the be-all and end-all of risk management.

The *pricing* of derivatives and structured financial instruments is the foundation, the conditio sine qua non, of all risk management. To this end, all derivatives under consideration, regardless of their complexity, are broken down into their basic components and systematically categorized as spot, futures and option transactions. This process is called *stripping*. Understanding the fundamental instruments and their valuation, as well as "stripping" poses a serious challenge for those involved, requiring skills which in the past were not called for to any extent in the financial world, such as the following

- Theoretical demands: statistics, probability theory, time series analysis, etc., essential to the understanding of concepts such as "value at risk" and the estimation of underlyings and their correlation; differential equations, martingale theory, numerical analysis, financial mathematics, etc., for pricing financial instruments.

- Trading demands: understanding the function of increasingly complicated securities and transactions (e.g. by stripping); making "imaginative" choices of derivatives or constructing new instruments to realize a particular strategy while taking into consideration the often surprising "side effects"; developing complicated hedging strategies; operating highly complex computer systems.

- IT demands: client-server computing, distributed systems, object-oriented programming, new operating systems, network engineering, real time information distribution, Internet and intranet architecture.

The theoretical and trading aspects in particular pervade all business hierarchies, since it is equally important for the trader/ treasurer to be able to understand the risk involved in his or her portfolio as it is for a managing director or the board of directors whose decisions are based on this information.

Thus, the aim of this book is to provide a detailed presentation of the methods and procedures used in pricing financial instruments and identifying, quantifying and controlling risk. The book is structured as follows:

In Part I, the legal environment in which banks operate when trading will be briefly discussed in order to provide additional motivation for risk management and risk controlling. Examples of the laws and guidelines making up this environment are the *Basel market risk paper*, the *G30 study*, the Minimal Requirements for Trading Activities of Credit Institutions in Germany (in short *MaH*), the *Capital Adequacy Directive* (in short *CAD*) and *Principle I* of the sixth amendment to the Banking Act (*KWG*) in Germany, which allows *internal models* in market risk management. The next step, namely internal models for credit risk and operational risk is planned for the near future, while the Basel Committee is working under extreme pressure to produce an amendment of the already thoroughly outdated *Basel Accord* of 1988.

The next chapter introduces the fundamental market parameters which govern the price and risk of financial instruments. These include the price of stocks, commodities, currencies, etc. and naturally interest rates. A stochastic model, know as the *random walk* and its generalization, the *Ito process*, will be used to model these fundamental risks, also known as *risk factors*, and will serve to describe both the deterministic and the random aspects of the risk factors. In doing so, an additional parameter, called *volatility* will, similar to the above mentioned risk factors, play an important role in the price and risk accorded to a financial instrument.

Finally, in Part I, the most common instruments for trading the risk factors are introduced. *Trading* is defined as the acceptance of risk or risk factors, through a financial instrument, in return for a certain yield from the instrument and/or specified payments from a contracting party. Conversely, a financial instrument can of course also be used to transfer risk to a contracting party in exchange for paying a specified yield and/or amount to the risk-taking party. In Part I, the *functionality* of several of the most important financial instruments will be introduced without entering into a discussion of their *valuation*, which is frequently very complicated.

Once the fundamental risks and the financial instruments used for trading them have been introduced, the building blocks for further analysis will have been defined. This will allow us to proceed with the actual topic of this book, namely determining the *value* and the *risk* of these building blocks. The specific valuation of the most common financial instruments will be dealt with in Part III, the determination of the risks associated with these instruments will be presented in Part IV, and possible methods for determining the risk factors will be the topic of Part V.

The same methods will be repeatedly applied to the pricing and risk management of the various instruments. In order to make this clear, the most important *methods* for pricing and hedging will be introduced in their full generality in Part II. As a result, this section will be rather theoretical and technical. The concepts elaborated upon in Part II will then be applied to rather simple financial instruments in Part III. This separate treatment of general methods and specific financial instruments will contribute to a clearer understanding of this extremely complex material.

2

Legal Environment

The financial services industry (banks, funds manager, insurance companies and other) faces the largest risks arising from financial instruments because trading in these instruments is their core business. Banking supervisory authorities and legislators have reacted to this situation with fundamentally new legal provisions imposing very high requirements on the risk management of banks. In 1993 the Washington-based organization *Group of Thirty* (*G30*) published a study on the risk management of derivatives entitled "Derivatives: Practices and Principles". This study is also known as the *G30 Report* and puts forward recommendations which form the basis of the *Minimum Requirements for Trading Activities* (*MaH*) in Germany and other such regulations. The following list shows the most important recommendations made in the G30 Report:

- Daily marking-to-market.
- Valuation using mean market prices taking into account special costs arising in the future.
- Identification of sources of return to understand risks taken.
- Daily determination of market risk and limit utilization.
- Simulation of worst case scenarios on a regular basis in order to understand how portfolios might behave in the event of a crisis.
- Predicting liquidity for derivative portfolios.
- Establishing an independent market risk management function in order to develop a unified approach to control risks taken.
- Due to the complexity of the products the final users (corporates) should, depending on the size of their business, use the same methods for quantifying and valuing risks as the suppliers.

The *Minimum Requirements for Trading Activities* in Germany specify that trading, processing and control, accounting and monitoring should be functionally separated.

They also require the implementation of a risk control and management system[1]: "To limit the risk arising from trading activities a system must be installed which allows the risk positions to be measured and controlled and the potential losses which may arise from them (risk controlling) to be analyzed. This system must take account of the size, complexity and risk potential of the trading activities. It has to document and quantify market price risk. It should also be integrated into a policy for risk monitoring and control encompassing all divisions of the bank. The system must be precisely documented, tested and regularly updated. Additional information on risk positions should be obtained by means of scenario analysis which also allows "worst case" situations to be taken into account."

The Minimum Requirements do not specify exact limits or quantitative guidelines on how to actually *measure* risk. This is due to differences in the size and the scale of risk involved in the trading activities of different financial institutions. It therefore depends mainly on the financial institution and its board of management as to how the Minimum Requirements for Trading Activities are actually implemented. The *Basel Committee on Banking Supervision* emphasizes the responsibility of the bank as follows[2]: "As a result of a study in 1993 the Basel committee refrained from its original intention of developing a measurement system to identify banks with extreme interest rate risks. This would have required a standardization of measurement procedures, which the Basel committee did not think possible in view of the wide variety of methods used in banking practice at present."

The Basel Committee on Banking Supervision has put forward requirements that are similar to the Minimum Requirements for Trading Activities. The 37-page paper entitled "Principles for the Management of Interest Rate Risk" was published on September 19, 1997. It basically concentrates on twelve principles. The sixth principle, for example, reads: "It is essential that banks have interest rate risk measurement systems that capture all material sources of interest rate risk and that assess the effect of interest rate changes in ways that are consistent with the scope of their activities. The assumptions underlying the system should be clearly understood by risk managers and bank management." Like the Minimum Requirements for Trading Activities this paper does not prescribe exact quantitative methods and is thus a further example of the trend towards qualitative supervision in the place of quantitative supervision.

The *Capital Adequacy Directive* (*CAD*), which has been formulated for the whole of Europe, is much more explicit. In Germany this directive has been implemented by the reform of the so-called *Principle I* in the sixth amendment to the Banking Act (KWG). The first seven sections of "Principle I" describe in detail how risk is to be measured and how much capital is needed to back the risks measured. For the measurement of interest rate risk, for example, the directive prescribes exactly how future cash flows resulting from financial instruments are to be assigned to so-called *maturity bands* or *time buckets*. It also sets out how cash flows in each maturity band influence the overall risk (taking account of the volatilities of interest rates for the different maturities), and to what extent offsetting cash flows in different maturity bands may be netted (taking account of correlations between the interest rates of different maturity bands). The

[1] Translated from [74].
[2] See their web pages at www.bis.org

exact provisions of the directive do not only refer to interest rate risks, but also to risks relating to the prices of stocks, currencies and commodities and even to certain risks relating to options (at least in the case of the so called *delta*, *gamma* and *vega* risks). For all these risk factors there are explicit rules not only on how high the required capital backing must be for a given risk but also how this risk is to be determined.

At first sight it appears as if this directive would lead to a "enforced conformity" of banks. The second phase of the directive (*CAD II*), however, allows for so-called *internal models*, which banks can use to measure their risk using methods different from those specified in directive. These internal models have to meet certain quantitative and qualitative criteria in order to receive the approval of the supervisory authorities. Apart from the quantitative criteria such as the use of a historical time series of at least one year, the assumption of a *liquidation period* for any portfolio of exactly ten days, and the indication of a maximum possible loss, which has a 99% probability of *not* occurring (to be proven by the so-called *backtesting*), it is above all the qualitative criteria that are difficult to meet and difficult to verify[3]. Some quotations from the "Amendment to the Capital Accord to Incorporate Market Risks" by the Basel Committee on Banking Supervision of January 1996 (modified September 1997) may illustrate this:

- There has to be a review of the mathematical-statistical methods and the underlying quantitative assumptions on a regular basis, the review should be documented in detail.

- Trading limits should be related to the bank's risk measurement model that is consistent over time.

- The bank's risk measurement system must be well documented, for example, through a risk management manual that describes the basic principles of the risk management system and that provides an explanation of the empirical techniques used to measure market risk.

- An independent review of the risk measurement system should be carried out regularly in the bank's own internal auditing process. A review should take place at regular intervals (ideally not less than once a year) and should specifically address, at a minimum:

  - the adequacy of the documentation of the risk management system and process;
  - the validation of any significant change in the risk measurement process;
  - the integrity of the management information system;
  - the verification of the consistency, timeliness and reliability of data sources used to run internal models, including the independence of such data sources;
  - the accuracy and appropriateness of volatility and correlation assumptions;
  - the accuracy of valuations and risk transformation calculations;

[3] All this is explained in detail in this book, see especially Part IV.

– the verification of the model's accuracy through frequent back-testing as described in the accompanying document "Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements".

Thus, even though suitable risk management systems are a necessity today mainly due to the increasing complexity of the financial markets, legal requirements – whether proposed or enacted – also play an important role in promoting the implementation of such systems.

3

Fundamental Risk Factors of Financial Markets

The fundamental risk factors in financial markets are the market parameters which determine the price of the financial instruments being traded. They include foreign currency exchange rates and the price of commodities and stocks and, of course, interest rates. Fluctuations in these fundamental risks induce fluctuations in the prices of the financial instruments which they underlie. They constitute an inherent market risk in the financial instruments and are therefore referred to as *risk factors*. The risk factors of a financial instrument are the market parameters (interest rates, foreign currency exchange rates, commodity and stock prices), which, through their fluctuation, produce a change in the price of the financial instrument. The above mentioned risk factors do not exhaust the list of the possible risk factors associated with a financial instrument nor do all risk factors affect the price of each instrument; for example, the value of a 5 year coupon bond in Czech Coruna is not determined by the current market price of gold. The first step in risk management is thus to identify the relevant risk factors of a specified financial instrument.

3.1 Interest Rates

Various different *conventions* are used in the markets to calculate interest payments. For example, interest rates on securities sold in the US money markets (*T-bills*, *T-bill futures*) are computed using *linear compounding*, whereas in the European money market, *simple compounding* is used. Interest rates in the capital markets are calculated using *discrete* or *annual* compounding while option prices are determined using the *continuous* compounding convention. While these conventions are not essential for a principle understanding of financial instruments and risk management they are of central importance for the *implementation* of any pricing, trading or risk management system. Before entering into a general discussion of interest rates in Section 3.1.5, we will therefore introduce the most important compounding conventions here.

3.1.1 Day Count Conventions

Before one of the many compounding conventions are applied to calculate the interest on a certain amount over a period between the time (date) t and a later time T, the number of days between t and T over which interest is accrued must first be counted. The beginning, the end and the length of this time period (usually measured in years) must be precisely specified. To do this there are again different conventions used in different markets, known as *day count conventions*, *DCC* for short.

These market usances are usually specified by made making use of a forward slash notation: the method for counting the days of the month are specified in front of the slash, the number of days of the year after the slash. A list of the most commonly used conventions is presented in Table 3.1. These conventions compute the length of an interest rate period as follows:

- *Act/Act*: The actual number of calendar days are counted and divided by the actual number of days in the year.
- *Actual/365f:* The actual number of calendar days between t and T are counted and divided by 365 to obtain the interest period in years regardless of whether the year concerned is a leap year. This distinguishes the Actual/365f convention from the Act/Act in years.
- *Actual/360*: The actual number of calendar days between t and T are counted and divided by 360 to obtain the interest period in years.
- *30/360:* The days are counted as if there were exactly 30 days in each month and exactly 360 days in every year. In addition, the following holds:
  - If the *beginning* of the interest period falls on the 31st of the month, the beginning is moved from the 31st to the 30th of the same month for the purpose of the calculation.
  - If the *end* of the interest period falls on the 31th of the month, it is moved forward for the purposes of the calculation to the 1st of the next month **unless** the *beginning* of the interest period falls on the 30th or 31st of a month. In this case, the end of the interest period is moved from the 31st to the 30th of the same month.
- *30E/360:* The days are counted as if there were exactly 30 days in each month and exactly 360 days in each year. In addition, the following holds:
  - If the *beginning* of the interest period falls on the 31st of the month, the beginning is moved from the 31st to the 30th of the same month for the purpose of the calculation.
  - If the *end* of the interest period falls on the 31st of the month, the end of the period is moved to the 30th of the same month for the purpose of the calculation (this differentiates this day count convention from the 30/360 convention).

Common Designation	**Alternative Designation**
Act/Act	Act, Act/365(l), Act/365(leap),Act/365leap
Act/365f	Act/365(f), Act/365(fixed), Act/365fixed
Act/360	
30/360	BondBasis, Bond, 30
30E/360	EuroBondBasis, EuroBond, 30E

Table 3.1: The commonly used day count conventions.

These conventions for calculating the time between two dates can also be expressed in formulas:

$$\begin{aligned}
Act/365f &: \frac{D_1 - D_2}{365} \\
Act/360 &: \frac{D_1 - D_2}{360} \\
30/360 &: J_2 - J_1 + \frac{M_2 - M_1}{12} \\
&\quad + \frac{T_2 - \min(T_1, 30) - \max(T_2 - 30, 0) * \text{feb}(T_1 - 29)}{360} \\
30E/360 &: J_2 - J_1 + \frac{M_2 - M_1}{12} + \frac{\min(T_2, 30) - \min(T_1, 30)}{360} \\
Act/Act &: J_2 - J_1 + \frac{D_2 - \text{Date}(J_2, 1, 1)}{\text{Date}(J_2 + 1, 1, 1) - \text{Date}(J_2, 1, 1)} \\
&\quad - \frac{D_1 - \text{Date}(1, 1, J_1)}{\text{Date}(J_1 + 1, 1, 1) - \text{Date}(J_1, 1, 1)}
\end{aligned} \tag{3.1}$$

The notation should be read as follows: D_1 denotes the start date of the interest period, D_2 the end date. D_1 consists of the number T_1 for the days, M_1 for the months and J_1 for the years; D_2 is defined analogously. The function "Date" delivers the running number for each given date. The counting of this running number begins at some time in the distant past, for example January 1, 1900. Then the date function (as defined in Microsoft Excel, for instance) yields the value 35728 for Date (1997, 10, 25). This is the running number corresponding to October 25, 1997. The function "feb(x)" is defined as equal to zero for $x \leq 0$, and 1 otherwise. The function $\min(x, y)$ yields the smaller of the two values x and y, and the function $\max(x, y)$ yields the larger of the two.

The computation of time periods using these day count conventions is demonstrated in Example 3.1 and in the Excel-Sheet USANCE.XLS on the accompanying CD-ROM. The time period has been intentionally selected to contain the 29th of February in order to demonstrate the difference between the *Act/Act* and *Act/365f* conventions. Furthermore, the interest period was chosen to end on the 31st of the month to illustrate the difference between the *30/360* and *30/E360* conventions.

start date Feb-15-00
end date Dez-31-00

	time period	
	days	years
Act/act	320	0.87431694
Act/365f	320	0.87671233
Act/360	320	0.88888889
30/360	316	0.87777778
30/E360	315	0.875

Figure 3.1: Determining the length of a time period using different day count conventions. The time period was chosen to yield a different length (in years) in each day count convention.

Convention	**Adjusted to:**
Following	the following business day
Modified Following	the following business day as long as this falls within the same month; otherwise the preceding business day
Preceding	the preceding business day
Modified Preceding	the preceding banking business day as long as it falls within the same month; otherwise the following business day

Table 3.2: Business day conventions.

3.1.2 Business Day Conventions

Establishing a day count convention is not sufficient to uniquely determine interest periods. The *value dates* of the cash flows must also be defined, i.e. the number of days following the end of an interest period T on which the interest payment must be settled. Several different conventions govern this calculation as well. Furthermore, there are conventions to account for weekends and holidays. If the value date falls on a *bank holiday*, for example, should payment be made on the day before or after the holiday? And finally, bank holidays themselves vary from country to country. The conventions governing these questions are called *business day conventions*, *BDC* for short. These conventions transfer the value dates of a cash flow away from weekends and bank holidays in accordance with the rules compiled in Table 3.2.

Table 3.3 shows as an example the adjustment of March 28, 1996 and April 1, 1997. These examples are interesting in that the days March 29, 1997 and March 30, 1997 fall on a weekend and the days March 28, 1997, March 30, 1997 and March 31, 1997 represent bank holidays (Easter holidays) according to the holiday calendar valid for banks in Frankfurt, Germany.

For interest rate instruments, a further distinction is made between whether the

Convention	Adjustment March 28, 1997	Adjustment April01, 1997
Following	April 01, 1997	April 02, 1997
Modified Following	March 27, 1997	April 02, 1997
Preceding	March 27, 1997	March 27, 1997
Modified Preceding	March 27, 1997	April 02, 1997

Table 3.3: Effects of business day conventions.

adjustment convention holds solely for the *payment* date of an interest period or for its *maturity* date as well. The *maturity* date determines the length of the interest rate period and thus affects the *amount* of the interest payment (if the length of an interest period is longer, the amount of the interest payment to be made is naturally higher). The *payment* date determines *when* the interest payments are actually made (usually one or two business days after the maturity date) and therefore affects how strongly a payment is discounted, in other words, today's *value* of the payment (a later payment is obviously worth less than an earlier one); it is thus relevant when the payment is actually made and not when it was due.

If the maturity date of a financial instrument is specified as *fixed,* i.e. non-moveable, it is not adjusted. However, the *payment* date is still adjusted in accordance with the business day convention for the instrument concerned. The *rollover day* of an interest rate instrument determines on which day and month of each year the *rollover* from one interest period into the next is to take place, i.e. when the maturity and payment dates of individual interest payments occur. Depending on the selected business day convention, a decision is made as to how the maturity and payment dates, derived from the rollover date, are to be adjusted. For example, federal bonds are commonly agreed to be *fixed.* This means that only the *payment* date is adjusted to the next valid business day while the *maturity* date is always the same as the rollover day. For swaps, on the other hand, both the payment and maturity dates are adjusted to the next valid business day subsequent to the rollover day.

All these conventions make trading substantially more complicated without causing a fundamental change in the properties of the instruments being traded. They are actually unnecessary but the markets are inconceivable without them because of strong historical ties. This is particularly true of holiday calendars, some of which even have religious roots.

3.1.3 Discount Factors

In order to concentrate on the essentials, a general notation will be observed in this book which holds for *all* compounding, day count, business day and other market conventions. To accomplish this, discount factors rather than interest rates will be employed throughout. The *discount factor* is the value by which a *cash flow* to be paid at a time T is multiplied in order to obtain the value of the cash flow at an earlier time t (for example, today). Since we (usually) would prefer to have a cash flow today rather than at some future date T, the cash flow paid at time T is worth less today than the same cash flow paid at time t. Thus, the discount factor is generally less than

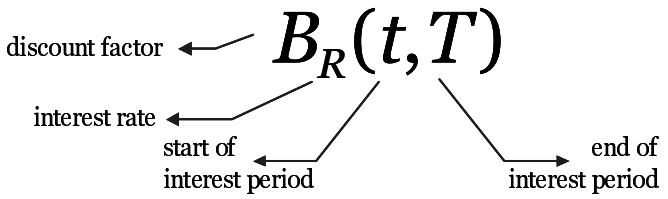


Figure 3.2: The general discount factor for the time span from t to T.

1 (but greater than 0). A discount factor for discounting from a time T back to an earlier time t at the interest rate R will be referred to using the notation indicated in Figure 3.2.

The reader is urged to become familiar with this notation as it will be used throughout the book. Since interest rates are different for different term lengths, the interest rate R is generally a function of both t and T, i.e. $R = R(t,T)$. In order to keep the notation simple, we will use the notation indicated in Figure 3.2 rather than $B_{R(t,T)}(t,T)$ on the understanding that the interest rate R refers to the rate corresponding to the times specified in the argument of the discount factor. If, as is occasionally the case, the interest rate is not dependent on the times (as in some option pricing models), we sometimes adopt the convention of denoting the constant interest rate by r and the corresponding discount factor by $B_r(t,T)$. The letter B has been used to denote the discount factor because a discount factor is nothing other than the price at time t of a zero-coupon **B**ond with a maturity T and nominal principal equal to one monetary unit (for example, 1 euro).

The discount factor yields the value of a *future* payment *today* (*discounting*). Conversely, the *future value* of a *payment today* (*compounding*) is obtained by multiplying the payment by $B_R(t,T)^{-1}$. The *interest* accrued between times t and T is thus the difference between the compounded value and the original amount, i.e. the original amount multiplied by the factor $(B_R(t,T)^{-1} - 1)$.

3.1.4 Compounding Methods

The explicit form of the discounting and compounding factors and of the interest accrued are shown for the four most common *compounding methods* in Table 3.4. At each stage in this book, the results expressed in the general notation given in the first line of Table 3.4 can be converted directly into the explicit expressions of the desired compounding method by replacing the general expression with the appropriate entries in Table 3.4.

The factors in Table 3.4 are obtained from intuitive considerations which are now described in detail for each compounding method. Note that interest rates are always quoted *per time*, for example 6% *per annum*, 1% *per month* or 0.02% *per day,* etc.

	Discount Factor	Compounding Factor	Interest Accrued
General	$B_R(t,T)$	$B_R(t,T)^{-1}$	$B_R(t,T)^{-1}-1$
Continuous	$e^{-R(t,T)}$	$e^{R(t,T)}$	$e^{R(t,T)}-1$
Discrete	$(1+R)^{-(T-t)}$	$(1+R)^{(T-t)}$	$(1+R)^{(T-t)}-1$
Simple	$[1+R(T-t)]^{-1}$	$1+R(T-t)$	$R(T-t)$
Linear	$1-R(T-t)$	$1+R(T-t)$	$R(T-t)$

Table 3.4: Interest rate factors in general notation and their specific form for the four most commonly used compounding methods.

Simple Compounding

For *simple compounding* the interest paid at the end of the term agreed upon is calculated by simply applying the rule of three: if an interest rate R per time unit has been agreed upon and the interest period $(T-t)$ spans n time units, the interest payable is simply the product n multiplied by R. If the period $(T-t)$ is measured in the same *unit* (e.g. years) as that used to quote the interest rate (e.g. interest rate is quoted per annum) then we simply have $n=(T-t)$ and the interest paid simply equals $R(T-t)$ as indicated in Table 3.4. If capital in the amount K_0 has been invested, the interest earned resulting from simple compounding is obtained by multiplying K_0 by $R(T-t)$. The capital held at maturity from an investment paying, for example, 6% per annum over a quarter of a year is thus

$$K = K_0(1+6\frac{\%}{\text{year}} * 0.25\text{year}) = K_0(1+6\% * 0.25) = K_0\left[1+R(T-t)\right].$$

The compounding factor is thus $1+R(T-t)$. The discount factor is the reciprocal of the compounding factor. Simple compounding is often used in money markets where interest periods $(T-t)$ are usually less than one year.

Discrete Discounting

Discrete compounding is used when the interest period $T-t$ is longer than the time unit for which the interest rate is quoted; for example, if $T-t=3$ years and the interest rate is given *per annum* (for example 6% per annum). The calculation proceeds as if at the end of each whole time unit (in this example, one year) the interest accrued is paid out and the entire sum (invested capital plus paid interest) is then reinvested at the same interest rate for the next time unit. The interest paid after the end of a single time unit is computed in accordance with the simple compounding convention. For example if the capital K_0 is invested for 3 years earning interest at a rate of 6% per annum, after 1 year the investor is in possession of an amount

$$K_1 = K_0(1+\underbrace{6\frac{\%}{\text{year}} * 1\text{year}}_{\text{simple compounding}}) = K_0(1+6\%) = K_0(1+R)$$

This capital is reinvested over the next time unit (the second year) in exactly the same manner. Thus, after two years the investor's capital has increased to

$$K_2 = K_1(1+R) = K_0(1+R)(1+R) = K_0(1+R)^2$$

In the second year, interest is accrued on the interest earned in the first year in addition to the initial investment capital. This is called *compounded interest.* Likewise, in the third year, the investor's capital is given by

$$K_3 = K_2(1+R) = K_0(1+R)^3.$$

and so on. The compounding factor thus obtained is $(1+R)^n$ where n is the number of time units which have passed during the interest period. Again, n is simply equal to $T-t$ if the time span is measured in the same unit (e.g. years) as is used to quote the interest rate (e.g. per annum). In this case the compounding factor is $(1+R)^{T-t}$ as in Table 3.4. Here, n is not necessarily a whole number. The calculation remains the same for $T-t = 3.5$, for example.

Of course there is nothing special about the choice of exactly one time unit (in our example one year) as the period after which interest is paid (and reinvested). We can do the same calculation when interest (each time calculated using simple compounding) is paid (and reinvested at the same rate) not only once in the course of a time unit (in our example, in the course of a year) but m times (e.g. quarterly or monthly). For example, if the interest payments are made quarterly (i.e. $m = 4$), the capital after the first payment, i.e. after a quarter of a year, at an interest rate of 6% per annum is given by

$$K_0(1+\underbrace{6\frac{\%}{\text{year}} * \frac{1}{4}\text{year}}_{\text{simple compounding}}) = K_0(1+\frac{1}{4}6\%) = K_0(1+\frac{1}{m}R)$$

Analogously after half a year

$$\underbrace{K_0(1+\frac{1}{m}R)}_{\text{captial after the first payment}} * (1+\underbrace{6\frac{\%}{\text{Jahr}} * \frac{1}{4}\text{Jahr}}_{\text{simple compounding}}) = K_0(1+\frac{1}{m}R)^2$$

And finally after an entire year, i.e. after the end of one unit of time with respect to which the interest rate is quoted

$$K_1 = K_0(1+\frac{1}{m}R)^m$$

Now, using this capital K_1 as the start capital at the beginning of the second year, an analogous argument can be used to obtain the following expression for the capital earned by the end of the second year

$$K_2 = K_1(1+\frac{1}{m}R)^m = K_0(1+\frac{1}{m}R)^{2m}$$

and in general after n years

$$K_n = K_0(1+\frac{1}{m}R)^{nm}$$

If the time period is measured in the same time unit used to quote the interest rate then $T - t = n$ (even for non-integer valued n) and we thus obtain

$$K_{(T-t)} = K_0(1 + \frac{1}{m}R)^{(T-t)m}$$

Continuous Compounding

In the case of *continuous compounding*, the calculation is performed as if interest payments were made after each *infinitesimal* small time increment (each payment calculated using simple compounding) with the accumulated interest being immediately reinvested at the same rate. The total capital thus accumulated on an investment over a period of $T - t$ time units is then

$$K_{(T-t)} = \lim_{m\to\infty} K_0(1 + \frac{1}{m}R)^{(T-t)m} = K_0 e^{R(T-t)} \tag{3.2}$$

The compounding factor is thus given by $e^{R(T-t)}$as indicated in Table 3.4. Here, *Euler's number* e, also called the *natural number*, arises. Its value is approximately

$$e = 2.718281828459$$

Euler's number to the power of some number x is called the *exponential function,* which is defined by the limit

$$\exp(x) := e^x := \lim_{m\to\infty}(1 + \frac{x}{m})^m$$

Linear Compounding

Linear compounding is justified for very short periods of time $T-t$. For such times, the product $R(T-t)$ is also very small. For example, if $R = 3\%$ per annum and the time to maturity $T-t$ is one month $= 0.083$ years, the product $R(T-t) = 0.0025$. The square of this product is considerably smaller, namely $R(T-t)^2 = 0.00000625$. Thus, in the case of linear compounding, only terms of order $R(T-t)$ are of any importance, i.e. all non-linear terms are simply neglected. If we represent the discount factor, which is always the inverse of the compounding factor, as a geometric series[1] neglecting all terms of higher order, we obtain the discount factor given in Table 3.4:

$$[1 + R(T-t)]^{-1} \approx \underbrace{1 - R(T-t)}_{\text{linear terms}} + \underbrace{(R(T-t))^2 \pm \cdots}_{\text{higher order terms are neglected!}}$$

[1]The expansion used here is $(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - x^5 \pm \cdots$. Such series expansions can be found in any book of mathematical formulas. The result is now obtained by substituting $R(T-t)$ for x.

start date Feb-15-00
end date Dez-31-00 **discount factor 0.9571**

	time period		zero rate			
	days	years	linear	simple	discrete	contin.
Act/act	320	0.87431694	4.91%	5.13%	5.14%	5.01%
Act/365f	320	0.87671233	4.89%	5.11%	5.13%	5.00%
Act/360	320	0.88888889	4.82%	5.04%	5.06%	4.93%
30/360	316	0.87777778	4.89%	5.11%	5.12%	4.99%
30/E360	315	0.875	4.90%	5.12%	5.14%	5.01%

Figure 3.3: Interest rates for the same discount factor based on different day count and compounding conventions.

Convention-Dependent Interest Rates

As different as the formulas for the discount factors in Table 3.4 may look: we emphasize that it is *not* the *numerical values* of the discount factors that are dependent on the compounding convention, but rather the *interest rates* themselves! After all, today's value of a monetary unit paid in the future must be independent of the convention used for discounting. All convention effects are "absorbed" in the interest rate. Requiring the discount factors for the different compounding methods as given in Table 3.4 to be numerically equal enables the conversion of interest rates from one convention into another:

$$e^{-R_{\text{continuous}}(T-t)} = \frac{1}{(1+R_{\text{discrete}})^{T-t}} = \frac{1}{1+R_{\text{simple}}(T-t)} = 1 - R_{\text{linear}}(T-t)$$

For example, the interest rate necessary to generate a discount factor in discrete discounting with the same numerical value as a given discount factor in continuous compounding is

$$R_{\text{discrete}} = e^{R_{\text{continuous}}} - 1 \quad \Longrightarrow \quad R_{\text{continuous}} = \ln(1 + R_{\text{discrete}})$$

However, not only the effects of the compounding methods but also the effects of the day count convention are absorbed in the interest rates. This is demonstrated in the accompanying Excel-sheet USANCE.XLS and in Example 3.3. The interest period is the same as in Example 3.1 but the different day count conventions generate different time lengths (measured in years). The discount factor must be the same for all conventions. The interest rates associated with this single discount factor, however, are strongly influenced by both the day count convention as well as the compounding convention. They vary between 4.82% and 5.14%.

As already mentioned, all of the conventions introduced here are actually unnecessary for understanding financial instruments. In order to concentrate on the essentials, discount factors rather than interest rates will be predominantly used in this book. For any concrete calculation, the reader should be able

- to write down explicitly the required discounting factor using Table 3.4 and

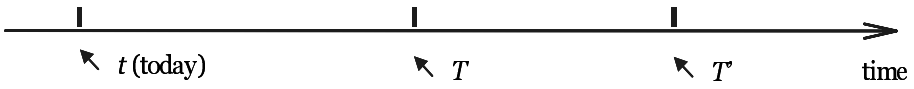


Figure 3.4: The sequence of the times t, T and T'

- to calculate the time length $T - t$ using Equation 3.1 (along with the Excel-sheet USANCE.XLS)
- after the exact dates t and T in the appropriate business day convention as specified in Table 3.2 have been determined.

In what follows, we will therefore work only with the general discount factor given in Figure 3.2.

3.1.5 Spot Rates

Spot rates are the *current* yields on securities which generate only one single payment (cash flow) upon maturity. A zero coupon bond is an example of such a security as are coupon bonds whose last coupon payment prior to maturity has already been made. The spot rates as a function of time to maturity T is called the *spot rate curve* or the *term structure*. These spot rate curves can be represented by the discount factors $B_R(t, T)$.

3.1.6 Forward Rates

Forward rates are the *future* spot rates from today's point of view which are consistent with the current spot rates in the sense of the following arbitrage argument: One monetary unit is to be invested today (time t) until a specified maturity date T' . The investor can consider the following two investment strategies:

1. Invest the monetary unit without adjusting the position until the maturity date T', or
2. Invest the monetary unit until some time T where $T < T'$, and at time T immediately reinvest the interest earned together with the original amount until the maturity date T'.

If the investor were able to fix *today* the interest rate for the future period between T and T' (see Figure 3.4) then the *total return* for both strategies must be the same. If this were not the case, the investor would have an opportunity to earn a profit without risk, i.e. an *arbitrage* opportunity (see Section 7). If the yield from the first strategy were higher, an investor could raise capital according to the second strategy and invest this capital according to the first. Financing the investment using the second strategy requires less interest than the total return on investing in the first. Thus, since all interest rates are fixed at time t, a profit would have been made without risk

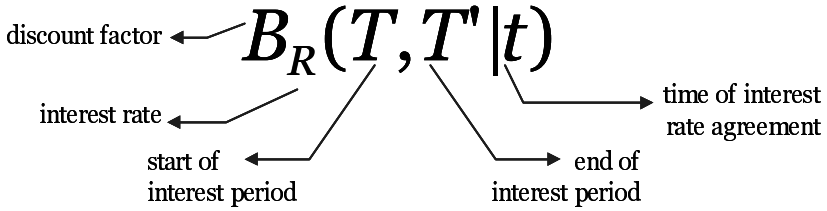


Figure 3.5: The general forward discount factor for the time between T and T' as seen at time t.

or investment capital at time t. Conversely, if the total return of the first strategy is lower than that of the second, the investor could finance the second investment strategy according to the terms of the first. Again, this gives the investor an arbitrage opportunity.

The interest rate fixed at time t for the future time period between T and T' which eliminates any possible arbitrage opportunity, i.e. which is consistent with the spot rates at time t, is called the forward rate. Analogous to the general notation for discount factors for the spot rates, the notation specified in Figure 3.5 will be used to denote the general form of the discount factor corresponding to the forward rates. This discount factor is also called the *forward discount factor*. As with spot rate discount factors, an implicit notation is used: the forward rate R depends on all time arguments of the discount factor, $R = R(T, T'|t)$. For the sake of simplifying the notation, "$B_R(T, T'|t)$" will be used instead of "$B_{R(T,T'|t)}(T, T'|t)$".

The condition eliminating the arbitrage opportunity can be summarized as follows: the compounding factor using the *spot* rate from t to T' must be equal to the compounding factor of the spot rate from t to T multiplied by the compounding factor of the forward rate from T to T'. The compounding factors are the inverse of the corresponding discount factors as indicated in Table 3.1. Consequently, the following equality must hold to prevent arbitrage:

$$\underbrace{B_R(t,T')^{-1}}_{\text{Spot}} = \underbrace{B_R(t,T)^{-1}}_{\text{Spot}} \underbrace{B_R(T,T'|t)^{-1}}_{\text{Forward}}$$

or equivalently,

$$\underbrace{B_R(t,T')}_{\text{Spot}} = \underbrace{B_R(t,T)}_{\text{Spot}} \underbrace{B_R(T,T'|t)}_{\text{Forward}} \quad \forall\ t \le T \le T' \tag{3.3}$$

The discount factors for the forward rates can thus be uniquely determined from the discount factors of the appropriate current spot rates:

$$B_R(T,T'|t) = \frac{B_R(t,T')}{B_R(t,T)} \qquad \forall\ t \le T \le T' \tag{3.4}$$

The last two equations are quite fundamental. They will be used repeatedly in what follows to simplify formulas involving terms with products and quotients of dis-

count factors. With the aid of these equations, many essential properties of financial instruments depending on interest rates can be derived without ever having to specify the compounding convention. This helps to clarify questions as to whether certain properties under consideration are intrinsic properties of the instruments themselves or merely results of the compounding convention being used. For example, the difference between *Macaulay* Duration and *Modified* Duration (see Section 6) is not an inherent property of bonds but merely an effect resulting from applying a particular compounding convention. In the case of continuous compounding, for example, there is no difference between the two!

At this point we provide an example demonstrating the usefulness of Equation 3.3 and Table 3.4. We give explicitly the forward rates as a function of the spot rates for each of the four common compounding conventions (for the sake of clarity, the time dependence of the interest rates is indicated in the formulas):

$$R(T,T'|t) = \begin{cases} \frac{R(t,T')(T'-t)-R(t,T)(T-t)}{T'-T} & \text{Continuous Compounding} \\ \frac{\left[1+R(t,T')\right]^{\frac{T'-t}{T-t}}}{\left[1+R(t,T')\right]^{\frac{T-t}{T'-t}}} - 1 & \text{Discrete Compounding} \\ \left[\frac{1+R(t,T')(T'-t)}{1+R(t,T)(T-t)} - 1\right] / (T'-T) & \text{Simple Compounding} \\ \left[1 - \frac{1-R(t,T')(T'-t)}{1-R(t,T)(T-t)}\right] / (T'-T) & \text{Linear Compounding} \end{cases}$$

This clearly demonstrates the advantage of using the general notation for discount factors 3.4 introduced above. In Example 3.6, the forward rates are calculated from the spot rates taken from the Excel sheet PlainVanilla.xls (see accompanying CD-ROM) using the above formulas for annual, discrete compounding over a period of 15 years.

3.2 Market Prices

Let $S(t)$ be the *spot price* at time t of a stock, a commodity, or a currency. The *dividend-adjusted* spot price $\widetilde{S}(t,T)$ at time t is the price net of the value of all dividends paid between the times t and T. It is given by

$$\widetilde{S}(t,T) = \begin{cases} S(t) & \text{no dividend} \\ S(t) - D(t')B_R(t,t') & \text{dividend payment } D \text{ due at time } t' \\ S(t)B_q(t,T) & \text{dividend yield } q \end{cases} \tag{3.5}$$

The *dividend adjustment* is thus accomplished by subtracting the value of dividends, discounted back to time t at the spot rate R, from the spot price, or – in the case of a dividend yield q – discounting the spot price from T back to t using the dividend yield q.

For currencies, q represents the risk-free interest rate in the foreign currency, for commodities, it is the difference between the *convenience yield* and the *cost of carry* (expressed as a yield, see later). When considering stock indices, the dividend payments

Maturity	Spot Rate	for	Forward Rate 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	2.20%	in	Year	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years
2	2.50%	1	2.80%	3.25%	3.67%	4.08%	4.36%	4.65%	4.83%	5.02%	5.20%	5.28%	5.37%	5.45%	5.49%	5.52%	5.54%
3	2.90%	2	3.70%	4.11%	4.51%	4.76%	5.03%	5.18%	5.34%	5.51%	5.56%	5.63%	5.70%	5.72%	5.74%	5.74%	
4	3.30%	3	4.51%	4.91%	5.11%	5.36%	5.47%	5.61%	5.77%	5.80%	5.84%	5.90%	5.90%	5.91%	5.90%		
5	3.70%	4	5.32%	5.41%	5.65%	5.71%	5.83%	5.98%	5.98%	6.01%	6.06%	6.04%	6.04%	6.02%			
6	4.00%	5	5.51%	5.82%	5.85%	5.96%	6.11%	6.10%	6.11%	6.15%	6.12%	6.11%	6.08%				
7	4.30%	6	6.12%	6.01%	6.11%	6.26%	6.21%	6.21%	6.24%	6.20%	6.18%	6.14%					
8	4.50%	7	5.91%	6.11%	6.31%	6.24%	6.23%	6.26%	6.21%	6.18%	6.14%						
9	4.70%	8	6.31%	6.52%	6.35%	6.31%	6.33%	6.26%	6.22%	6.17%							
10	4.90%	9	6.72%	6.36%	6.31%	6.33%	6.25%	6.21%	6.15%								
11	5.00%	10	6.01%	6.11%	6.21%	6.13%	6.10%	6.05%									
12	5.10%	11	6.21%	6.31%	6.17%	6.13%	6.06%										
13	5.20%	12	6.41%	6.15%	6.10%	6.02%											
14	5.25%	13	5.90%	5.95%	5.90%												
15	5.30%	14	6.00%	5.89%													
16	5.33%	15	5.78%														

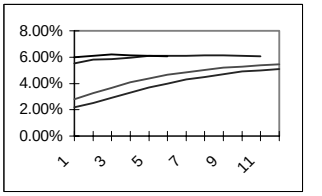


Figure 3.6: Forward rates for periods starting in $T = 1, 2, \ldots, 15$ years for terms $T' - T = 1, 2, \ldots, 15$ years. From line to line the start points T of the forward periods change by one year. From column to column the lengths $T' - T$ of the forward periods change by one year. The inset graphic shows, from top to bottom, the current term structure along with the forward term structures in 1, 5 and 10 years.

from the assets in the index are commonly averaged out to result in a dividend yield q of the index rather than considering each dividend of each stock as an individual payment. However, *performance indices*, such as the German *DAX*, which require that dividends be reinvested in the index should not be adjusted for dividend payments[2].

3.3 An Intuitive Model for Financial Risk Factors

3.3.1 Random Walks as the Basis for Pricing and Risk Models

The random walk is a mathematical model which is frequently used to characterize the random nature of real processes. The concept of the random walk has attained enormous importance in the modern financial world. Most option pricing models (such as the Black-Scholes model) and several methods used in modern risk management (for example, the variance-covariance method) and, of course, Monte Carlo simulations are based on the assumption that market prices are in part driven by a random element which can be represented by a random walk. It is therefore worthwhile to acquire an understanding of random walks if only to develop an intuitive comprehension of *stochastic processes* which are at the heart of this book.

A *random walk* can be described as follows: starting from some point in space, we travel a random distance in a randomly selected direction. Having arrived at a new point, another such step of random length and direction is taken. Each individual step in the procedure has a length and direction and thus can be represented as a vector

[2]Except for effects caused by taxation (corporate tax).

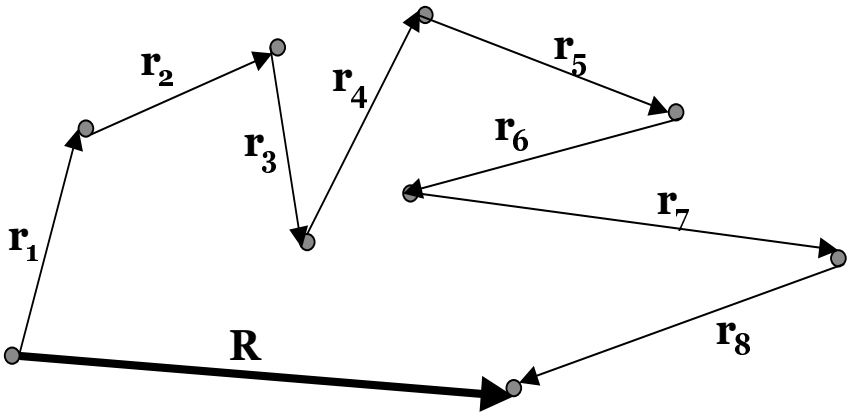


Figure 3.7: A random walk in two dimensions.

as shown in Figure 3.7. The completed random walk is a series of such vectors. Each base point of a vector is the end point of its predecessor.

At this point, we ask the following important question: what is the distance from the starting point after having completed a random walk consisting of n steps[3]? In other words: how large is the "end-to-end distance" represented by the length of the vector $\mathbf{R}$ in Figure 3.7? The length and direction of the vector $\mathbf{R}$ are random since $\mathbf{R}$ is the sum of random steps. As a result, only statistical statements are available to describe the properties of this vector. For example, the length of this vector cannot be determined with certainty but we could calculate its mean length. This requires a large number of random walks with the same number of steps. For each of these random walks, the square of the Euclidean norm of the end-to-end vector $\mathbf{R}$ is determined and used to calculate the mean $\langle \mathbf{R}^2 \rangle$. The mean end-to-end distance is than defined as the square root of this value. A Monte Carlo simulation (see Section 12) could be carried out to generate the random walks and obtain an estimate for the mean $\langle \mathbf{R}^2 \rangle$ by measuring $\mathbf{R}^2$, the square of the end-to-end vector, of each simulated random walk and then take the average of these. In doing so, it can be observed that the square of the end-to-end vector is, on average, proportional to the number of steps taken in the random walk[4].

$$\mathrm{E}\left[\mathbf{R^2}\right] \approx \langle \mathbf{R^2} \rangle \propto n$$

Here $\mathrm{E}[x]$ denotes the expectation of a random variable x and $\langle R^2 \rangle$ denotes the mean of the squared Euclidean norm of $\mathbf{R}$. This holds irrespective of the dimension of the space in which the random walk occurs. The same result holds for a random walk on a line, in a plane as in Figure 3.7 or in a 15-dimensional Euclidean space [42].

The expectation of the end-to-end vector itself is equal to zero, i.e. $\mathrm{E}[\mathbf{R}] = 0$. This is immediate since the end-to-end vector $\mathbf{R}$ points in any direction with equal likelihood

[3] Figure 3.7 for instance shows a random walk with $n = 8$.
[4] The symbol "$\propto$" means "is proportional to".

and therefore all these vectors cancel each other out in the average. In consequence, the variance of the end-to-end vector is given by [5]:

$$\mathrm{var}(\mathbf{R}) \equiv \mathrm{E}\left[(\mathbf{R}-\mathrm{E}[\mathbf{R}])^2\right] = \mathrm{E}\left[\mathbf{R^2}\right] \propto n$$

The variance of the end-to-end vector is thus also proportional to n. The standard deviation, defined as the square root of the variance, is consequently proportional to$\sqrt{n}$. This is the reason why the uncertainty in future market prices increases proportionally to the square root of time since, as will be shown in Section 3.3.2, the time corresponds to the number of steps when a random walk is used to model and interpret price movements. Many well-known rules in financial mathematics, for example, that the overnight value at risk is to be multiplied by a factor consisting of the square root of the liquidation period or that the conversion of daily volatilities into monthly volatilities is accomplished by multiplying by the square root of the number of days in the month, have their origin in the fact that the variance of the end-to-end vector of a random walk is proportional to the number of random steps!

The relation $\mathrm{E}[\mathbf{R^2}] \propto n$ is so fundamental, that it is worthwhile to understand its theoretical meaning. As mentioned above, a random walk consists of a series of 'step' vectors. According to the rules of vector addition, the end-to-end vector is merely the sum of these vectors: $\mathbf{R} = \sum_{i=1}^{n} \mathbf{r_i}$. Since each individual step vector, $\mathbf{r_i}$ points in any direction with the same likelihood, the expectation of each step is $\mathrm{E}[\mathbf{r_i}] = 0$ just as for the end-to-end vector. Since each of the steps are independent of one another (in particular, they are uncorrelated), we have

$$\mathrm{E}[\mathbf{r_i} \cdot \mathbf{r_j}] = \underbrace{\mathrm{E}[\mathbf{r_i}]}_{=0} \cdot \underbrace{\mathrm{E}[\mathbf{r_j}]}_{=0} = 0 \quad \forall\, i \neq j, i, j = 1, ..., n.$$

With this information about the individual steps, we immediately obtain

$$\mathrm{E}\left[R^2\right] = \mathrm{E}\left[\sum_{i=1}^{n} \mathbf{r}_i \cdot \sum_{j=1}^{n} \mathbf{r}_j\right] = \sum_{i,j=1}^{n} \mathrm{E}\left[\mathbf{r}_i \cdot \mathbf{r}_j\right] = \sum_{i=1}^{n} \mathrm{E}\left[\mathbf{r}_i \cdot \mathbf{r}_i\right] + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \underbrace{\mathrm{E}\left[\mathbf{r}_i \cdot \mathbf{r}_j\right]}_{0} = n\, b^2$$

where the constant b denotes the mean length of a single step:

$$b^2 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\mathbf{r_i^2}]$$

and as such is the constant of proportionality in the relation $\mathrm{E}[\mathbf{R^2}] \propto n$. At no point did the dimension enter into the above derivation. The equation $\mathrm{E}[\mathbf{R^2}] = b^2 n$ thus holds in any dimension and is an expression of a fundamental property of random walks, namely their *self-similarity*: the statistical properties of a random walk are always the same, regardless of the degree of detail with which they are observed. In other words, a step in a random walk can itself be represented as the end-to-end vector of a random walk with "smaller" steps. Likewise, the end-to-end vector of a random walk can itself be considered a single step of a "coarser" random walk (*coarse graining*).

[5] In this book we will denote the variance of a quantity x by $\mathrm{var}[x]$.

As yet, only the so called moments (expectation, variance, etc.) of the probability distribution $p(\mathbf{R})$ have received mention. However, the distribution of $\mathbf{R}$ itself can be determined as well. In general, the end-to-end vector of a random walk in a d-dimensional space has a normal distribution. The concrete expression of this fact in one dimension (the most important case for the financial applications) is given by [42]

$$p(\mathbf{R}) = \frac{1}{\sqrt{2\pi \mathrm{var}(\mathbf{R})}} \exp\left[\frac{(\mathbf{R} - \mathrm{E}[\mathbf{R}])^2}{2\mathrm{var}(\mathbf{R})}\right] \tag{3.6}$$

where

$$\mathrm{E}[\mathbf{R}] = 0 \qquad \mathrm{var}(\mathbf{R}) = b^2 n \tag{3.7}$$

Equations 3.6 and 3.7, together with the principle of self-similarity are the quintessential properties of the theory of random walks introduced here. They comprise all that is necessary for their application in the field of finance. For example, the "*normal distribution assumption*" for price movements is an immediate consequence of the theory of random walks. Contrary to popular belief, the normality of the distribution of relative price changes need not be *assumed*. It follows automatically that if market prices (or more precisely, the logarithm of the market prices, see below) behave as random walks, they must be normally distributed with a density function as given in 3.6. In financial literature, random walks are often *defined* having normally distributed individual step vectors. By doing so the "normal distribution assumption" is "injected" into the definition of the random walk. This assumption is equally unnecessary! As mentioned above, a random walk is defined as a series of completely random steps. Absolutely no assumptions have been made on the distribution of the individual steps or their sum. The fact that the end-to-end vector has a normal distribution follows *automatically* as a consequence of the central limit theorem.

3.3.2 Risk Factors as Random Walks

A *random variable* $z(t)$, whose values change randomly with time t is called a *stochastic process*. The process is called *Markov* if its future behavior is influenced solely by its value at the present time t. This means intuitively that the future behavior is independent of the path taken to reach the present value. Assuming that the current value of a risk factor, such as a stock price or an interest rate, contains all the information about its historical development (this is called *weak market efficiency*), it follows that the subsequent values taken on by such a risk factor depend only on the current price and other external effects, such as politics, but not on the past prices or rates. Market prices can then assumed to be Markov processes.

In order to derive a model for the Markov process $S(t)$ representing the time-evolution of a risk factor, we assume that the process can be split into a random and a deterministic component. The deterministic component is called the *drift*. We will begin our discussion with an analysis of the random component.

The derivation of the model describing the random component given here is fundamentally different from that which is commonly found in the related literature and is based on the general properties of the random walk, Equation 3.6 and Equation 3.7. The literature usually begins with the introduction of a model and proceeds with

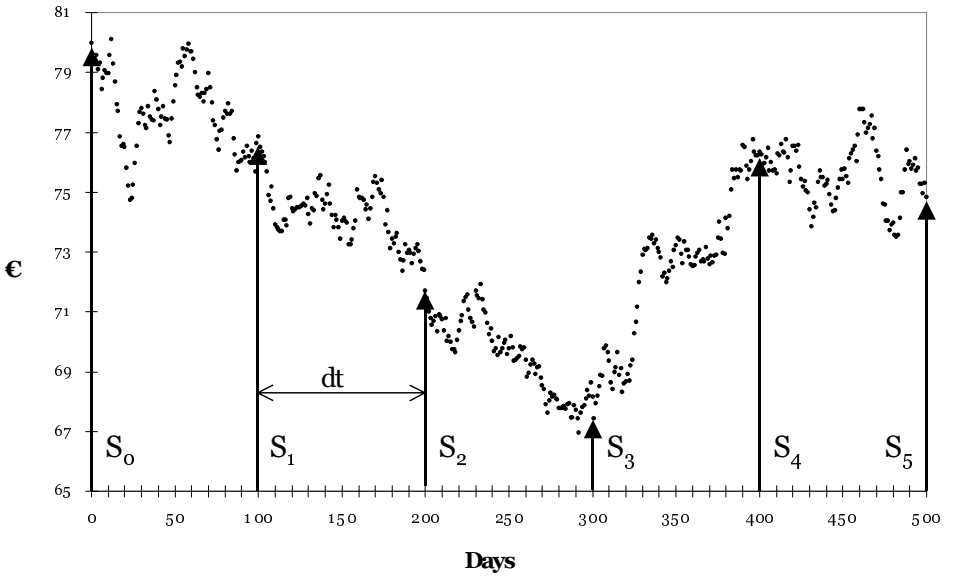


Figure 3.8: End-of-day values of a stock price over a period of 500 trading days. The values after $0, 100, 200, \ldots, 500$ days are denoted by $S_0, S_1, S_2, \ldots, S_5$.

a presentation of calculations and results following from the model assumption. It is more intuitive, however, to begin with a derivation of the model itself, i.e. to illustrate the steps and describe the motivation leading to its construction. Thus, this section serves the dual purpose of introducing a model for future analysis and an example of the modelling process. What follows is a detailed discussion of how we can construct a model describing a real phenomenon by means of abstraction (and intuition!) which is simple enough to allow a mathematical and/or computational analysis but still complex enough to capture the essential features of the real phenomenon.

A real process, for example the closing prices of a stock on 500 days, might look like the points shown in Figure 3.8. We propose to model this process. To do so, several questions must first be answered:

- *What is the fundamental idea behind the model?*
  As mentioned above, the random properties of many processes can be traced back to the general concept of the random walk. We therefore use the random walk as the basis for our model.

- *A random walk in which dimension?*
  A market price can rise or fall, i.e. can change in only two directions. A random walk in d-dimensional space allows for an upward or downward change in d linearly independent directions, i.e. there are $2d$ possible changes in direction. Thus, the dimension required for the description of just upward or downward

changes is $d = 1$.

- *Which real parameter is described by the number of steps n in the random walk?* In order to observe a change in price (in other words for a step in a random walk to be taken), one thing must occur: time must pass. If the price is observed in regular, fixed time intervals dt (for example, every 100 days as in Figure 3.8, or daily or hourly, etc.), then the amount of time passing between steps is dt. If the entire random walk occurs between t (=today) and a future date T then

$$T - t = ndt \tag{3.8}$$

Since dt is a constant, the number of steps n is proportional to the time in which the random walk occurs, i.e. proportional to $T - t$.

- *Which real parameter should be modeled by a random walk?* At first glance, we might take the market price of a risk factor. The market price evolution $S_5 - S_0$ over the entire period in Figure 3.8 can be decomposed into individual steps as follows :

$$S_5 - S_0 = (S_0 - S_1) - (S_2 - S_1) + (S_3 - S_2) + (S_4 - S_3) + (S_5 - S_4),$$

or more generally,

$$S_n - S_0 = \sum_{i=1}^{n} dS_i \quad \text{with} \quad dS_i = S_i - S_{i-1}$$

If the market price itself were a random walk, then as a result of the self similarity property, the individual steps dS_i would also be random walks. The price differences however are real cash amounts given in euros, for example. This would mean that a security costing 1000 euros would experience the same fluctuations (in euros) as one costing only 10 euros. This is surely not the case. It would make much more sense to consider *relative* fluctuations. Our next candidate for a step in our random walk could therefore be the ratio S_i/S_{i-1}. The ratio of the last price to the first is given by

$$\frac{S_5}{S_0} = \frac{S_1}{S_0}\frac{S_2}{S_1}\frac{S_3}{S_2}\frac{S_4}{S_3}\frac{S_5}{S_4},$$

or more generally,

$$\frac{S_n}{S_0} = \prod_{i=1}^{n} \frac{S_i}{S_{i-1}}$$

This is the *product* of the individual steps and not their sum. A random walk however, is a vector and as such must always be the sum of its component steps. In light of this fact, the ratios $S_i/S_{i-1}$are completely unsuitable for the steps of a random walk as they are not even vectors! However, the ratios S_i/S_{i-1}, which make economic sense, can be utilized by converting the products into sums by

taking the *logarithm* of both sides of the above equation. The *functional equation* for the logarithm is given by

$$\ln(a \times b) = \ln(a) + \ln(b) \quad , \qquad \ln(a/b) = \ln(a) - \ln(b)$$

Thus taking the logarithm of both sides of the above product yields

$$\ln(\frac{S_5}{S_0}) = \ln(\frac{S_1}{S_0}) + \ln(\frac{S_2}{S_1}) + \ln(\frac{S_3}{S_2}) + \ln(\frac{S_4}{S_3}) + \ln(\frac{S_5}{S_4}),$$

or more generally,

$$\ln(\frac{S_n}{S_0}) = \sum_{i=1}^{n} \ln(\frac{S_i}{S_0 i - 1})$$

and thus,

$$\ln(S_n) - \ln(S_0) = \sum_{i=1}^{n} d\ln(S_i) \quad \text{with} \quad d\ln(S_i) = \ln(S_i) - \ln(S_{i-1})$$

This looks exactly like our first attempt with the sole exception that the market price S has been replaced with its logarithm $\ln(S)$. This small change makes it possible to satisfy both the economic requirement that the proportional changes in the market price be modeled and the mathematical requirement that a random walk be the sum of its component steps.

We have thus completed the fundamental construction of our model:

The random component of a market price is modeled by interpreting the logarithm of the price as a one-dimensional random walk with the number of random steps being proportional to the length of time during which the random walk takes place.

Now we are in a position to apply what we know about random walks to draw conclusions about the evolution of market prices. It follows from Equation 3.6 that the end-to-end vector $\mathbf{R} = \ln(S(T)/S(t))$ is normally distributed with expectation and variance as given in Equation 3.7. Because in our model the time and the number of steps in the random walk are related as in Equation 3.8, the time dependence of the end-to-end vector's variance can be calculated as

$$\mathrm{E}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = 0 \quad , \quad \mathrm{var}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = b^2 n = \sigma^2(T-t) \tag{3.9}$$

The logarithm of the market price is normally distributed, its variance being proportional to the time. The constant of proportionality σ^2 is well defined, even in the limiting case of the time intervals dt becoming infinitesimal. The *standard deviation*, defined as the square root of the variance, is thus proportional to the square root of the time. The constant of proportionality σ is called *volatility* . Making use of Equations 3.9 and 3.8 gives us

$$\sigma^2 = \frac{b^2}{dt} = \frac{1}{T-t}\mathrm{var}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] \tag{3.10}$$

Note again that the normal distribution has not been *assumed* at any point in this argument. Nor has it been assumed that the standard deviation increases proportionally to the square root of the time. The model consists solely of the statement above, that the logarithm of the market price is modeled as a random walk. The rest follows automatically! Owing to the self-similarity property, each step of the random walk, regardless of its length can be interpreted as a random walk itself. It follows that all changes $d\ln(S)$ in the logarithm of the market price over an infinitesimally small time interval dt are also normally distributed with an expectation of zero and the following variance:

$$\begin{aligned} \text{var}\left[d\ln\left(S(t)\right)\right] &= \text{var}\left[\ln\left(S(t+dt)\right) - \ln\left(S(t)\right)\right] \\ &= \text{var}\left[\ln\left(\frac{S(t+dt)}{S(t)}\right)\right] = \sigma^2 dt \end{aligned} \tag{3.11}$$

Therefore the process $d\ln(S(t))$ is very similar to a very prominent and well understood process appearing time and again in physics, biology and other sciences, called the *Wiener process* which, for instance, describes what is known as *Brownian motion* in physics. Such a Wiener process W is a stochastic process which changes randomly by an amount dW over a time interval dt. These changes dW are normally distributed with a mean of zero and with a variance equal to the length of the time interval passed during the change, ie. with a variance equal to dt. Or written more compactly:

$$dW \sim X\sqrt{dt} \quad \text{with} \quad X \sim \text{N}(0,1) \tag{3.12}$$

Here, as always in this book, the notation $\text{N}(x,y)$ denotes the *normal distribution* with mean x and variance y, i.e. $\text{N}(0,1)$ denotes the *standard normal distribution*[6]. The sign "$\sim$" in these circumstances is to be read as "is distributed as". Equation 3.12 thus means in words: "dW is distributed as $\sqrt{dt}$ times a random number X. This random number X in turn is distributed according to the standard normal distribution." The definition 3.12 together with the property 3.11 allows us to write the process for the infinitesimal random change $d\ln(S(t))$ very compactly as

$$d\ln\left(S(t)\right) = \sigma\, dW$$

We have now arrived at the point where the discussion in the literature usually begins.

The only model parameter which has as yet been introduced is the volatility σ. This, however, is insufficient to provide an adequate description of the behavior of risk factors. For example, two parameters (denoted by u and d) are required for the binomial model dealt with in Section 11. All risk factors will be required to have an additional essential property, namely that they be martingales[7]. This subject will receive detailed discussion in Section 14. The intuitive meaning of the martingale property is that the current value of a risk factor can be represented as the appropriately discounted expected future value. In order to ensure that this property is satisfied when modeling market movements with a random walk, two parameters are required.

[6] The normal distribution and the standard normal distribution are presented in detail in Section A.4.3.

[7] At least as long as the risk factors are tradeable, see section 14.

Random Variable	$x = \ln\left(\frac{S(t+dt)}{S(t)}\right)$	$x = \left(\frac{S(t+dt)}{S(t)}\right)$
Distribution	**Normal Distribution**	**Lognormal Distribution**
Density	$\frac{1}{\sqrt{2\pi\sigma^2 dt}} e^{-\frac{(x-\mu dt)^2}{2\sigma^2 dt}}$	$\frac{1}{x\sqrt{2\pi\sigma^2 dt}} e^{-\frac{(\ln(x)-\mu dt)^2}{2\sigma^2 dt}}$
$\mathbf{P(x \leq a)}$	$\frac{1}{\sqrt{2\pi\sigma^2 dt}} \int_{-\infty}^{a} e^{-\frac{(x-\mu dt)^2}{2\sigma^2 dt}} dx$	$\frac{1}{\sqrt{2\pi\sigma^2 dt}} \int_{-\infty}^{\ln(a)} e^{-\frac{(x-\mu dt)^2}{2\sigma^2 dt}} dx$
Expectation	μdt	$e^{(\mu+\frac{\sigma}{2})dt}$
Variance	$\sigma^2 dt$	$e^{2\mu dt}(e^{2\sigma^2 dt} - e^{\sigma^2 dt})$

Table 3.5: Statistical properties of the logarithm of a risk factor and of the risk factor itself.

To provide a motivation[8] for the second parameter in our random walk, we consider the following: the end-to-end vector in our 1-dimensional random walk is given by $\mathbf{R} = \ln[S(T)] - \ln[S(t)]$. If our random walk is to serve as a model for a stock price which, in the long term, should show a positive mean return, then the expectation of the end-to-end vector should not be equal to zero but should grow with time. This is incorporated into the model by introducing a deterministic term, called the *drift*:

$$d\ln[S(t)] = \mu dt + \sigma dW \tag{3.13}$$

Since no additional random component has been introduced into the model, $\ln(S(t))$ retains the normal distribution with variance $\sigma^2(T-t)$ after the passage of time from t to T. The expectation, however, is now given by $\mu(T-t)$. The addition of the drift into our random walk has the consequence that the mean of the random walk is no longer 0 but rather $\mu(T-t)$. What consequences does this have for the *mean return*? Continuous compounding with yield R implies that a security worth $S(t)$ at time t will be worth $S(T) = S(t)e^{R(T-t)}$ after the passing of a time interval of length $T-t$. Solving this equation for R and taking expectations gives:

$$R = \frac{1}{T-t}\ln\left(\frac{S(T)}{S(t)}\right) \Rightarrow \quad \mathrm{E}[R] = \frac{1}{T-t}\mathrm{E}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = \frac{\mu(T-t)}{T-t} = \mu \tag{3.14}$$

The drift parameter μ is thus merely the mean return expressed in terms of continuous compounding.

Note that the expectation of the *logarithm* of the market price is required for the determination of the mean return not the expectation of the market price itself. A random variable whose logarithm has a normal distribution has a *lognormal distribution*. The price $S(t)$ is thus lognormally distributed. The most important properties of $\ln(S)$ and S summarized in Table 3.5, see also Section A.4.

[8] This is only a motivation based on the example and not a general explanation for a drift. For example, when considering an interest rate as a risk factor, there is no reason to believe that the change in the interest rate should show a "mean return". The more general and profounder explanation for the drift is that the risk factors must modeled in such a way that there exists a probability measure in which all (tradeable and properly normalized) financial instruments are martingales.

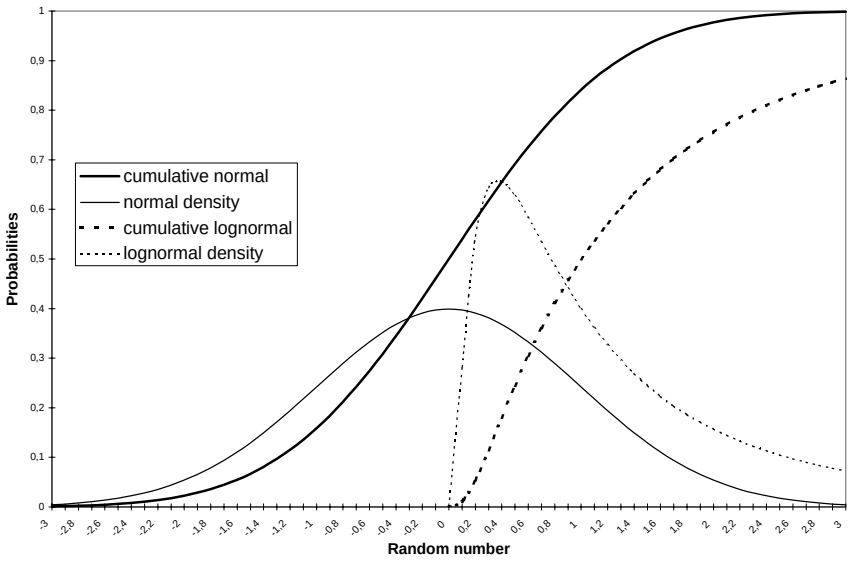


Figure 3.9: The distribution from Table 3.5 with $\mu dt = 0$ and $\sigma\sqrt{dt} = 1$. With these parameters the normal distribution has mean 0 and variance 1 while the lognormal distribution has mean $\sqrt{e} \approx 1,65$ and variance $e^2 - e \approx 4,67$.

In Figure 3.9, the density functions and cumulative probabilities of the normal and lognormal distribution are shown. Here, the parameters μdt and $\sigma\sqrt{dt}$ have been set equal to 0 and 1, respectively.

The random walk model just derived can, of course, be generalized by allowing for non-constant drifts and volatilities. If σ is expressed as in Equation 3.10 in terms of the variance and if the yield is left in its most general form as a function of time t and the market price $S(t)$ at time t, Equation 3.13 can be expressed in the following generalized form:

$$d\ln(S(t)) = \mu(S(t), t)dt + X\sqrt{\text{var}[d\ln(S(t))]} \quad \text{with } X \sim \text{N}(0,1) \tag{3.15}$$

This equation is now in the general form of a stochastic diffusion process as given by Equation 3.16 and is the starting point for more general stochastic models for the market parameters. These will receive a detailed treatment in the following chapter. Concrete examples of such generalized stochastic processes can be found in Section 30.1.

3.4 Ito Processes and Stochastic Analysis

In the previous section, risk factors in financial markets were introduced and a motivation and derivation of an intuitive model (the random walk) was provided to describe them. Now, this section will be devoted to the more theoretical fundamentals underlying these concepts, namely to stochastic analysis. *Stochastic analysis* is the branch of mathematics dealing with the investigation of stochastic processes. Particularly close attention will be paid to results which find application in finance. This subject is naturally quite theoretical, but should give the interested reader a deeper understanding of relationships between different fundamental concepts in financial mathematics. Nevertheless, the reader who is less interested in the mathematical details may skip over the rest of Chapter 3 and continue directly on to Chapter 4.

3.4.1 General Diffusion Processes

All stochastic processes in this book which will be used to model risk factors satisfy – as long as there is only one single random variable involved – an equation of the following form:

$$dS(t) = a\,(S,t)\,dt + b(S,t)\,dW \qquad \text{with} \quad dW \sim X\sqrt{dt}\ , \quad X \sim \mathrm{N}(0,1) \tag{3.16}$$

Here, W denotes – as always – a Brownian motion and X a standard normally distributed random variable. Processes satisfying an equation of this type are called *diffusion processes* or *Ito processes*. These quite general stochastic processes have long been the subject of research in the field of stochastic analysis.

The parameters $a(S,t)$ and $b(S,t)$ are called the drift rate and the volatility of the Ito process. They may depend on the time t, on the stochastic process S or on both. The interpretation of the variable S depends on the particular application under consideration. In the simple models derived in Chapter 3.3, for example in Equation 3.13, the logarithm of the risk factor was modeled as the stochastic variable. These models are of the form 3.16 with the stochastic variable being given by $\ln(S)$, where $b(S,t) = \sigma$ and $a\,(S,t) = \mu$.

The first moments of the conditional probability distribution of the general Ito process are

$$\begin{aligned}
\mathrm{E}\,[dS(t)] &= a(S,t)\,dt \\
\mathrm{var}\,[dS(t)] &= \mathrm{E}\left[(dS - \mathrm{E}\,[dS])^2\right] = \mathrm{E}\left[(b(S,t)\,dW)^2\right] = b(S,t)^2 dt \\
\mathrm{E}\left[dS(t)^2\right] &= \mathrm{E}\left[(a(S,t)dt + b(S,t)dW)^2\right] \approx b(t,S)^2 dt + \text{higher order terms in } dt
\end{aligned}$$

3.4.2 Ito's Lemma

A very important question is which stochastic differential equation is satisfied by a function $f(S)$ of a stochastic variable S, i.e. what change df is induced by an infinitesimal change dS in S. This question is answered by Ito's famous lemma, the proof

of which will be sketched in this section. Typically, when we are interested in small changes in a function f, a *Taylor series* expansion is calculated for df. The only subtlety in our case (and this is critical) is that, because of Equation 3.16, a *stochastic* variable is now involved:

$$\begin{aligned} df(S,t) &= \frac{\partial f}{\partial S}dS + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}(dS)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 + \cdots \\ &= \frac{\partial f}{\partial S}\underbrace{[a\,(S,t)\,dt + b(S,t)\,dW]}_{dS} + \frac{\partial f}{\partial t}dt \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\left[b(S,t)^2\,dW^2 + 2a\,(S,t)\,b(S,t)\,dW\,dt\right] + O(dt^2) \end{aligned}$$

In this case we can *not*, as is usually the case, neglect the non-linear terms since the Brownian motion is of the order $dW \sim \sqrt{dt}$ and thus dW^2 is *not* small in comparison with dt. In fact, it can be shown that the square of Brownian motion has the following properties:

$$\begin{aligned} \mathrm{E}\left[dW^2\right] &= dt \\ \mathrm{var}\left[dW^2\right] &\sim dt^2 \end{aligned}$$

Thus, the variance of dW^2 vanishes in the linear order of dt. This means that, in linear order, dW^2 is no longer stochastic but can be set equal to its expectation, i.e.

$$dW^2 \approx dt \tag{3.17}$$

This is the central idea of Ito's lemma. Making this substitution and *then* ignoring all terms of order greater than dt, making use of the fact that $dW \sim \sqrt{dt}$ immediately gives the equation for infinitesimal (limit $dt \to 0$) changes in the function $f(S,t)$

$$df(S,t) = \left[\frac{\partial f}{\partial S}a\,(S,t) + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}b(S,t)^2\right]dt + \frac{\partial f}{\partial S}\,b(S,t)\,dW \tag{3.18}$$

This is Ito's famous lemma. This equation has the same structure as Equation 3.16. The function f is thus also an Ito process with its drift rate and volatility given by

$$\begin{aligned} a_f(S,t) &= \frac{\partial\, f}{\partial S}a(S,t) + \frac{\partial\, f}{\partial\, t} + \frac{1}{2}b(S,t)^2\frac{\partial^2 f}{\partial\, S^2} \\ b_f(s,t) &= \frac{\partial\, f}{\partial\, S}b(s,t) \end{aligned}$$

Equation 3.18 appears rather complicated. In practice, it is sufficient to remember the essential idea, namely Equation 3.17. The equation can then be derived from a "mechanical" application of Taylor's theorem up to the linear term in dt and quadratic term in dW and making the substitution $dW^2 = dt$. A useful form of Equation 3.18 can be obtained by observing that the terms involving $\frac{\partial f}{\partial S}$ can be collected and expressed in terms of the process dS in accordance with Equation 3.16 :

$$df(S,t) = \left[\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}b(S,t)^2\right]dt + \frac{\partial f}{\partial S}dS(t) \tag{3.19}$$

The Process for the Risk Factor Itself Ito's lemma is a powerful tool which will be applied frequently in this book. For example, we can immediately obtain the process for the market price from the process describing the *logarithm* of the market price given by Equation 3.13. To do so, we define the stochastic variable[9] $y = \ln(S(t))$ and choose for f the function given by $f(y,t) = e^{y(t)}$. The original process $y(t) = \ln(S(t))$ satisfies the Equation 3.13, i.e.

$$dy(t) = \mu dt + \sigma\, dW$$

The required derivatives of f can be easily calculated and applying Ito's lemma we obtain

$$f(y,t) = e^y \quad \Rightarrow \quad \frac{\partial f}{\partial y} = f \quad , \quad \frac{\partial f}{\partial t} = 0 \, , \quad \frac{\partial^2 f}{\partial y^2} = f$$

$$df(y,t) = \left(f(y,t)\mu + \frac{1}{2}\sigma^2 f(y,t) \right) dt + f(y,t)\,\sigma\, dW$$

Setting $S(t) = f(y,t)$ yields:

$$dS(t) = S(t) \left(\mu + \frac{\sigma^2}{2} \right) dt + S(t)\sigma\, dW \tag{3.20}$$

In the literature, an equivalent approach is sometimes taken to arrive at this equation. A random walk model for the risk factor $S(t)$ is introduced in the first place (and not for its logarithm as done here) with a subsequent application of Ito's lemma to derive the process for $\ln(S(t))$. Explicitly, one begins with the process

$$dS(t) = S(t)\widetilde{\mu}\, dt + S(t)\sigma\, dW \tag{3.21}$$

obtaining from Ito's lemma[10]

$$d\ln(S(t)) = (\widetilde{\mu} - \frac{\sigma^2}{2})\, dt + S(t)\sigma\, dW \tag{3.22}$$

This result corresponds to Equations 3.20 and 3.13 in our development with a somewhat modified drift.

$$\widetilde{\mu} = \mu + \frac{\sigma^2}{2} \tag{3.23}$$

The model is thus exactly the same, only the drift must be reinterpreted. In our derivation, the drift μ can directly be interpreted as the mean return[11] of the stock price S (see Equation 3.14), whereas an intuitive interpretation of $\widetilde{\mu}$ is plainly more difficult.

The formulation of 3.20 and 3.21, in which the risk factor (and not its logarithm) is directly modeled as a stochastic process, of course also has the form of a diffusion process as in Equation 3.16, with the choice of a and b given by $b(S,t) = S(t)\sigma$ and $a\,(S,t) = S(t)\widetilde{\mu}$ or equivalently, $a(S,t) = S(t)(\mu + \sigma^2/2)$.

[9] Instead of S, another letter is used for the stochastic variable here in order to avoid confusion in the notation.

[10] Equation 3.18 with $f(S,t) = \ln(S(t))$

[11] In anticipation of later chapters note that μ is the mean return in the *real* world (as opposed to the risk-neutral world).

The Process for the Risk Factor over a Finite Time Interval

Equation 3.16 describes the *infinitesimal* change in S and thus determines the *differential* of S. We are therefore dealing with a (partial) *differential equation*. Because it contains the stochastic component dW, it is referred to as a *stochastic partial* differential equation, often abbreviated by *SPDE*. Special cases such as Equation 3.20, for example, are SPDEs as well.

With the aid of Ito's lemma and the general diffusion process, Equation 3.16, an equation for *finite* changes in S (over a finite, positive time span δt) can be derived by solving the SPDE 3.20 (which holds for *infinitesimal* changes dS). For this purpose, we use the process 3.16with $a\,(y,t) = 0$ and $b(y,t) = 1$, i.e. simply[12] $dy(t) = dW(t)$. Now we construct a function S of the stochastic variable y by

$$S(y,t) := S_0 \exp\left(\mu t + \sigma y\right)$$

where $y(t) = W(t)$ is the value of the Wiener process at time t, and S_0 is an arbitrary factor. Ito's lemma gives us the process for S induced by the process dy:

$$\begin{aligned} dS &= \left[\underbrace{\frac{\partial S}{\partial y}}_{\sigma S} \underbrace{a(y,t)}_{0} + \underbrace{\frac{\partial S}{\partial t}}_{\mu S} + \frac{1}{2} \underbrace{b(y,t)^2}_{1} \underbrace{\frac{\partial^2 S}{\partial y^2}}_{\sigma^2 S} \right] dt + \underbrace{\frac{\partial S}{\partial y}}_{\sigma S} \underbrace{b(y,t)}_{1} dW \\ &= (\mu + \frac{\sigma^2}{2}) S\, dt + \sigma S\, dW \end{aligned}$$

This corresponds exactly to the process in Equation 3.20. This means that the process S thus constructed satisfies the stochastic differential equation 3.20, i.e. is a solution of this SPDE. Simply making the substitution $t \to t + \delta t$ we obtain

$$\begin{aligned} S(t+\delta t) &= S_0 \exp\left(\mu t + \mu \delta t + \sigma y(t+\delta t)\right) \\ &= S_0 \exp\left(\mu \delta t + \sigma W(t+\delta t)\right) \\ &= S_0 \exp\left(\sigma W(t) + \mu \delta t + \sigma \delta W\right) \end{aligned}$$

where in the second step we absorbed $\exp(\mu t)$ in the (still arbitrary) S_0 and in the third step we adopted the notation δW for a change in a Brownian motion after the passing of a finite time interval δt:

$$\delta W := W(t+\delta t) - W(t) \Longrightarrow \delta W \sim N(0, \delta t) \tag{3.24}$$

The first term in the exponent refers to (already known) values at time t. It can also be absorbed into the (still arbitrary) factor S_0, i.e.

$$S(t+\delta t) = S_0 \exp\left(\mu \delta t + \sigma \delta W\right)$$

Finally, the (still arbitrary) S_0 is chosen so that $S(t+\delta t) \stackrel{\delta t \to 0}{=} S(t)$ holds. This corresponds to the initial condition for the solution of the SPDE. Thus, we obtain the change in S corresponding to Equation 3.20 for a finite, positive time span δt:

$$S(t+\delta t) = S(t) \exp\left(\mu \delta t + \sigma \delta W\right) \quad \text{where} \quad \delta W \sim \mathrm{N}(0, \delta t) \tag{3.25}$$

[12] In order to avoid confusion in the notation, we denote the stochastic variable by y.

Analogously, Equation 3.21 gives the corresponding change in S over a finite, positive time span δt as:

$$S(t+\delta t) = S(t)\exp\left((\widetilde{\mu} - \frac{\sigma^2}{2})\delta t + \sigma\delta W\right) \quad \text{where} \quad \delta W \sim \mathrm{N}(0,\delta t) \tag{3.26}$$

3.4.3 Transition Probabilities, Forward and Backward Equation

Since S in Equation 3.16 is a stochastic process, its *exact* evolution over time cannot be described even by solving the stochastic differential equation. This becomes clear in the above examples where the solutions to Equations 3.25 and 3.26 still contain a random component $\delta W \sim \mathrm{N}(0,\delta t)$. However, probabilities can be given for the event that the stochastic process at time t' takes on values between a and b on the condition that the value of S at an earlier time t is known. A formal notation for this probability is

$$P(a \le S(t') \le b \,|S(t) = S\,) = \int_a^b p(S', t'\,|S, t\,)dS' \tag{3.27a}$$

where the *transition probabilities* denoted by $p(S', t'\,|S, t\,)$ indicate the probability that the stochastic variable at time t' is exactly equal to S', on the condition that it was equal to S at time t. The transition probability[13] contains all information about the associated stochastic process 3.16. This fact prompts a more detailed investigation of transition probabilities.

The Forward Equation

The time evolution of the transition probabilities driven by the stochastic process 3.16 can be described using two differential equations. One of these equations involves the derivative with respect to the future variables S', t' and answers the question: if the stochastic process is equal to S at time t, what distribution does it have at a later time $t' > t$? This equation is called the *forward equation*, also known as the *Fokker-Planck equation*. It is explicitly given by[14]

$$\frac{\partial p}{\partial t'} - \frac{1}{2}\frac{\partial^2}{\partial S'^2}\left[b(S', t')^2 p\right] + \frac{\partial}{\partial S'}\left[a(S', t')p\right] = 0 \tag{3.28}$$

The initial condition that the value of the process at time $t' = t$ is exactly equal to S can be formally expressed with the help of the *Dirac delta function*

$$p(S', t' = t\,|\,S, t\,) = \delta(S' - S)$$

The solution to the forward equation with this initial condition describes the widening of the probability distribution of the process over time starting from the "sharp" delta

[13] This concept of transition probabilities is quite similar to the concept of (Feynman-) propagators used in quantum field theories of elementary particle physics.

[14] To simplify the notation, we neglect here the arguments of p, a habit we will continue to observe often in the remainder of this chapter. If p stands in an equation without an argument, it should be read as $p(S', t'\,|S, t\,)$.

function. The forward equation is useful in situations where S has an established value at a fixed time and we are interested in the probability distribution of S at a later time, in other words, when information available now is to be used to calculate *forward* in time.

Die Backward Equation

In mathematical finance, we are more often interested in the opposite situation: information at a future point in time is known (for example, at the maturity of an option) and we wish to calculate values backwards to an earlier point in time (today, say). To do so, we require a differential equation involving derivatives with respect to the earlier time variables S and t which answers the question: given that the stochastic variable is equal to S' at time t', how is the stochastic process distributed at an earlier point in time $t < t'$? This question can be answered using the *backward* or *Kolmogorov equation* This is given explicitly by

$$\frac{\partial p}{\partial t} + \frac{1}{2}b(S,t)^2\frac{\partial^2 p}{\partial S^2} + a(S,t)\frac{\partial p}{\partial S} = 0 \tag{3.29}$$

The associated initial condition is given by

$$p(S', t' \,|S, t = t' \,) = \delta(S - S')$$

Note that for all process whose parameters a and b are not dependent on S, the only difference between the forward and backward equation is the sign of the term involving the second derivative.

A Derivation of the Forward and Backward Equations

Both the forward equation 3.28 and the backward equation 3.29 are of such a fundamental nature that it is possible to complete their derivation on the basis of first principles. This derivation involves quite extensive calculations (essentially, taking products of Taylor series) which the reader need not necessarily work through (the reader who does not wish to take the time is recommended to continue on to Section 3.4.4). However, experience shows that one tends to feel somewhat ill at ease when such fundamental equations simply appear out of thin air and are accepted without a sound explanation. Therefore, the derivation of the forward and backward equation is presented explicitly here in greater detail than is commonly found in the literature on mathematical finance. This gives the reader the opportunity to obtain a real understanding of these two fundamental equations.

We begin by deriving the *forward* equation. We split the probability that the stochastic process will travel from (S,t) to (S',t') into the probability that by time $t' - \delta t$ it will have arrived at $S' - \delta S$, $S' + \delta S$ or S' and then will proceed to S' in the remaining time δt with the probabilities δp_+, δp_- and δp_0, respectively, thus

$$\begin{aligned} p(S', t' \,|S, t\,) = {} & p(S' - \delta S, t' - \delta t \,|S, t\,)\delta p_+(S' - \delta S, t' - \delta t) \\ & + p(S', t' - \delta t \,|S, t\,)\delta p_0(S', t' - \delta t) \\ & + p(S' + \delta S, t' - \delta t \,|S, t\,)\delta p_-(S' + \delta S, t' - \delta t) \end{aligned} \tag{3.30}$$

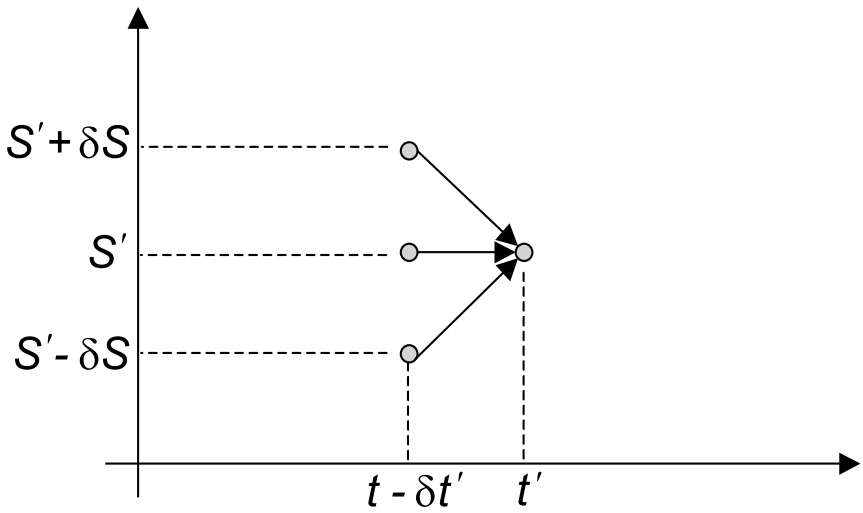


Figure 3.10: The trinomial tree used to derive the forward equation.

where

$$\delta p_{\pm}(S,t) := p(S \pm \delta S, t+\delta t \,|S,t\,)$$
$$\delta p_0(S,t) := p(S, t+\delta t \,|S,t\,) = 1 - \delta p_+(S,t) - \delta p_-(S,t)$$

Intuitively δp_+ is the probability that the process will increase by δS in the short time span δt, δp_- the probability that it will decrease by δS in the time span δt, and δp_0 is the probability that it will remain constant in the time span δt. Other possibilities for this stochastic process within this time span are not admitted. This split of the probabilities is graphically represented by the trinomial tree in Figure 3.10.

The relationship between these δp and the stochastic process 3.16 results from the requirement that the first two moments, i.e. the expectation and the variance of the discrete process defined for the trinomial tree, agree with the first two moments of the continuous Ito process in the limit as $\delta t \to dt \to 0$. The expectation of δS is

$$\mathrm{E}\,[\delta S] = \delta p_+ \delta S + \delta p_0 0 - \delta p_- \delta S = (\delta p_+ - \delta p_-)\delta S$$

By definition, its variance is given by

$$\begin{aligned}\mathrm{var}\,[\delta S] &= \mathrm{E}\left[(\delta S - \mathrm{E}\,[\delta S])^2\right] \\ &= \delta p_+(\delta S - \mathrm{E}\,[\delta S])^2 + \delta p_-(-\delta S - \mathrm{E}\,[\delta S])^2 + \delta p_0(0 - \mathrm{E}\,[\delta S])^2\end{aligned}$$

Substituting $\mathrm{E}[\delta S] = (\delta p_+ - \delta p_-)\delta S$ and $\delta p_0 = 1 - \delta p_+ - \delta p_-$ yields, after several

simple algebraic manipulations

$$\begin{aligned}\mathrm{var}\,[\delta S] &= [\delta p_+(1-\delta p_+ + \delta p_-)^2 + \delta p_-(1+\delta p_+ - \delta p_-)^2 \\ &\quad + (1-\delta p_+ - \delta p_-)(\delta p_+ - \delta p_-)^2]\delta S^2 \\ &= (\delta p_+ + \delta p_-)(1-(\delta p_+ + \delta p_-))\delta S^2 \\ &\approx (\delta p_+ + \delta p_-)\delta S^2\end{aligned}$$

where only terms up to order $\delta p \delta S^2$ were considered in the last step. The requirement that the expectation and the variance be equal to those of the Ito process 3.16 as $\delta t \to dt \to 0$, i.e. up to linear order in δt means that

$$\begin{aligned}(\delta p_+ - \delta p_-)\delta S &= \mathrm{E}\,[\delta S] \overset{\delta S \to dS}{\longrightarrow} \mathrm{E}\,[dS] = a(S,t)dt \\ (\delta p_+ + \delta p_-)\delta S^2 &\approx \mathrm{var}\,[\delta S] \overset{\delta S \to dS}{\longrightarrow} \mathrm{var}\,[dS] = b(S,t)^2 dt\end{aligned}$$

This is obviously achieved by choosing

$$\delta p_\pm(S,t) = \frac{1}{2}\frac{\delta t}{\delta S^2}\left[\pm a(S,t)\delta S + b(S,t)^2\right] \tag{3.31}$$

And, in order to retain the equality $\delta p_0 = 1 - \delta p_+ - \delta p_-$ we have

$$\delta p_0(S,t) = 1 - \frac{\delta t}{\delta S^2} b(S,t)^2$$

Substituting these values in our original equation 3.30 and expanding *all* terms appearing in this expression in a Taylor series about the point (S',t') up to order δt finally yields the forward equation. As a result of Ito's lemma we have $\delta S^2 \sim \delta t$. This means that all terms of order 1, δS, δS^2 and δt are retained and all higher order terms can be neglected. Note that $\delta t/\delta S^2$ in Equation 3.31 is of order 1. Explicitly, the Taylor series for p is given by:

$$\begin{aligned}p(S' \pm \delta S, t' - \delta t\,|S,t) &= p(S',t'\,|S,t) \pm \frac{\partial p}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2 - \frac{\partial p}{\partial t'}\delta t + \cdots \\ p(S',t' - \delta t\,|S,t) &= p(S',t'\,|S,t) - \frac{\partial p}{\partial t'}\delta t + \cdots\end{aligned}$$

But now, all terms involving δp must also be expanded in consideration of Equation 3.31:

$$\begin{aligned}\delta p_+(S' - \delta S, t' - \delta t) &\approx \delta p_+(S',t') - \frac{\partial \delta p_+}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 \delta p_+}{\partial S'^2}\delta S^2 - \frac{\partial \delta p_+}{\partial t'}\delta t + \cdots \\ &\approx \frac{1}{2}\frac{\delta t}{\delta S^2}\left[b^2 + a\delta S\right] - \frac{1}{2}\frac{\delta t}{\delta S^2}\left[\frac{\partial b^2}{\partial S'} + \frac{\partial a}{\partial S'}\delta S\right]\delta S \\ &\quad + \frac{1}{4}\delta t\left[\frac{\partial^2 b^2}{\partial S'^2} + \frac{\partial^2 a}{\partial S'^2}\delta S\right] - \frac{1}{2}\frac{\delta t}{\delta S^2}\left[\frac{\partial b^2}{\partial t'} + \frac{\partial a}{\partial t'}\delta S\right]\delta t + \cdots \\ &\approx \frac{1}{2}\frac{\delta t}{\delta S^2}b^2 + \frac{1}{2}\frac{\delta t}{\delta S^2}\left[a - \frac{\partial b^2}{\partial S'}\right]\delta S \\ &\quad + \frac{1}{2}\left[\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'} - \frac{\partial a}{\partial S'}\right]\delta t\end{aligned}$$

where in the last step, the terms are written in increasing order and those of order $\sim \delta S \delta t$ or smaller are neglected. Analogously we obtain

$$\begin{aligned}\delta p_-(S'+\delta S, t'-\delta t) \approx \frac{1}{2}\frac{\delta t}{\delta S^2}b^2 - \frac{1}{2}\frac{\delta t}{\delta S^2}\left[a - \frac{\partial b^2}{\partial S'}\right]\delta S \\ + \frac{1}{2}\left[\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'} - \frac{\partial a}{\partial S'}\right]\delta t\end{aligned}$$

and finally, for the probability of remaining at S':

$$\delta p_0(S', t'-\delta t) \approx \delta p_0(S', t') - \frac{\partial \delta p_0}{\partial t'}\delta t + \cdots \approx 1 - \frac{\delta t}{\delta S^2}b^2 + \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'}\delta t$$

These expansions substituted in Equation 3.30 give

$$\begin{aligned}p \approx & \left[p - \frac{\partial p}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2 - \frac{\partial p}{\partial t'}\delta t\right] \\ & \times \left(\frac{1}{2}\frac{\delta t}{\delta S^2}b^2 + \frac{1}{2}\frac{\delta t}{\delta S^2}\left[a - \frac{\partial b^2}{\partial S'}\right]\delta S + \frac{1}{2}\left[\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'} - \frac{\partial a}{\partial S'}\right]\delta t\right) \\ & + \left[p - \frac{\partial p}{\partial t'}\delta t\right]\left(1 - \frac{\delta t}{\delta S^2}b^2 + \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'}\delta t\right) \\ & + \left[p + \frac{\partial p}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2 - \frac{\partial p}{\partial t'}\delta t\right] \\ & \times \left(\frac{1}{2}\frac{\delta t}{\delta S^2}b^2 - \frac{1}{2}\frac{\delta t}{\delta S^2}\left[a - \frac{\partial b^2}{\partial S'}\right]\delta S + \frac{1}{2}\left[\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\delta t}{\delta S^2}\frac{\partial b^2}{\partial t'} - \frac{\partial a}{\partial S'}\right]\delta t\right)\end{aligned}$$

where all values appearing in the equation are evaluated at (S', t'). Taking the product and neglecting all terms of order $\sim \delta S \delta t$ or smaller, we obtain, in linear order of δt:

$$\begin{aligned}p \approx p + p\delta t\left(\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\partial a}{\partial S'}\right) - \frac{\partial p}{\partial S'}\delta S\frac{\delta t}{\delta S^2}\left[a - \frac{\partial b^2}{\partial S'}\right]\delta S \\ + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2\frac{\delta t}{\delta S^2}b^2 - \frac{\partial p}{\partial t'}\delta t\end{aligned}$$

This immediately yields the forward equation:

$$\begin{aligned}0 &= \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}b^2 + p(\frac{1}{2}\frac{\partial^2 b^2}{\partial S'^2} - \frac{\partial a}{\partial S'}) - \frac{\partial p}{\partial S'}(a - \frac{\partial b^2}{\partial S'}) - \frac{\partial p}{\partial t'} \\ 0 &= \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}b^2 + \frac{\partial p}{\partial S'}\frac{\partial b^2}{\partial S'} + \frac{1}{2}p\frac{\partial^2 b^2}{\partial S'^2} - p\frac{\partial a}{\partial S'} - \frac{\partial p}{\partial S'}a - \frac{\partial p}{\partial t'} \\ 0 &= \frac{1}{2}\frac{\partial^2}{\partial S'^2}(b^2 p) - \frac{\partial}{\partial S'}(a\,p) - \frac{\partial p}{\partial t'}\end{aligned}$$

The derivation of the *backward* equation is analogous but requires considerably less effort: since the variable in a trinomial tree has only three possibilities to travel from a point S' at time t' over the next time step δt, the probability that the variable willbe

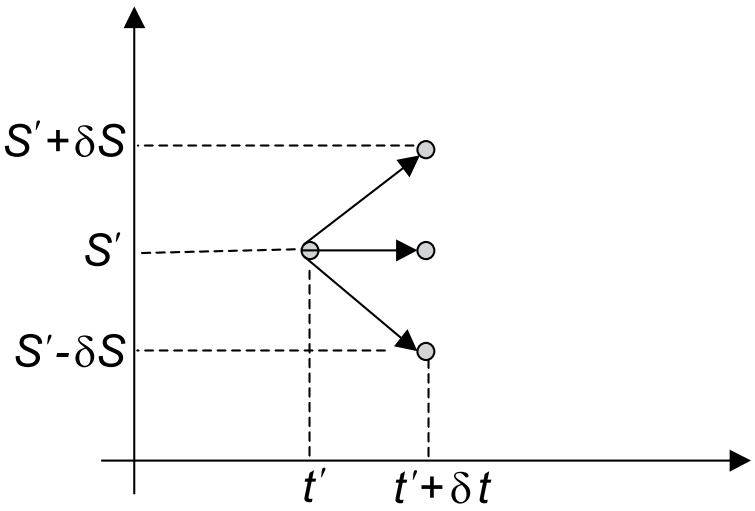


Figure 3.11: The trinomial tree used to derive the backward equation.

equal to S' at time t' is equal to the sum of the probabilities that it will travel to either $S' + \delta S$, $S' - \delta S$ or S' under the condition that it started at S at time t, see Figure 3.11. Thus we can state

$$\begin{aligned} p(S', t' \,|S, t) = {} & p(S' + \delta S, t' + \delta t \,|S, t)\delta p_+(S', t') \\ & + p(S', t' + \delta t \,|S, t)\delta p_0(S', t') \\ & + p(S' - \delta S, t' + \delta t \,|S, t)\delta p_-(S', t') \end{aligned}$$

This equation means in words: the probability of being S' at time t' equals

- the probability of being $S' + \delta S$ at time $t' + \delta t$ multiplied by the probability of having made an "up move" in the time interval from t' to $t' + \delta t$,
- plus the probability of being $S' - \delta S$ at time $t' + \delta t$ multiplied by the probability of having made a "down move" in the time interval from t' to $t' + \delta t$
- plus the probability of being S' at time $t' + \delta t$ multiplied by the probability of having made a "null move" in the time interval from t' to $t' + \delta t$.

We can now see immediately that the backward equation is more easily derived than its counterpart since the probabilities δp are needed at the point (S', t'), and consequently, a Taylor series expansion about this point is unnecessary. We only need

the Taylor series expansion of p given by

$$\begin{aligned} p = &\left[p + \frac{\partial p}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2 + \frac{\partial p}{\partial t'}\delta t\right]\frac{1}{2}\frac{\delta t}{\delta S^2}\left[b^2 + a\delta S\right] \\ &+ \left[p + \frac{\partial p}{\partial t'}\delta t\right]\left[1 - \frac{\delta t}{\delta S^2}b^2\right] \\ &+ \left[p - \frac{\partial p}{\partial S'}\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2 + \frac{\partial p}{\partial t'}\delta t\right]\frac{1}{2}\frac{\delta t}{\delta S^2}\left[b^2 - a\delta S\right]) \end{aligned}$$

All terms appearing in the above expression are evaluated at the point (S', t'). Taking the products and neglecting all terms of order $\sim \delta S\delta t$ or smaller gives, in linear order of δt,

$$p = p + \frac{\partial p}{\partial S'}\delta S\frac{\delta t}{\delta S^2}a\delta S + \frac{1}{2}\frac{\partial^2 p}{\partial S'^2}\delta S^2\frac{\delta t}{\delta S^2}b^2 + \frac{\partial p}{\partial t'}\delta t$$

Which, after rearranging the terms, immediately yields the backward equation 3.29.

3.4.4 Forward and Backward Equation in the Black-Scholes World

We consider the simple example of the process given by Equation 3.21, namely $dS(t) = S(t)\widetilde{\mu}\,dt + S(t)\sigma dW$, i.e. we simply have

$$\begin{aligned} a(S', t') &= S'\widetilde{\mu} \\ b(S', t') &= S'\sigma \end{aligned}$$

This process, or its equivalent 3.13, is, as has been shown above, the simplest process of relevance in mathematical finance. Reasonable and often even analytical expressions for the prices of financial instruments can be obtained by solving these equations. For example, we will later show that such processes form the basis for the famous Black-Scholes option pricing formula. We thus speak of the *Black-Scholes world* when referring to the description of market parameters using such processes. With these assumptions, the backward equation becomes

$$\frac{\partial p}{\partial t} + \frac{1}{2}\sigma^2 S'^2\frac{\partial^2 p}{\partial S^2} + \widetilde{\mu}S'\frac{\partial p}{\partial S} = 0$$

and likewise the forward equation is simply[15]

$$\frac{\partial p}{\partial t'} - \frac{\sigma^2}{2}S'^2\frac{\partial^2 p}{\partial S'^2} + \left[\widetilde{\mu} - \sigma^2\right]S'\frac{\partial p}{\partial S'} + \left[\widetilde{\mu} - \sigma^2\right]p = 0$$

[15] The derivation of this equation is a simple application of the product rule:

$$\begin{aligned} \frac{\partial p}{\partial t'} - \frac{1}{2}\frac{\partial^2}{\partial S'^2}\left[S'^2\sigma^2 p\right] + \frac{\partial}{\partial S'}\left[S'\widetilde{\mu}\,p\right] &= 0 \\ \frac{\partial p}{\partial t'} - \frac{\sigma^2}{2}\frac{\partial}{\partial S'}\left[2S'p + S'^2\frac{\partial p}{\partial S'}\right] + \widetilde{\mu}\left[p + S'\frac{\partial p}{\partial S'}\right] &= 0 \\ \frac{\partial p}{\partial t'} - \frac{\sigma^2}{2}\left[2p + 2S'\frac{\partial p}{\partial S'} + S'^2\frac{\partial^2 p}{\partial S'^2}\right] + \widetilde{\mu}\left[p + S'\frac{\partial p}{\partial S'}\right] &= 0 \end{aligned}$$

The solution to this differential equation with the above initial condition $p(S', t \mid S, t) = \delta(S' - S)$ provides an explicit example for a transition probability:

$$p(S', t' \mid S, t) = \frac{1}{S'\sqrt{2\pi\sigma^2(t'-t)}} \exp\left\{ \frac{-\left[\ln(S'/S) - (\widetilde{\mu} - \sigma^2/2)(t'-t)\right]^2}{2\sigma^2(t'-t)} \right\} \quad (3.32)$$

The corresponding equation and its solution for the equivalent process 3.20 can of course be found by simply making the substitution $\widetilde{\mu} = \mu + \sigma^2/2$ in accordance with 3.23.

4

Financial Instruments: A System of Derivatives and Underlyings

As mentioned in the introduction, *trading* can be defined as an agreement between two parties in which one of the two consciously accepts a financial risk in return for the receipt of a specified payment or at least the expectation of such a payment at same future time from the counterparty. *Financial instruments*, also called *financial products*, are instruments which make such a *risk mitigation* or *risk transfer* possible. The purpose of this section is to present a classification of such instruments in a system of underlyings and derivatives, specifically for interest rate instruments. *Interest rate risk* is by far the most complex market risk. Correspondingly, *interest rate instruments*, ie. instruments having interest rates as their underlying risk factors, are among the most complex financial instruments traded on the market. Instruments on other risk factors such as stocks or foreign exchange rates can be classified analogously and will be discussed in detail in later sections of the book.

The main characteristics of interest rate instruments is that they can be represented as a right to one or several payments (*cash flows*) which may be either fixed or variable and will occur at some specified future dates. The amount of these cash flows is determined from the length of the respective interest periods and the associated interest rates quoted according to the convention particular to the market under consideration. The following parameters may be used to define the cash flow structure of an instrument, in particular, to establish the times on which payments are to be made: valuation date, rollover date, maturity date, payment date, reset or fixing date, frequency of the coupon payments per year (e.g., 1, 3, 6 and 12 months), long or over-long first period, partial final period, number of interest periods, conventions (day count conventions, business day conventions, etc.)

We differentiate between the start of an interest period, the maturity of the interest period and the payment date. The start of the interest period is the date upon which interest begins to accumulate. The *maturity* is the date on which interest stops

accumulating, in other words, it is the end of the interest period. The *payment date* is the date on which the payment is actually made.

The *rollover date* is the date on which the current interest period ends and the next begins. Knowledge of the rollover date is quite important since it specifies the first date as seen from today after which regular cash flows occur with a specified frequency. Payments are made subsequent to this date until maturity. This is the reason why many trading systems do not fix the maturity date when constructing cash flow systems; if fixed interest periods are assumed, the maturity date is simply the payment date at the end of the last interest period (adjusted to the correct business day in accordance with the appropriate day count convention).

4.1 Spot Transactions

On the *spot market*, transactions take place in which financial instruments are immediately (more precisely, as of the current *value date*) exchanged and paid for. Basically, the spot market for interest rate instruments consists of the short-term *money market* and the long-term *capital market*. Therefore, interest rate instruments are traditionally categorized into two groups, *money market securities* and *capital market securities.*

4.1.1 Money Market Securities

In contrast to the capital market, where securities with terms of up to 30 years or more are traded, the money market encompasses all short-term instruments, with *terms* (i.e. times to maturity) of up to one year. Occasionally, instruments with terms of up to two years are referred to as money market instruments. The money market itself can be divided into two categories, depending on whether or not the money market instrument pays regular interest. Figure 4.1 shows examples of commonly traded money market instruments.

Discount Papers

Papers which are traded at a discount of their nominal value and do not regularly pay interest are referred to as *discount papers*. The interest rate at issue is determined by the difference between the issue price and the nominal value. The buyer pays the current price equal to the nominal less the discount and receives the entire nominal at maturity in return. Typical examples of discount papers are *treasury bills* and *commercial papers* in the US or the UK or *BTFs* (*bons du Tresor a taux fixe*) in France.

Commercial papers, or *CPs* for short, are short-term, unsecured, tradable obligations issued by banks and corporations. In Germany, commercial papers can be issued with terms ranging from seven days up to two years. A commercial paper program represents the general framework for an agreement between the issuer and the banks appointed to place the paper. It gives the issuer the right but not the obligation to issue short-term papers into the market at any time. Such a program has the character of a continual emission since the CP can be issued at various times in various amounts

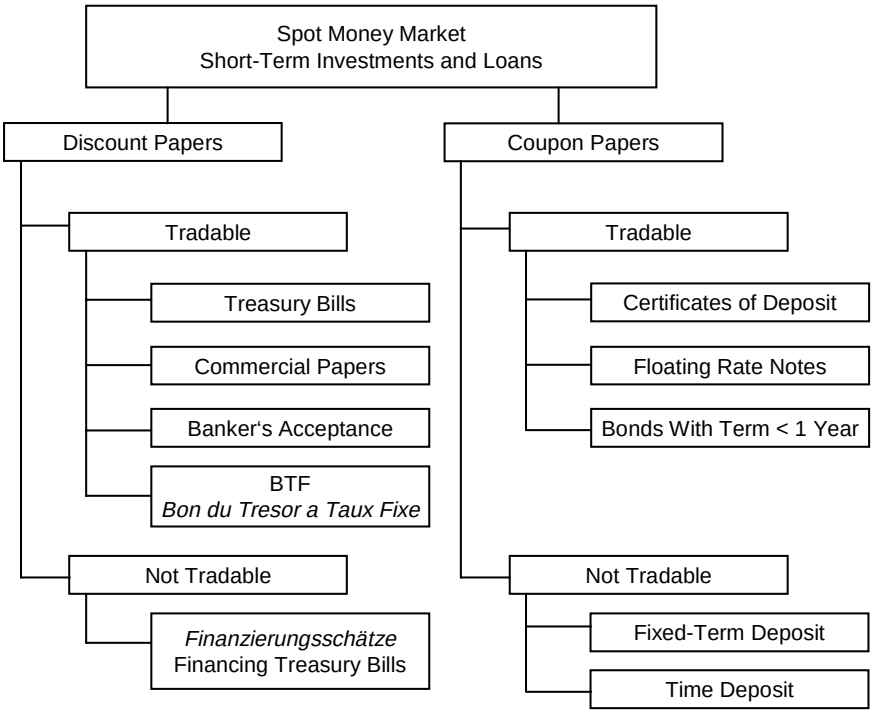


Figure 4.1: Examples of common money market instruments.

over a longer period of time. The terms of commercial papers have an upper limit of two years and are closely related to the so-called *medium-term note* program under which partial obligations with a term of at least two years are issued. The commercial paper program fills the gap to the medium-term note program.

Interest Bearing Securities

Papers having one or more regular interest payments are referred to as *interest bearing securities* or *coupon bearing securities*.They are issued at face value. *Certificates of deposit*, *floating rate notes*, *reverse floaters* and short-term capital market papers are included in this category.

Certificates of deposit (*CD*), are tradable money market papers issued by banks having a term between 30 days and 4 years. Essentially, CDs are securitized time deposits in banks.

Floating rate notes (also called *floaters* or *FRNs*) are *bonds* whose interest payments vary depending on a current market rate. The interest rates are regularly adjusted to a reference rate (for example, the *LIBOR* or *EURIBOR*), in general, to a money market rate. Hence, floaters, despite often having longer terms are also assigned to the category of money market papers. As opposed to other floaters (reverse floaters, for example),

ID (WKN)	113494	263119	315903
Issuer	German Government	Hamburgische Landesbank	Rheinboden Hypothekenbank
Face Value	EUR 1,000	EUR 5,000	DM 10,000
Reference Rate	3-Month EURIBOR	6-Month EURIBOR	6-Month LIBOR
Spread	-0.25% p.a.	0.041% p.a.	0.00% p.a.
Interest Payment	20th of Mar, Jun, Sep, Dec	11th of Feb, Aug	2nd of Jan, Jul
Repayment on	Sep 20, 2004	Aug 11, 2009	Jan 02, 2008

Table 4.1: Examples of floaters in Germany

these normal floaters are referred to as plain vanilla. *Plain vanilla* is an American term for simple papers without any particular attributes such as options or convertibility rights. These simple floaters are to be considered as floating rate notes in their most basic form. Table 4.1 provides some examples of floaters in Germany.

The most important difference between a floater and a *fixed rate* instrument is the nominal interest rate. While for a fixed rate paper, the nominal interest rate is fixed over the entire lifetime of the paper, the nominal rate of the floater fluctuates. The nominal interest rate changes according to the current level of the reference market rate; a decrease in short-term rates results in a corresponding decline in the interest yield of a floater. Investors expecting a decline in money market rates will therefore usually close out their positions in plain vanilla floaters.

A more complex variant of plain vanilla floaters are *reverse floater*, sometimes referred to as *bull floating rate notes*. Like plain vanilla floaters, reverse floaters are papers paying interest at a variable rate, depending on the level of the current money market rate. Also like normal floaters, the interest is paid in regular intervals, e.g. every six months. However, while the interest yields of a plain vanilla floater increases with rising money market rates, the opposite is true of the reverse floater (hence the name). For a reverse floater, the current money market rate is *subtracted* from a fixed base rate.

Long-term papers on the capital market such as federal bonds and obligations, mortgage bonds, jumbo bonds and municipal bonds are indeed issued with terms than money market papers. However, the residual *time to maturity* naturally decreases constantly with the passing of time. Therefore, these papers, initially designated as capital market papers, are comparable with money market papers in the last year of their lifetime.

Time and Notice Deposits

The name *time deposit* signifies a deposit at a bank which is invested for a certain time (*fixed deposits*) or with an agreed upon term of notice (*deposits at notice*). *Fixed deposits* are deposits at a bank or savings institution with individual, fixed and unalterable maturities, amounts and interest rates, all agreed upon in advance. The

conditions depend on current interest rates as well as the term of the deposit. The amount deposited also plays a deciding role; the more invested, the higher the interest paid. The minimal sum which may be invested is usually DM 10,000 or EUR 5,000.

Investors gain somewhat more flexibility in comparison to fixed deposits by placing their savings in *deposits at notice*. For these deposits, a period is agreed upon which the investor must give notice in advance if he or she wants to withdraw the invested funds. Like fixed deposits, the deposits at notice offered by banks usually start at a minimum principal of DM 10,000 or EUR 5,000. Withdrawal notice can be one day, 48 hours, seven days or even three months. In comparison to fixed deposits, deposits at notice play a subordinate role.

Trading Conventions for Money Market Instruments

Similar to papers sold on the capital markets, the yields of money market papers are quoted publicly. *Yields* are always per year, in other words, they are *annualized.* However, different conventions for calculating yields are observed on the different international markets, as was illustrated in Section 3.1. The annualized yields are influenced by the day count convention and compounding method used. The day count conventions vary in the calculation of the number of days in the year as well as the days of the month. The usual day count convention in the money markets is the *Act/360* convention. Examples of instruments whose interest is calculated in accordance with this day count convention are commercial papers, treasury bills and floating rate notes in the US. In Germany, quotes have been published in accordance to this convention since July 1, 1990 for inter-bank trading. *Act/360* is employed elsewhere in Europe as well. Floaters are also calculated employing this method. The method *Act/360* is also referred to as the *international* or *French* method. It is frequently called the *euro interest convention* as well. Alternatively, *Act/365* is used in several markets. Sterling commercial papers, for instance, are quoted in accordance with this method; it is sometimes referred to as the *English* method.

In addition, the *30/360* convention is employed in particular in trading bonds and obligations; with the introduction of the euro, it is to be phased out and replaced with the *Act/Act* convention. The *30/360* convention assumes that each year has 360 days, each month 30 days. The *30/360* convention is also referred to as the *German* method. The *Act/Act* method calculates with the actual number of days of the month and year. In Germany, the *30/360* convention is still applied, in particular for mortgage bonds issued before 1999. Some existing issues have already changed over to the *Act/Act* convention (for example, all government bonds and all jumbo bonds) New issues in the euro zone are computed in accordance with the *Act/Act* convention.

If yields are determined for money market instruments which pay interest several times within a single year (or in its lifetime if this happens to be less than one year in duration), a conversion into an annual yield (*annualization*) can be accomplished with or without *compounding interest effects.* Yields which do not take compounded effects into consideration (linear and simple compounding) are referred to as *nominal yields* or *nominal rates.* These methods of computing yields predominate in most money markets. We already provided a detailed description of the different compounding conventions in Section 3.1.

Instrument	Yield Calculation	Day Count	Typical Term on Issue
Treasury Bill	Discount Rate	Act/360	13,26,52 Weeks
EURIBOR	MMY	Act/360	1,2,3,...,12 Months
Commercial Paper USD (CP)	Discount Rate	Act/360	7 Days to 2 Years
Commercial Paper GBP (CP)	Discount Rate	Act/365	7 Days to 2 Years
Certificate of Deposit USD (CD)	MMY	Act/360	7 Days to 1 Year
Federal Financing Treasuries	Discount Rate	Act/Act	1 or 2 Years

Table 4.2: Some examples of international money market instruments.

On the international money markets, money market papers are traded in accordance with linear (*discount rate*) or simple compounding (money market yields).

Discount papers are generally traded on the basis of the discount rate (*DR).* The discount rate is not a yield since, in contrast to a yield, it is not calculated by reference to invested capital, but on the nominal principal of the instrument. Examples of discount papers traded in foreign markets are treasury bills in the United States

The *money market yield* (*MMY*) or *CD-equivalent yield* is computed for money market instruments having interest payments. An example is the *certificate of Deposit* (*CD*). Table 4.2 provides an overview of the most important details of several money market instruments on the international money markets.

4.1.2 Capital Market Securities

Capital market papers are long-term interest rate instruments with a *term* (time to maturity) longer than one year. Papers with terms longer than 30 years are available, for instance in Germany. Typically, however, such papers have a term of ten years. The prices and yields of these long-term instruments are computed with compounded interest. Figure 4.2 illustrates some examples of common capital market papers.

Long-term interest rate instruments can also be divided into discount instruments and coupon bearing securities. The only difference between long-term to short-term interest rate instruments is that multiple interest payments are involved, instead of only one for most money market instruments.

Zero coupon bonds (or *zero bonds* for short) are tradable interest instruments which can be classified as discount securities. Like financing treasury notes, zero bonds are interest instruments without regular interest payments. The issue price is significantly lower than the redemption price. The difference between both determines the yield to maturity. We can distinguish between *zeros issued with discount*, which are issued well below par, e.g. at 60%, and *capital growth bonds*, which are repaid well above par, e.g. 230%. Zero coupon bonds with discount are issued most frequently.

Non-tradable long-term interest instruments are for instance the two-year financing treasury bonds (*Finanzierungsschätze*) and the federal savings bonds type B (*Bundesschatzbriefe Typ B*) in Germany . The latter have annual interest rates, which rise during the time to maturity according to a fixed plan (*step up coupon*). The interest is automatically reinvested by the issuer. The investor can only gain access to the

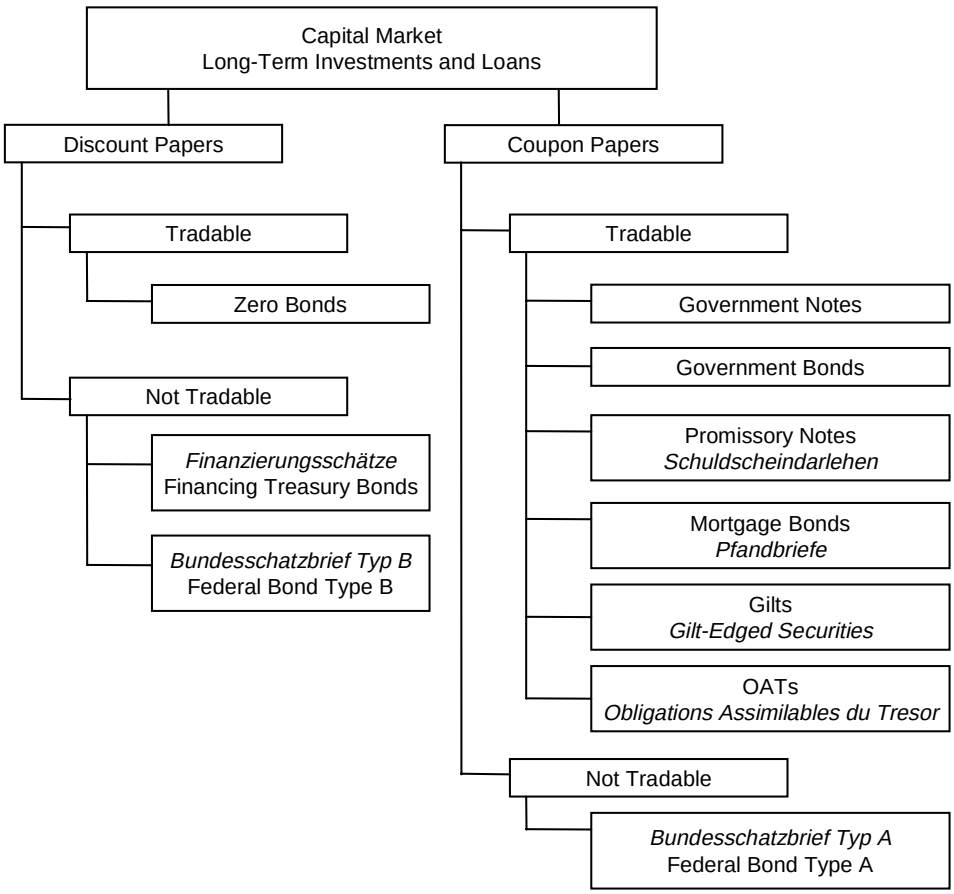


Figure 4.2: Examples of common capital market securities.

interest and reinvestment yields from interest payments at maturity of the paper. For this reason, the payment structure of Bundesschatzbrief Typ B is comparable to the payment structures of zero bonds.

The dominating long-term interest instruments are clearly the tradable instruments. The interest instruments introduced below are exclusively *straight bonds* (one exception is *Bundesschatzbrief Typ A*). Straight bonds are instruments with the following features:

- Fixed *coupon* interest rate, which is constant during time to maturity
- Redemption at par (i.e. 100%)
- Fixed life span
- Payment of the full price at issue or purchase

- No embedded option rights

Treasury notes (in Germany called *Bundesschatzanweisungen*) are securities issued by the government with maturity periods up to of two years. Treasury notes are fixed-rate coupon-bearing securities with 100% repayment at maturity. *Bundesobligationen* (or *BOBL* for short) have times to maturity of five years and are also issued by the German government. The interest is paid annually and redemption is also made at face value. Both instruments are medium-term and therefore comparable to *US Treasury notes.*

Mortgage bonds and *jumbo bonds* are interest instruments issued by mortgage banks and public sector banks used for refinancing mortgage loans. Public mortgage bonds and *municipal bonds* are issued for financing tasks of public institutions (cities, communities). Mortgage bonds and municipal bonds pay annual interest and are repayed in full at maturity.

Government bonds are securities issued by the government with terms up to 30 years, annual interest payments and redemption at par. Examples of these are the *Bundesanleihen* in Germany, *US Treasury bonds* in the United States or the French *OATs* (*obligation assimilable du Trésor*).

Schuldscheindarlehen in Germany are *promissory notes* or *loans* issued by banks, states and the federal government which are not traded on an exchange. A promissory loan is a transferable contract between the lender and borrower. A promissory loan, from a legal point of view, does not fall into the category of financial instruments. The lifetime can be as long as 15 years.

The only existing long-term non-tradable interest instrument in Germany is the *Bundesschatzbrief Typ A*, which is issued by the German government. The major difference in comparison with type B is that type A has annual interest payments and a life of only six years. Type A also has a step-up coupon (annually rising interest rate).

On many international government bond markets the interest payment is made semi-annually. This convention can be found especially in Anglo-American bond markets. Table 4.3 gives an overview.

4.1.3 Swaps

A *plain vanilla swap* is an agreement in which two parties agree to exchange a series of fixed interest payments for a series of floating interest payments. While the *fixed leg* of the swap has the character of a fixed rate bond, the *floating leg* refers to a variable, short-term interest rate (the 6-month EURIBOR for instance) which is "fixed" at some agreed upon future dates.

Thus, swaps combine two portfolios (the legs) and the interest is usually earned on the same nominal principal for both legs. However, different compounding and payment conventions might be observed. A standard or plain vanilla swap consists of series of coupon payments, called the *coupon strip* in exchange for a series of EURIBOR payments, the *EURIBOR strip.*

For a plain vanilla swap in euros, both strips have the same maturity but may roll over at different dates. The fixed leg pays interest annually in accordance with the

National Bond Market	Coupon Frequency	Yield Computation
Australia	Semi-annual	Semi-Annual Yield
Belgium	Annual	Annual Yield
Federal Republic of Germany	Annual	Annual Yield
France	Annual	Annual Yield
United Kingdom	Semi-Annual	Semi-Annual Yield
Japan	Semi-Annual	Simple Yield-to-Maturity
Italy	Semi-Annual	Semi-Annual Yield
Canada	Semi-Annual	Semi-Annual Yield
Netherlands	Annual	Annual Yield
Austria	Annual	Annual Yield
Sweden	Annual	Annual Yield
Switzerland	Annual	Annual Yield
Spain	Annual	Annual Yield
United States	Semi-annual	Semi-Annual Yield

Table 4.3: Interest payment modes in international bond markets.

30/360 compounding convention. The floating leg, in contrast, pays semi-annually and the *Act/360* compounding convention is observed. The *TARGET holiday calendar* is used to determine bank holidays.

4.2 Forward Transactions

Derivative securities, or short *derivatives*, can be divided into *conditional* and *unconditional* forward transactions. A conditional forward transaction grants one party of the contract certain rights whereas the other party assumes certain obligations. In contrast, an unconditional forward transaction is an agreement that is binding on both parties. Warrants, options, and instruments similar to them, e.g. caps and floors, can be classified as conditional derivatives. Futures and forwards, on the other hand, are assigned to the unconditional derivatives.

Forward Rate Agreements and Forwards

Forward rate agreements (abbreviated *FRA*s) are the oldest and thus probably the most simple interest rate instruments. A forward rate agreement is an agreement between to parties to lend or borrow short-term money at a fixed rate of interest at some time (usually a few months) in the future. The agreed upon fixed rate is referred to as the *FRA rate* or *forward rate*. If the reference interest rate, e.g. EURIBOR, exceeds the agreed FRA rate at maturity, the buyer realizes a profit and receives a cash settlement payment from the seller. However, if the reference rate at maturity is lower, the buyer suffers a loss and compensates the seller (in form of a cash settlement). It follows that the buyer pays the FRA rate and receives in return the current interest rate from the seller.

Aa FRA is usually purchased in combination with a loan. The *buyer* of a FRA protects him or herself against *rising* interest rates. If the reference interest rate (for example, the EURIBOR) at maturity lies above the FRA rate, the buyer makes a profit (which can be used to compensate for the higher interest rate on the loan). If the reference rate lies under the FRA rate, the buyer takes a loss (but on the other hand has to pay lower interest on the loan).

Conversely, the *seller* of a FRA protects him or herself against *falling* interest rates. He or she could for instance sell a FRA to lock in (at least for the FRA period) a certain interest yield on a floating rate investment. If the reference rate (e.g. EURIBOR) at the FRA's maturity is lower than the FRA rate, the seller receives a cash settlement which can be used to increase the (then low) yield from the investment. If the EURIBOR at FRA maturity is higher than the FRA rate, the seller has to pay a cash settlement (but then will have a high yield from his investment)

If the reference rate at the fixing time is equal to the FRA rate, neither a profit nor a loss is realized.

Two additional factors should be taken into account with respect to FRAs. Both parties of the contract face a counterparty credit risk. This is because depending on the interest rate level at maturity of the FRA, either party receives or is obliged to make a cash settlement payment. The second aspect is that no liquid funds are exchanged at the transaction or during the life of the contract. Only at maturity of the FRA will the contract be settled if the current interest rate does not equal the FRA rate.

If the underlying in the contract is not an interest *rate* (e.g. LIBOR, EURIBOR) but a fixed rate security (for example, a government obligation, government bond, mortgage bond, promissory note, etc.) then such a forward transaction is simply referred to as a *forward*.

Financial Futures

Although *commodity futures* have been traded on organized exchanges since 1860, *financial futures* are a rather recent addition to the market. In the US, active trading in *commodity* futures began at the beginning of the 19th century. It was no coincidence that the *Chicago Board of Trade* (*CBOT*) was founded as early as 1848; this was where wheat contracts were first traded. In the course of the following years, additional commodity exchanges were founded, for example in New York and London.

Trade in *financial* forward transactions is relatively recent; trading in these instruments commenced in Chicago in 1972 on the *International Monetary Market* (*IMM*). The first transactions were concluded for foreign currencies. In Germany, futures have been traded since the establishment of the *Deutsche Terminbörse* (*DTB,* now called the *EUREX*) in 1990.

The basic idea behind *futures* is identical to that of forward rate agreements and forwards. Futures are a further development in forward transactions, which go back as far as the 17th century. Hence, futures have the same payoff profile as forwards. The essential difference between forwards and futures is that contract elements are not individually negotiated. A future is a *standardized* forward transaction. The underlying security, the volume, the time of settlement, and other payment and delivery conditions are standardized and are set by the exchange (for example, *EUREX*, *LIFFE*).

Futures are usually not exercised but closed out prior to maturity by entering into a trade opposite to the original one.

Similar to forward transactions, both parties to a futures contract take on a counterparty risk. The counterparty risk in the case of futures, however, is lower by orders of magnitude; it is in fact negligible. Futures markets have developed two mechanisms for the elimination of counterparty risk. The first involves the settlement of a potential payment not at maturity but daily. If, for example, the forward price of the futures underlying increases, the holder of a futures contract makes a profit. For a forward transaction, the profit in the form of a settlement payment cannot be realized until the maturity of the forward transaction, i.e. at the end of the forward's lifetime. For futures, in contrast, the profit is credited *directly* to a so-called *margin account* on a daily basis. Similarly, in the case of a loss, the corresponding amount is debited from the account. In the language of the futures market, these daily profit and loss payments are referred to as *daily settlement*. The daily sum credited or debited to the margin account is called the *variation margin*. Since the time to repayment is reduced to one day by the daily settlement, the counterparty risk is accordingly smaller.

In addition to the variation margin, a second mechanism reducing the counterparty risk is the *risk-based margin* which every market participant has to set aside. These are securities which must be set aside by the investors to cover potential settlement risk. Clear rules establish the amount of margin required by each exchange. These rules are based on price fluctuations associated with futures. For example, the EUREX requests payment of a so-called *additional margin (*originally called the *initial margin*) of 1,400 euros (as of October 26, 2000) for every Bund future contract (contract size of 100,000 euros) entered into, to guarantee the ability to make any settlement payments later on. If the value of the contract increases in the course of a business day, the profit is credited to the margin account. On the other hand, if the value of the contract falls, the loss is charged to the account. Should the margin account balance fall below 1,400 euros, the market participant must deposit additional funds to ensure the minimum level of 1,400 euros per contract. This subsequent payment is referred to as a *margin call.* If the market participant fails to meet this obligation, his or her position is perforce closed out. These two precautionary measures serve to reduce the counterparty risk to an (almost) negligible amount.

A further difference exists between forward transactions and futures. While two parties close a contract directly with one another in the case of a forward, the exchange plays the role of a *clearing house* for futures and *guarantees* the settlement of the contract. The market participant thus need not estimate the counterparty risk of an arbitrary trading partner but solely that represented by the exchange itself. The counterparty risk is thus transferred to the exchange.

A further advantage of futures lies in their liquidity (market depth) in the markets. The standardization results in a concentration on trading in only a few instruments. This high liquidity leads to a lower demand-dependent price fluctuations and to lower transaction costs.

4.3 Options

With FRAs, forwards, futures and swaps, a party enters into a contractual *obligation*. In contrast to this, the buyer of an *option* obtains a *right*. For an FRA, for example, a payment is to be made on a fixed date agreed upon in advance in an amount which is determined solely by the market, automatically, so to speak. Options, on the other hand, furnish the buyer with a *claim* to payment or delivery which is conditional on market events and terms agreed upon in advance. The buyer of the option then has the choice of exercising this right or allowing it to expire without exercise. The seller of the option always has an obligation.

Options are divided into two basic types: options to *buy* and options to *sell*, referred to as *calls* and *puts*, respectively.

Options which are traded on an exchange are called *exchange traded options*. On the EUREX, for example, options on stocks, the *DAX* index, the Bobl future, and the Bund future are traded. The EUREX is the largest options exchange in Europe. In addition to options traded on the various exchanges, individually designed options, called *OTC-options* (*OTC* is an abbreviation for *over the counter*), are traded directly between financial institutions and corporations rather than on an exchange. All the features of an OTC option can be freely specified between the counterparties, independent of the standardization required for options traded on an exchange. Options or *option rights* can also be embedded in other instruments. These options are referred to as *embedded options* or *embeddos*. The most common types are warrant bonds, which are composed of straight bonds and attached *warrants*. Shortly after issue the warrants are usually detached from the bonds and traded independently as securities, viz. as warrants. Additionally, there exist securities with attached rights that resemble warrant bonds. In these instruments, the attached rights also exhibit option-like properties. In this case, however, these cannot be detached from the underlying security. Examples are *callable* and *puttable* bonds.

Generally speaking, an option is an agreement between two contract parties in which the seller guarantees the buyer the right to buy (call) or sell (put) an *underlying* at a previously agreed upon price either within a certain time period (*American option*) or on a certain date (*European option*). The seller of the option, guarantees this right in exchange for payment of a certain cash amount called the *option premium*. The price for which the underlying is to be bought or sold is called the *strike* or *strike price*. The date up to which (American options) or on which (European options) the option may be exercised by the buyer is call the *maturity date*.

An option is said to be *at the money* when its strike is equal to the current underlying price. If the current underlying price is such that the option holder would make money by exercising the option *now*, then the option is said to be *in the money*. If the current underlying price is such that the holder would *not* make money by exercising the option *now*, then the option is said to be *out of the money*.

Sometimes there is confusion among investors about the expressions put and call. For stocks, a long position in calls wagers on an increase in the stock price, whereas the holder of a long put position hopes for a decline in the stock price. For interest rate instruments, the situation is somewhat more complicated. For example: A bank has issued a floater and wishes to hedge itself against rising short-term rates (6-month

EURIBOR). One possibility would be to take a long position in a call on the 6-month EURIBOR. The higher the 6-month EURIBOR increases, the more valuable the call. The underlying is the interest rate. The same hedging result is obtained when taking a long position in puts on an interest rate *instrument* (for example, on the 3-month euro future). The future *loses* value if short-term rates *rise*. The further the price of the future falls the higher the value of the put on the future). The underlying of the put is the future and not the interest rate. The desired result is attained by both strategies: the bank accomplishes the hedge against rising short-term rates since it can compensate for the higher interest rates it has to pay for the floater through the profits resulting from the long call on the 6-month EURIBOR or the long put on the 3-month Euro *future.*

There has been a marked increase in demand for caps and floors in recent years since an increasing number of investors wish to protect themselves against fluctuations in the money market rates or the term structure curve, or to profit from these fluctuations. While an investor would wish to invest in caps if he or she expected increasing money market rates, an increasingly steep interest rate term structure or an upward parallel shift of the term structure, floors are interesting for the investor wishing to profit from falling money market rates, a flatter term structure or a downward parallel shift of the term structure.

A *cap* is an agreement between the seller of the cap and the buyer according to which the seller pays the buyer the difference in the interest earned on a nominal with respect to a *reference* rate (e.g. a floating market rate such as EURIBOR) and that earned on an agreed upon fixed *strike* rate, should this fixed rate be *exceeded* by the reference rate. Different modes of payment are used in practice: in some cases, the discounted payoff is made at the maturity of the option ("early"), otherwise, payment is made upon maturity of the respective EURIBOR period("standard"). Viewed analytically, a cap can be interpreted as a series (portfolio) of European call options on a reference rate with various maturities. The buyer of a cap (long cap) profits from an increase in interest rates. Caps are employed primarily to hedge against interest rate risk arising in consequence of variable *financing* (floaters, for instance). They can also be used when speculating on rising interest rates. Moreover, variations in the term structure effect a corresponding change in the value of the option.

Floors are the counterpart of a cap. While caps limit a floating rate from above, the floor limits them from below (lower interest rate limit). A floor is an agreement between the seller of the floor (short floor) and the buyer (long floor) in which the seller pays the buyer the difference in interest on a nominal calculated with respect to a floating market rate (for example, the 6-month EURIBOR) and a fixed *strike* rate established in the floor, should the market rate fall *below* this fixed rate. Floors are employed in hedging against risks involved in floating rate *investments* (such as floating rate notes). From the analytical point of view, a floor can be interpreted as a series (a portfolio) of European put options on the reference market rate with various maturities.

Collars are a combination of caps and floors. The purchase of a collar (long collar) corresponds to simultaneously buying a cap (long cap) and selling a floor (short floor) where the strike of the floor is lower than that of the cap. With the purchase of a collar, the buyer obtains the right to receive a settlement payment from the seller should the

reference rate exceed the cap rate, must however make payment to the seller should the reference rate sink below the floor rate. In this way, the buyer of a collar ensures that for him the interest rate effectively stays within a specified range whose upper and lower limits are established by the cap and floor rates, respectively. The purpose of a collar is to reduce the costs of the cap involved. The long position in the collar receives a premium through the short floor position which lowers the price of the long cap position. However, an investor can then only profit from falling interest rates up to the lower interest rate limit of the collar.

Swaptions are interest rate instruments having a swap as an underlying. They thus give the holder the right to enter into the underlying swap. The strike of the swap is quoted in the form of the specified *swap rate.* Both "cash" and "swap" are typical as modes of settlement; upon exercise of the swap, either a cash payment is made in settlement or the underlying swap is actually entered into.

In Part III, the structure and function of these and other financial instruments will be discussed in far greater detail. Perhaps the most important point in understanding financial instruments is their valuation and in this connection, the concept of *hedging.* Hence, the most important methods of pricing and hedging will be introduced in the following Part II before we subsequently apply them to today's most common financial products in Part III.

Part II

Methods

5

Overview of the Assumptions

To apply the common methods for pricing and risk management we need to make assumptions which are necessary for the construction of the associated models. A complete list of all model assumptions made in this book are summarized here in order to provide an overview of the numerous conditions and assumptions arising in the various methods. For each pricing and risk management method discussed in the following chapters, we will specify which of these assumptions are needed for their application.

1. Given identical costs and risks, investors always prefer the strategy earning the greatest profit.
   This is one of the fundamental motivations for trading. No one deliberately forgoes a profit.

2. There are no arbitrage opportunities.
   Should opportunities exist allowing a profit to be made without risk, they would be exploited by arbitrageurs until an "equilibrium" is reestablished in the market.

3. The markets have infinite liquidity.
   Unlimited quantities of financial instruments are available for purchase and sale at any time without affecting their price. As is indicated in its the formulation, this assumption becomes increasingly questionable the more illiquid the market for the instrument becomes. In practice, liquidity risks are reflected in broad bid/ask spreads. Thus, Assumption 5 usually implies Assumption 3.

4. There is no counterparty risk.
   The price of an instrument is solely a function of market parameters (prices of the risk factors, underlyings, volatility) and is in no way dependent on the instrument's seller. This is a good approximation for products traded on an exchange which also acts as a clearing house. This, of course, is not the case when dealing with OTC transactions. The quantitative treatment of credit and counterparty risks is presently a rapidly growing and increasingly important branch of risk management. We refer the reader to Chapter 20.

5. There is no "friction" in the market.
   Market friction is a general term referring to all costs involved in trading. These include transaction costs, bid/ask spreads, (opportunity) costs of margins, taxes, etc. All such costs not belonging to the actual *investment* are neglected. This is by no means a bad approximation for large institutions.

6. Continuous trading is possible.
   This means that the time between two trades can be arbitrarily small as can the differences in price and the number of instruments traded. The assumption that trading can take place in arbitrarily short time intervals is an approximation for the simple reason that no exchange is *always* open for trade (weekends, bank holidays). Furthermore, infinitely many adjustments in a continuous hedging strategy would cause infinitely high transaction costs if Assumption 5 did not hold. This means that Assumption 5 must hold if Assumption 6 is to make sense. The assumption that arbitrarily small price differences are possible can be problematic for options which are well out of the money. For example, the smallest possible price change of bond futures ("tick size") usually is 0.01 which is small in comparison to the futures quote, say 99.8. Therefore, the assumption of continuously changing prices is good for futures. However, for an *option* on the future which is far out of the money, a change in the futures price by one tick can result in a change in the price of the option of 50% of its value. Thus, such options do not have continuous price changes.

7. The logarithm of the relative price change of a risk factor is a random walk.
   In light of Section 3.3, this implies that the price of a risk factor is lognormally distributed.

8. Interest rates are not stochastic.
   Intuitively, this means that the evolution of interest rates over time does not involve a random component but is completely deterministic. In such a world, future interest rates are known today.

9. Interest rates are constant.
   This means that future interest rates are not only known today, they have today's value as well. Interest rates do not change in the course of time.

10. The volatilities of risk factors are not stochastic.
    Intuitively, this means that the volatilities as a function of time do not involve a random component but are completely deterministic. Future volatilities are known today.

11. The volatilities of risk factors are constant.
    This means that future volatilities are not only known today, they have the same value as today as well. Volatilities do not change in the course of time.

12. Dividend yields of risk factors are constant.
    In many pricing models for derivatives, the dividend (yield) of an underlying is an input parameter which is assumed to be constant throughout the term of the derivative contract.

13. Correlations between the risk factors are not stochastic.
    As with volatilities, the evolution of the correlations over time has no random component. Future correlations are known today. As in the case of volatilities, this is not often the case in reality.

6

Present Value Methods, Yields and Traditional Risk Measures

Present value methods determine the value of a financial instrument by discounting all future cash flows resulting from the instrument. Applying this method requires few assumptions. Only Assumptions 1, 2, 3, 4 and 5 from Section 5 are necessary.

6.1 Present Value and Yield to Maturity

Suppose that between the time t and maturity T of a security, there are n dates t_i on which cash flows $C(t_i), i = 1, ..., n$ are due. The *present value* $V_R(t,T)$ of a security [1] is the sum of all *future* cash flows due between t and T, where each cash flow $C(t_i)$ is discounted at the spot rate belonging to the maturity t_i of the corresponding cash flow:

$$V_R(t,T) = \sum_{i=m+1}^{n} B_R(t,t_i)\, C(t_i) \tag{6.1}$$

where

$$\underbrace{T_E = t_0 < t_1 < \cdots < t_{m-1} < t_m}_{\text{Past}} \leq t < \underbrace{t_{m+1} < t_{m+2} < \cdots < t_n = T}_{\text{Future}}$$

Here, T_E denotes the security's date of issue. If all the cash flows are known, the valuation of the security follows from the current zero coupon bond rates (spot rates) R. Whenever dealing with interest rate instruments, we use the notation $t_0 = T_E$ for the date of issue, t_m for the date of the last cash flow before the present value date t, and $t_n = T$ for the maturity date. Consequently, all future cash flows will occur

[1] In this book, the value of a financial instruments will usually be denoted by V. In general, V will often have an index (R, for example) to signify its dependence on a specific parameter.

between times t_{m+1} and t_n while all past cash flows have occurred between times t_0 and t_m.

Since the reader may be unfamiliar with the general notation for the discount factor used throughout this book, examples are provided in this section expressing the present values and sensitivities explicitly for continuous and discrete compounding. This is accomplished by substituting the general expression for the discount factor with the expressions for the desired compounding method found in Table 3.4. Substituting in Equation 6.1 with continuous compounding yields a present value of

$$V_R\,(t,T) = \sum_{i=m+1}^{n} e^{-R(t_i-t)} C(t_i)$$

Likewise, for discrete compounding

$$V_R(t,T) = \sum_{i=m+1}^{n} \frac{C(t_i)}{(1+R)^{(t_i-t)}}$$

An important parameter of a bond is its *yield to maturity*, abbreviated by *YTM*. This is the compounding rate obtained if the bond is held until maturity while reinvesting all cash flows received immediately at the same rate, these also being held until maturity. If the interest rate term structure were perfectly flat until the bond matures (i.e. constant with yield $\overline{R}$) this would be equal to the yield to maturity. Thus, instead of discounting using the current zero coupon bond yields, the cash flows can be discounted using the yield to maturity to obtain the current present value, called the *dirty price* of the bond. Therefore, $\overline{R}$ is given implicitly by Equation 6.1: if all future cash flows are discounted at the same yield to maturity $\overline{R}$, the resulting value must be equal to the current present value of the security, i.e.

$$\sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i)\, C(t_i) = \sum_{i=m+1}^{n} B_R(t,t_i)\, C(t_i) \tag{6.2}$$

This establishes the relationship between the YTM of the instrument and the current spot rates. In general, the yield to maturity cannot be calculated analytically from this equation; $\overline{R}$ must usually be computed using a numerical iteration procedure.

The yield of a security over a term less than its maturity is not equal to its YTM (this is already indicated by the name "yield *to maturity*"!) but rather must be equal to the current spot rate corresponding to that term (and the credit rating of the instrument); otherwise, the investor is presented with an arbitrage opportunity. If, for instance, short maturity periods pay less interest than long periods, an investor could employ the following strategy: a 3-month loan is used to by a 10-year bond. During the three months, the investor holds the 10-year bond and earns interest at the YTM rate (because of the interest accrued, coupon payments in this time are irrelevant to the argument). After three months, the bond is sold and the loan repaid. As the interest earned from the bond in the three months exceeds the interest paid on the 3-month loan, this strategy would lead to a profit without risk or investment capital. In order to eliminate this arbitrage, the price of the bond today is adjusted by the market (i.e.

higher than its forward price) in such a way that, during the three months, the bond does not earn interest according to its yield to maturity but in accordance with the 3-month interest rate.

6.2 Internal Rate of Return and Net Present Value

The *internal rate of return* (*IRR*), denoted by R_E, is defined as the interest rate which, when applied to *all* (including past) cash flows associated with an instrument or portfolio, makes the of the sum of all these cash flows equal to zero. Here, past cash flows are compounded at the internal rate of return while those in the future are discounted at the same rate, thus

$$\begin{aligned} 0 = \sum_{j=1}^{m} \underbrace{B_{R_E}(t_j, t)^{-1} C(t_j)}_{\text{Past Cash Flows}} + \sum_{i=m+1}^{n} \underbrace{B_{R_E}(t, t_i) C(t_i)}_{\text{Future Cash Flows}} \\ \text{where } t_1 < \cdots < t_m \leq t \leq t_{m+1} < \cdots < t_n = T \\ \text{and } \ 1 \leq m, m+1 \leq n \end{aligned} \tag{6.3}$$

This implicit definition of the internal rate of return can be used to solve for R_E via a numerical iteration procedure. Neither spot rates nor present values appear in the above condition, only the cash flows. For instruments paying a floating rate, the cash flows are affected by the market, but for fixed rate instruments, all cash flows are determined in advance and thus for such instruments the internal rate of return is completely independent of the current market. It is, however, influenced by the market situation at the time the instrument was purchased since the purchase price $C(t_1)$ of the instrument depends on the market situation and appears in the above equation (with a negative sign since it is a payment).

Consider for example a simple bond with a term of 6.5 years paying an annual rate of 6% on its nominal. Let's assume that the bond paid its last coupon 6 months previous to the present date and was purchased 2 months ago for 102.5% (of its nominal value) plus the accrued interest for the four months (in this example, 2%). All future cash flows (coupons and redemption upon maturity) are incoming and thus positive. In this example, the internal rate of return must be chosen so that the discounted future cash flows are equal to the purchase payment compounded at the same rate.

The internal rate of return can be interpreted as the constant interest rate at which the portfolio yield compounds the investments in the portfolio for the life of the portfolio (i.e. from purchase to sale or to the due date of the last cash flow)[2].

Starting from the above equation defining the internal rate of return, we can define a further quantity, known as the *net present value* or *NPV*. The NPV of an instrument is its current present value plus all the past cash flows compounded at the internal rate

[2]Often in definitions of the internal rate of return (and of the NPV) the future cash flows are discounted, while the past cash flows are *not* compounded. This inconsistency makes a reasonable interpretation of the NPV and IRR practically impossible. The definitions of NPV and IRR introduced here are more general and accurate.

of return:

$$NPV(t,T) = \sum_{j=1}^{m} \underbrace{B_{R_E}(t_j,t)^{-1} C(t_j)}_{\text{Past Cash-Flows}} + \underbrace{\sum_{j=m+1}^{n} B_R(t_i,t) C(t_i)}_{\text{Present Value}}$$
$$\text{where } t_i < \cdots < t_m \leq t \leq t_{m+1} < \cdots < t_n = T$$
$$\text{and } 1 \leq m, m+1 \leq n \tag{6.4}$$

The difference between this and Equation 6.3 is that *future* cash flows are not discounted at the IRR but at the *spot rates* and thus represent the present value. According to Equation 6.3, the sum of the *past* cash flows compounded at the IRR is exactly the negative value of the sum of the *future* cash flows discounted at the IRR. Thus, the NPV can be interpreted as the difference between the sum of the future cash flows discounted at the IRR on the one hand and at the spot rates on the other:

$$NPV(t,T) = \sum_{j=m+1}^{n} \left[B_R(t_i,t) - B_{R_E}(t_j,t) \right] C(t_i)$$

According to the definition of the YTM given in Equation 6.2, the sum of future cash flows discounted at the YTM instead of the spot rates also yields the present value. Thus, if the YTM equals the IRR, the NPV will be zero:

The instrument's internal rate of return equals its yield to maturity if and only if the net present value is zero.

If the IRR is higher (lower) than the YTM, the instrument will have a positive (negative) NPV.

In Figure 6.1 (from the sheet IRR in the Excel workbook PlainVanilla.xls on the accompanying CD-ROM), the YTM, IRR and NPV have been calculated for a portfolio. Only the cash flows for the portfolio are required. The actual *instruments* generating these cash flows are irrelevant. In the column "Discounted Using Spot Rate", the *future* cash flows have been discounted at the current spot rates. The sum of these values yields the present value of the portfolio. In the column "Discounted Using YTM", the *future* cash flows are discounted at the YTM. In accordance with Equation 6.2, the YTM was adjusted until the sum of the discounted cash flows was equal to the present value calculated using the current spot rates (we have utilized the Excel function "Goal Seek" to make the field "PV Difference" equal zero). In the column "Discounted/Compounded Using IRR" *all* cash flows were discounted or compounded, respectively, at the IRR, adjusting the IRR until the sum of all cash flows in this column equaled zero (see Equation 6.3). Finally, the NPV was determined according to Equation 6.4 as the sum of the present value and the past cash flows compounded at the IRR. This is the difference between the future cash flows discounted at the IRR and YTM, respectively. The NPV is negative and the IRR of the portfolio lies significantly under the current YTM.

The IRR and NPV are suitable for measuring the *performance* of a portfolio (or a trader), but are by no means parameters relevant to valuation or risk management. This is due to the fact that the *past* cash flows influence these parameters. But the

Yield Curve			Portfolio Cash Flows			
Payment Date	Time 30/E360	Spot Rate	Cash Flow	Discounted Using Spot Rates	Discounted Using YTM	Discounted/ Compounded Using IRR
Dez-31-99	-1.142					0
Jun-30-00	-0.642		-323,784			-333,439
Dez-31-00	-0.142		50,000			50,325
Feb-21-01			-189,774			-189,774
Jun-30-01	0.358	2.20%	50,000	49,612	48,769	49,186
Dez-31-01	0.861	3.20%	50,000	48,662	47,093	48,067
Jun-30-02	1.358	4.50%	60,000	56,518	54,590	56,382
Dez-31-02	1.861	5.50%	60,000	54,310	52,714	55,099
Jun-30-03	2.358	6.30%	60,000	51,949	50,922	53,858
Dez-31-03	2.861	7.10%	70,000	57,526	57,367	61,404
Jun-30-04	3.358	7.90%	70,000	54,225	55,417	60,022
Dez-31-04	3.861	8.50%	70,000	51,086	53,513	58,656
Jun-30-05	4.358	9.00%	-15,000	-10,303	-11,077	-12,286
Dez-31-05	4.861	9.30%		0	0	0
Jun-30-06	5.358	9.50%	40,000	24,596	27,554	31,297
Dez-31-06	5.861	9.60%		0	0	0
Jun-30-07	6.358	9.70%		0	0	0
Dez-31-07	6.861	9.80%	30,000	15,796	18,614	21,912
Jun-30-08	7.358	9.90%	-15,000	-7,489	-8,991	-10,709
Values as at		**YTM and IRR**			**7.20%**	**4.69%**
Feb-21-01		**Past Cash Flows**				-472,887
		Future Cash Flows		**446,487**	**446,487**	**472,887**
		NPV		**-26,400**		

Figure 6.1: Cash flow table of a portfolio.

value of an instrument (and thus its risk resulting from a potential change in its value) is independent of its history. The market establishes a *single* price for an instrument independent of the cash flows that may have been generated by the instrument previously. Since the risk of an instrument is by definition a reflection of the potential *future* change in the its price, the same argument holds for risk management.

For example, a single instrument could have different NPVs and IRRs, if a trader bought the instrument in the past at a cheaper price than another trader at the same time, or if two traders bought the same instrument at the same price but on different dates. It is in fact quite unlikely that the same instrument in different portfolios will have exactly the same NPV and IRR since for this to happen, the instruments must be purchased at exactly the same time and price (or the price difference must correspond exactly to compounding at the IRR for the period between the differing purchase dates).

6.3 Accrued Interest, Residual Debt and Par Rates

The *residual debt* of an instrument is the amount of outstanding debt on which interest must be paid, i.e. the open claims against the issuer of the instrument. The *accrued*

interest is the outstanding interest having accumulated at a contractually agreed fixed interest rate K in the time between the last payment date t_m and t (=today). Accrued interest only makes sense when dealing with fixed rate instruments or for those instruments whose interest rates are established at the beginning of each interest period. In our general notation

$$\text{Accrued interest} = N(t_m)\left[B_K(t_m,t)^{-1}-1\right] \text{ where } t_m \le t < t_{m+1}$$

with $N(t_m)$ denoting the residual debt as of the last coupon payment. For instruments which have no amortization payments in the time between today and maturity, the residual debt immediately after a payment date is always equal to the face value N. However, if repayments have already been made by time t_m, the residual debt is accordingly smaller.

Since a fixed interest rate K results in a debt at time t consisting of the residual debt as of the preceding payment date t_m plus the interest accrued since that time, the residual debt at time t is the residual debt from the last payment date compounded up to today at the rate K:

$$\underbrace{N(t)}_{\substack{\text{Residual Debt}\\ \text{at Time } t}} = \underbrace{N(t_m)}_{\substack{\text{Residual Debt}\\ \text{at Time } t_m}} + \underbrace{N(t_m)\left[B_K(t_m,t)^{-1}-1\right]}_{\text{Accrued Interest}} = B_K(t_m,t)^{-1}N(t_m) \tag{6.5}$$

The *clean price* $V_{\text{clean}}(t,T)$ of a financial instrument is defined as the present value less the accrued interest

$$V_{\text{clean}}(t,T) = \sum_{i=m+1}^{n} B_R(t,t_i)C(t_i) - N\left[B_K(t_m,t)^{-1}-1\right]$$

An interest rate instrument (or a portfolio consisting of interest rate instruments) is quoted *at par* when its present value is equal to the residual debt of the instrument, i.e. when

$$N(t) = \sum_{i=m+1}^{n} B_R(t,t_i)C(t_i) \quad \text{where} \quad t_i < \cdots < t_m \le t \le t_{m+1} < \cdots < t_n = T \tag{6.6}$$

or equivalently, when the clean price is equal to the residual debt at the previous payment date $N(t_m) = V_{\text{clean}}(t,T)$. This is a generalization of the usual definition for bonds and applies to all interest rate instruments and even to portfolios with arbitrary cash flows. A simple bond (with the redemption of the full nominal being made at maturity) is quoted at par if the clean price is equal to its face value.

As mentioned before, the advantage of Equation 6.6 is that it holds for all interest rate instruments (including floating rate instruments) and portfolios consisting of interest rate instruments; although the determination of the open claims (current residual debt) of the portfolio may become quite tedious. If the present value of the portfolio, i.e. the sum of its cash flows discounted at current spot rates, is equal to the sum of the open claims, then the portfolio is quoted at par.

The *par rate* of a fixed rate instrument (for example, bonds, fixed rate mortgages, etc.) is defined as the rate K which the instrument should have as its fixed rate in

		replace		**vary rate so that the sum of discounted futures cash flows equals**		
		Spot Rate	Coupon Rate	Present Value	Residual Dept	Sum of Compounded Past Cash Flows
by	IRR	**x**				**x**
	Par Rate		**x**		**x**	
	YTM	**x**		**x**		

Figure 6.2: Summary of the different yields.

order for the instrument to be quoted at par. Par rates are only defined for fixed rate instruments. For such instruments, the residual debt and the cash flows are dependent on K. The at-par condition given by Equation 6.6 (with Equation 6.5 for $N(t)$) solved for K yields the par rate of the instrument as a function of the current spot rate. Of particular importance are the par rates of coupon bonds paying an annual coupon calculated using discrete compounding. Graphing these par rates as a function of time to the maturity of the coupon bonds gives a curve called the *par yield curve*.

To summarize, Figure 6.2 comprises a list of all interest rates introduced in this section. The table indicates how each of the three interest rates just presented, namely the IRR, the par rate and the YTM, can be calculated from those established by the market (spot rate) and the instrument (coupon rate). For example, the combination "replace spot rate by YTM" means that the future cash flows should be discounted at the YTM rather than the spot rates. This rate (i.e. the YTM) is then varied until the sum of discounted future cash flows equals the target variable, in this case the present value.

From this table it is apparent that the par rate is a function of the spot rates but *not* of the coupon rate (the par rate replaces the coupon rate). Conversely, the YTM is a function of the coupon rate and *not* explicitly of the spot rates (these are replaced by the YTM). Naturally, the YTM depends implicitly on the spot rates since the spot rates affect the present value (whose value should be attained by adding up the future cash flows discounted at the YTM). Thus, the par rate of an instrument changes when the spot rate changes, in contrast to the YTM which changes when either the coupon rate or the spot rates change. From this behavior, it follows that the par rate and the YTM can only be equal for a particular combination of spot rates (determined by the market) and coupon rates (given by the instrument):

Par rate and yield to maturity of an interest rate instrument are equal if and only if the instrument is quoted at par.

To conclude this section, we want to "prove" this statement: the par rate of a fixed rate instrument is defined as the coupon rate K which the instrument should have for the present value to be equal to the residual debt. On the other hand, the present value can be expressed in terms of the YTM through Equation 6.2. For an instrument

whose coupon rate is equal to the par rate we have on a payment date, i.e. for $t = t_m$:

$$N(t_m) = \sum_{i=m+1}^{n} B_{\overline{R}}(t_m, t_i) C(t_i)$$

Multiplying both sides of the equation by the coupon rate compounding factor $B_K(t_m, t)^{-1}$, we find that the left-hand side corresponds to the residual debt at time t (see Equation 6.5). Only on the condition that K is equal to the YTM can we make use of Equation 3.4 to transform the residual debt into the present value at time t (in accordance with Equation 6.2) :

$$\begin{aligned} \underbrace{B_K(t_m,t)^{-1} N(t_m)}_{\text{Present Value for } \bar{\mathrm{R}} = \mathit{Par\ Rate}} &= B_K(t_m,t_i)^{-1} \sum_{i=m+1}^{n} B_{\overline{R}}(t_m,t_i) C(t_i) \\ &= \underbrace{\sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i) C(t_i)}_{\text{Present Value for } \bar{\mathrm{R}} = \mathit{YTM}} \end{aligned} \tag{6.7}$$

Thus, it follows that if the coupon rate of an instrument is equal to its YTM, the present value is equal to the residual debt, i.e. the at-par condition given by Equation 6.6 is satisfied and therefore, the coupon rate must be equal to the par rate. Conversely, the at-par condition can only be satisfied for all times t if the par rate is equal to the YTM.

6.4 Traditional Sensitivities of Interest Rate Instruments

6.4.1 Average Lifetime and Macaulay Duration

The simplest feature characterizing a security whose value derives from future cash flows is its mean (residual) lifetime. The mean is the weighted average taken over the time periods until each future coupon payment. The weight accorded to a particular period is exactly equal to the contribution of the corresponding cash flow to the present value of the security. This mean lifetime is called the *Macaulay duration.*

$$\begin{aligned} D_{\text{Macaulay}}(t) &= \sum_{i=m+1}^{n} (t_i - t) \frac{B_R(t,t_i)\, C(t_i)}{V_R(t,T)} \\ &= \sum_{i=m+1}^{n} (t_i - t) \left[\frac{B_R(t,t_i)\, C(t_i)}{\sum_{k=m+1}^{n} B_R(t,t_k)\, C(t_k)} \right] \end{aligned} \tag{6.8}$$

The Macaulay duration has an interesting property: If a time interval of length dt (during which no cash flow payment was due) has passed since the last calculation of

the Macaulay duration, the new Macaulay duration is simply given by the difference of the old value less the elapsed time dt.

$$D_{\text{Macaulay}}(t+dt) = D_{\text{Macaulay}}(t) - dt \quad \text{if} \quad t+dt < t_{m+1}$$

The derivation of this result is trivial. We only need to make use of the fact that the sum over all weights equals one:

$$\begin{aligned} D_{\text{Macaulay}}\ (t+dt) &= \sum_{i=m+1}^{n} (t_i - t - dt) \left[\frac{B_R(t,t_i)C(t_i)}{V_R(t,T)} \right] \\ &= \underbrace{\sum_{i=m+1}^{n} (t_i - t) \left[\frac{B_R(t,t_i)C(t_i)}{V_R(t,T)} \right]}_{D_{\text{Macaulay}}\ (t)} - dt \underbrace{\left[\frac{\sum_{i=m+1}^{n} B_R(t,t_i)C(t_i)}{\sum_{k=m+1}^{n} B_R(t,t_k)C(t_k)} \right]}_{1} \end{aligned}$$

6.4.2 Modified Duration and Convexity

The methods for pricing financial instruments introduced in this book are applied in risk management when determining the value of a portfolio in any given scenario. This can be extremely time consuming. In order to obtain a quick risk assessment without repeating time-consuming pricing calculations for each different scenario, *approximation techniques* have been developed which are all based on the same procedure:

Let V be the price of a financial instrument and R the price of the underlying risk factor (for instance an interest rate), also called the underlying. One obvious measure for the risk of this instrument is the sensitivity of its price with respect to changes in the underlying. The price is generally a complicated function of R which, however, admits a *Taylor series* representation.

$$V_{R+dR}(t,T) = V_R(t,T) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k V_R(t,T)}{\partial R^k} (dR)^k$$

In such a Taylor series, the price of the instrument whose underlying has changed by a factor of dR is calculated by adding the infinite series above to the price corresponding to the original value of the underlying R. The terms of the sum involve all powers $(dR)^k$ of the change in the underlying dR. If this change is small (for example, 1 basis point = 0.01% = 0.0001), its powers are particularly small for large k (for example, $(0.0001)^2 = 0.00000001$, $(0.0001)^3 = 0.000000000001$, etc.). Thus, for small changes in R, we can obtain a good approximation of the price by truncating the series after the first few terms[3]. In this way the infinite sum is approximated by

$$dV_R(t,T) = \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k V_R(t,T)}{\partial R^k} (dR)^k \approx \frac{\partial V_R(t,T)}{\partial R} dR + \frac{1}{2} \frac{\partial^2 V_R(t,T)}{\partial R^2} (dR)^2 \qquad (6.9)$$

[3]The method of expanding a function in a Taylor series for small changes in its parameters, neglecting all but the parameter terms up to a certain power (here, the second power) is well established and frequently used in practice with verifiable correctness (assuming the variables under consideration are continuous). This is not simply an academic "mathematical trick" but a method which finds constant application in the financial world. Almost all hedging strategies as well as the variance-covariance methods used in calculating the value at risk, for example, find their basis in this procedure. The same method has already been applied when we presented Ito's lemma in Section 3.4.2.

where $dV_R = V_{R+dR} - V_R$ denotes the difference between the old and new price of the instrument. In the traditional terminology of interest rate management and if R represents the yield to maturity, the coefficients of the first two powers of dR (divided by the original price V) are referred to as the *modified duration* and the *convexity*. In modern risk management, these coefficients are called *delta* and *gamma*, where R is any arbitrary underlying.

More precisely, the modified duration is defined as the (negative) derivative of the price with respect to $\overline{R}$ divided by the price, i.e. the (negative) relative sensitivity of the security to linear changes in the yield to maturity, see Equation 6.10. The minus sign reflects the effect that bond prices decline as interest rates rise. Assuming that the cash flows themselves are independent of the interest rate, Equation 6.2 gives the following expression for the modified duration:

$$D_{\text{mod}} = -\frac{1}{V}\frac{\partial V}{\partial \overline{R}} = -\sum_{i=m+1}^{n} \frac{C(t_i)}{V}\,\frac{\partial B_{\overline{R}}(t,t_i)}{\partial \overline{R}} \tag{6.10}$$

This is the general expression for all compounding methods. For example, for continuous compounding it reads explicitly

$$\begin{aligned} D_{\text{mod}} &= -\frac{1}{V}\frac{\partial V}{\partial R} = -\frac{1}{V}\sum_{i=m+1}^{n}(t_i - t)C(t_i)e^{-\overline{R}(t_i-t)} \\ &= -\sum_{i=m+1}^{n}(t_i - t)\left[\frac{C(t_i)e^{-\overline{R}(t_i-t)}}{\sum_{j=1}^{n} C(t_j)e^{-\overline{R}(t_i-t)}}\right] = D_{\text{Macaulay}} \end{aligned}$$

The last expression allows an intuitive interpretation: it is a weighted average of the "future cash flow maturities" $t_i - t$, where the maturity of each cash flow contributes to the average with the weight specified in brackets. This weight is just the proportional contribution of the corresponding cash flow to the present value of the security. The modified duration is thus, for continuous compounding, equal to the Macaulay duration. This result makes it clear that the Macaulay duration (which was defined as the average cash flow maturity) contains information about the sensitivity of the security with respect to yield changes and can thus be used as a risk measure.

For annual compounding, the modified duration given in Equation 6.10 becomes

$$\begin{aligned} D_{\text{mod}} &= \frac{1}{V}\frac{\partial V}{\partial R} = -\frac{1}{V}\sum_{i=m+1}^{n}(t_i - t)\frac{C(t_i)}{(1+\overline{R})^{(t_i-t+1)}} \\ &= \frac{1}{1+\overline{R}}\sum_{i=m+1}^{n}(t_i - t)\left[\frac{C(t_i)/(1+\overline{R})^{(t_i-t)}}{\sum_{i=m+1}^{n} C(t_j)/(1+\overline{R})^{(t_j-t)}}\right] = \frac{D_{\text{Macaulay}}}{1+\overline{R}} \end{aligned}$$

Thus, for annual compounding the modified duration is *not* equal to the Macaulay duration (the weighted average of the cash flow maturities). Instead, the Macaulay

duration must be modified by a factor $1/(1+\overline{R})$. This is the reason for the name *modified* duration. Note that a modification is unnecessary for continuous compounding. It is thus solely an effect of the compounding method applied.

Regardless of whether the compounding is computed continuously or annually, the unit of both the Macaulay duration and the modified duration is "years" since both deal with an average duration multiplied, in the case of the modified duration and annual compounding, by the dimensionless factor $1/(1+\overline{R})$.

According to Equation 6.9, the modified duration is defined precisely so that it gives a linear approximation of the relative price change when multiplied by the change in the yield to maturity:

$$dV \approx \frac{\partial V}{\partial \overline{R}} d\overline{R} = -V D_{\text{mod}}\, d\overline{R} \quad \Longrightarrow \quad \frac{dV}{V} = -D_{\text{mod}}\, d\overline{R} \tag{6.11}$$

Setting $d\overline{R} = 1\%$ *per year* makes the numerical value of relative price dV/V (measured in percent) equal to the *numeric value* of the modified duration. This provides an intuitive interpretation of the modified duration: the numeric value of the modified duration is equal to the percentage change in the price when the yield changes by 1% per year. This interpretation of the numeric value is often (incorrectly) used as the definition of the modified duration itself.

Substituting one basis point for the change in yield, i.e. $d\overline{R} = 0,01\%$ per year instead of 1% per year, we obtain the *absolute* price change $dV = -V D_{\text{mod}} dR$ referred to as the *base point value*.

In order to obtain a more precise measure of the sensitivity of a security with respect to its yield to maturity, the linear approximation can be improved by including the second order term of the above mentioned Taylor series of the price function (Equation 6.9) in our considerations. This term measures the sensitivity of the present value with respect to the square of the change in yield. Intuitively, it describes the curvature of the price curve and is therefore referred to as the *convexity*. Assuming that the cash flows themselves are independent of the YTM and making use of Equation 6.2, the convexity of the present value is given by

$$\frac{1}{V}\frac{\partial^2 V}{\partial\, \overline{R}^2} = \sum_{i=m+1}^{n} \frac{C(t_i)}{V}\, \frac{\partial^2\, B_{\overline{R}}(t,t_i)}{\partial\, \overline{R}^2} \tag{6.12}$$

This general expression holds for all compounding methods. When the interest is compounded continuously, it reads explicitly

$$\frac{1}{V}\frac{\partial^2 V}{\partial \overline{R}^2} = \frac{1}{V}\sum_{i=1}^{n}(t_i - t)^2 C(t_i) e^{-\overline{R}(t_i - t)} = \sum_{i=1}^{n}(t_i - t)^2 \left[\frac{C(t_i)/e^{-\overline{R}(t_i - t)}}{\sum_{j=1}^{n} C(t_j)/e^{-\overline{R}(t_j - t)}} \right]$$

For annual compounding, the explicit expression is given by

$$\begin{aligned}\frac{1}{V}\frac{\partial^2 V}{\partial \overline{R}^2} &= -\frac{1}{V}\sum_{i=m+1}^{n}\frac{\partial^2}{\partial \overline{R}^2}\frac{(t_i-t)C(t_i)}{(1+\overline{R})^{(t_i-t+1)}} \\ &= \frac{1}{(1+\overline{R})^2}\sum_{i=m+1}^{n}(t_i-t)(t_i-t+1)\left[\frac{C(t_i)/(1+\overline{R})^{(t_i-t)}}{\sum_{j=m+1}^{n}C(t_j)/(1+\overline{R})^{(t_j-t)}}\right]\end{aligned}$$

As is the case for the duration, the cash flow weights appear in the brackets. The convexity can thus be interpreted as the weighted average of the *squared* cash flow maturities.

As yet, the sensitivities have only been considered with respect to the yield to maturity. We could also calculate the sensitivities with respect to the current spot rates. To do so, the original form of the present value given by Equation 6.1 is differentiated with respect to the spot rates valid for the cash flow maturities, called the *key rates*. In general, the derivative of a price with respect to the current spot rate is referred to as the *yield delta*. In analogy to the modified duration, the *key rate durations* are defined as the negative yield deltas with respect to the key rates divided by the price:

$$D_i^{\text{key}} = -\frac{1}{V}\underbrace{\frac{\partial V}{\partial R(t_i)}}_{\text{Yield Delta}} = -\frac{1}{V}\underbrace{C(t_i)\frac{\partial B_R(t,t_i)}{\partial R(t_i)}}_{\text{Yield Delta}} \tag{6.13}$$

A linear approximation of the proportional change in the present value of the instrument resulting from a change in the term structure of interest rates is thus

$$\frac{dV}{V} = \frac{1}{V}\sum_i \frac{\partial V}{\partial R(t_i)}dR(t_i) = -\sum_i D_i^{\text{key}}dR(t_i) \tag{6.14}$$

Note that the change in the key rates $R(t_i)$ need not be the same for each i; the yield curve could be "bent". This method allows risk calculations for far more complicated scenarios in comparison to just using modified duration which, as we recall, is based on a flat term structure of interest rates at the level of the yield to maturity. The modified duration can therefore only show the effect of a parallel shift of the term structure.

6.4.3 Summation of Traditional Sensitivities

In contrast to the "modern" sensitivities (delta, gamma, vega, etc.) of options etc., which give the *absolute* sensitivity (the amount of *money* by which the value changes when a parameter is modified), the traditional sensitivities indicate a *relative* sensitivity (the *percentage* change in the value of an instrument when a parameter is modified). In consequence, modern sensitivities of instruments can simply be added up in order to obtain the total sensitivity of all instruments in a given portfolio with respect to a particular parameter (net delta, net gamma, etc.). This is not the case for the traditional sensitivities.

The second and more important difference between traditional and modern sensitivities is that the modern sensitivities refer to *market* parameters such as spot rates

which are relevant for most instruments. The traditional sensitivities, however, most often refer to *instrument-specific* properties, such as yield to maturity. For example, how should we interpret the modified duration of a portfolio consisting of a 5-year bond and a 10-year bond? The sensitivity with respect to the 5-year YMT or the 10-year YMT? Despite these difficulties, it is possible to define total traditional sensitivities for an entire portfolio. To do so, we need to perform the following four steps:

1. First, a mark-to-market valuation of the entire portfolio is performed in accordance with 6.1 by discounting all future cash flows at the current spot rates.

2. The YTM of the entire portfolio is determined according to Equation 6.2. All cash flows are discounted at the same interest rate. This rate is varied until the present value calculated in Step 1 is attained.

3. In the equations for the sensitivities of each of the individual instruments, the YTM of the each instruments is replaced by the YTM of the entire portfolio. For example, consider the modified duration of a zero coupon bond at time t with maturity T which, in discrete compounding, is given by „$(T-t)/(1+R)$", where R denotes the YTM of the zero coupon bond. This YTM is replaced by the YTM of the entire portfolio yielding the sensitivity of the zero bond with respect to the YTM of the portfolio.

4. Each sensitivity of a particular instrument (with respect to the YTM of the portfolio) is multiplied by the present value of the corresponding instrument. The sum of these products divided by the present value of the portfolio is the desired sensitivity of the portfolio.

The last two steps can be combined in one equation. The modified duration and convexity of a portfolio consisting of two instruments A and B with respect to the YTM $\overline{R}_{A+B}$ of the portfolio, and the key rate duration of this portfolio with respect to a spot rate R (which is the same for all instruments in the portfolio) are[4]:

$$\begin{aligned} D_{\text{mod}}(A+B)_{\overline{R}_{A+B}} &= \frac{1}{A+B}\left(A\, D_{\text{mod}}(A)_{\overline{R}_{A+B}} + B\, D_{\text{mod}}(B)_{\overline{R}_{A+B}}\right) \\ \text{Convexity}(A+B)_{\overline{R}_{A+B}} &= \frac{1}{A+B}\left(A\,\text{Convexity}(A)_{\overline{R}_{A+B}} + B\,\text{Convexity}(B)_{\overline{R}_{A+B}}\right) \\ D_{\text{key}}(A+B)_R &= \frac{1}{A+B}\left(A\, D_{\text{key}}(A)_R + B\, D_{\text{key}}(B)_R\right) \end{aligned} \tag{6.15}$$

We have denoted the present value of instrument A by simply A, and analogously for instrument B. The situation for the key rate duration is somewhat simpler since this

[4]The derivation of these equations is trivial and will only be shown for the modified duration:
$D_{\text{mod}}(A+B)_{\bar{R}_{A+B}} = -\frac{1}{A+B}\frac{\partial(A+B)}{\partial \bar{R}_{A+B}} = \frac{1}{A+B}\left(-\frac{\partial A}{\partial \bar{R}_{A+B}} - \frac{\partial B}{\partial \bar{R}_{A+B}}\right)$
$= \frac{1}{A+B}\left(\underbrace{-\frac{1}{A}\frac{\partial A}{\partial \bar{R}_{A+B}}}_{D_{mod}(A)} A + \underbrace{-\frac{1}{B}\frac{\partial B}{\partial \bar{R}_{A+B}}}_{D_{mod}(B)} B\right)$

is a sensitivity with respect to the spot rates. Spot rates are not instrument-specific parameters and consequently the first two steps need not be performed.

Traditional sensitivities can be calculated in this way even for modern instruments such as futures. A future corresponds to a portfolio consisting of the cash flows of the underlying occurring subsequent to the future's maturity plus the cash flow resulting from the future's delivery price upon its maturity. The YTM is calculated for a portfolio with these cash flows. Each individual cash flow is regarded as a zero bond and the modified durations and/or the convexities of these zero bonds are added up according to the above equations.

7

Arbitrage

Arbitrage considerations alone are sufficient for deriving relations such as the put-call parity or determining forward prices. Such arguments require only very few assumptions; we need only assumptions 1, 2, 3, 4 and 5 from Section 5 to be satisfied.

7.1 Forward Contracts

Forward contracts are transactions in which the purchase or sale of an underlying S at a later date T for a fixed price, called the *delivery price* K, is agreed upon as of the current value date t. In this book, the value of such a contract is denoted by $f_S(t, T, K)$ or, for futures contracts (see below), by $F_S(t, T, K)$. If the delivery price at time t is chosen so that the value of the forward or futures contract is zero, then the delivery price is also referred to as the *forward price*. Throughout this book, the forward price of an underlying with a price $S(t)$ at the current time t will be denoted by $S(t, T)$ where the later time T is the maturity of the corresponding forward contract.

7.1.1 Forward Price and Cash & Carry Arbitrage

The following *Cash & Carry Arbitrage* can be used to determine the value of a forward contract $f_S(t, T, K)$ and subsequently the forward price $S(t, T)$ of an underlying with a spot price $S(t)$. Consider the following two portfolios:

- Portfolio structure
  - Portfolio **A** consists of a forward contract on an underlying for which the delivery price K is to be paid at time T and a cash sum invested at an interest rate R to yield precisely the delivery price at time T.
  - Portfolio **B** consists of the underlying and a loan in the amount of the present value of the dividends to be received at time t' between the present time and time T.
- Value of the portfolio at time t

  - Portfolio **A** at time t is worth $f_S(t,T,K)$, the value of the forward contract plus the value $B_R(t,T)K$ of the cash.
  - Portfolio **B** at time t is worth $S(t)$, the value of the underlying, plus the negative (since it is a loan) present value of the dividend payment to be made at the (later) time t', i.e. $-B_R(t,t')D(t')$. The total value of the portfolio is thus the dividend-adjusted spot price as given in Equation 6.3. (The argument is completely analogous for the case of a dividend yield).

- Value of the Portfolio at time T

  - Portfolio **A** at time T is worth $S(T)-K$, the value of the forward contract, plus $B_R(t,T)KB_R(t,T)^{-1}=K$, from the cash, thus having in total a value of $S(T)$. In Portfolio **A**, the delivery price of the forward contract is paid using the cash and the accrued interest earned. The portfolio now consists of the underlying at its current spot price $S(T)$.
  - Portfolio **B** at time T is worth $S(T)+D(t')B_R(t',T)^{-1}$, the value of the underlying plus the dividend paid at t' and then compounded up to time T and, in addition, the amount $-B_R(t,t')D(t')B_R(t,T)^{-1} = -D(t')B_R(t',T)^{-1}$ from the loan. Thus the whole portfolio is worth in total $S(T)$. In Portfolio **B**, the loan is repaid using the dividend payment received at time t' and invested up to time T. As for Portfolio **A**, Portfolio **B** contains exactly the underlying at time T.

Thus, both portfolios have the same value at time T. Their value at earlier times must therefore also be equal since otherwise the investor would be presented with the following arbitrage opportunity: if Portfolio **A,** for example, was worth more than Portfolio **B** at time t, an investor could sell **A** ("go short") and buy **B** with the proceeds keeping the difference. At time T, the value of both portfolios must be the same, as demonstrated above. Thus the short position can be offset exactly by the long position without any additional payment. Thus, a profit can be made without risk or capital by investing according to this strategy at time t. Analogously, if **B** was worth more than **A,** an investor could (short) sell **B** and buy **A.** In order to eliminate such arbitrage opportunities, the value of both portfolios at time t must be equal:

$$f_S(t,T,K) + K\,B_R(t,T) = \overbrace{S(t) - B_R(t,t')D(t')}^{\tilde{S}(t,T)}$$
$$\Rightarrow$$
$$f_S(t,T,K) = \tilde{S}(t,T) - K\,B_R(t,T)$$

The *forward price* $S(t,T)$ of the underlying is now defined as the delivery price K for which the forward contract is worth nothing at time t. This means that for $K = S(t,T)$ we have $f_S = 0$ in the above equation. Thus we obtain for the forward

price the explicit expression:

$$S(t,T) = \frac{\tilde{S}(t,T)}{B_R(t,T)} = \begin{cases} \frac{S(t)}{B_R(t,T)} & \text{No Dividends} \\ \frac{S(t)}{B_R(t,T)} - \frac{D(t')}{B_R(t',T|t)} & \text{Dividend Payment } D \text{ at Time } t' \\ \frac{B_q(t,T)}{B_R(t,T)} S(t) & \text{Dividend Yield } q \end{cases} \tag{7.1}$$

where in the second step Equation 3.5 is used for the dividend-adjusted spot price, and also Equation 3.4 for the case of a dividend *payment*. In this case the forward rate for the period between the dividend date t' and the maturity T is needed.

Conversely, using the forward and spot prices the present value $d(t,T)$ at time t of dividend payments or dividend yields accrued in the time period $T-t$ can be determined. These are referred to as *implied dividends* and *implied dividend yields*, respectively:

$$\begin{aligned} d(t,T) &\equiv S(t) - \tilde{S}(t,T) = S(t) - B_R(t,T)S(t,T) \\ B_q(t,T) &\equiv \frac{\tilde{S}(t,T)}{S(t)} = B_R(t,T)\frac{S(t,T)}{S(t)} \end{aligned} \tag{7.2}$$

7.1.2 The Stochastic Process for the Forward Price

Having established the forward price as a function of the risk factor S as given by Equation 7.1, we make the random walk assumption 3.13 and proceed as in Section 3.4.2, applying Ito's lemma (see Equation 3.18) to obtain a stochastic process for the forward price $S(t,T)$. Making use of Equation 7.1 we define in the case of a dividend yield

$$f(y,t) := S(t,T) = \frac{B_q(t,T)}{B_r(t,T)} e^{y(t)}, \qquad \text{with} \quad y(t) = \ln(S(t))$$

The derivatives needed for the application of Ito's lemma are given by

$$\begin{aligned} &\frac{\partial f}{\partial y} = f = S(t,T), \quad \frac{\partial^2 f}{\partial y^2} = f = S(t,T) \\ &\frac{\partial f}{\partial t} = e^{y(t)} \frac{\partial}{\partial t}\left[\frac{B_q(t,T)}{B_r(t,T)}\right] = S(t)\frac{\partial}{\partial t} e^{(r-q)(T-t)} = -(r-q)S(t,T) \end{aligned}$$

where the last two steps hold for continuous compounding. Substituting this in the Ito formula yields the process for the forward price

$$\begin{aligned} dS(t,T) &= \left(S(t,T)\mu + S(t)\frac{\partial}{\partial t}\left[\frac{B_q(t,T)}{B_r(t,T)}\right] + \frac{1}{2}\sigma^2 S(t,T) \right) dt + S(t,T)\sigma\, dW \\ &= S(t,T)\left(\mu - r + q + \frac{1}{2}\sigma^2\right) dt + S(t,T)\sigma\, dW \end{aligned} \tag{7.3}$$

where, again, continuous compounding is used in the last step.

In anticipation of later chapters we note the following: we will later show that the valuation of derivatives necessitates a particular choice for the drift μ. This choice will

be referred to as the "risk-neutral" value (see for example Equation 11.25). With this choice for the drift, the process for the forward price is reduced to

$$dS(t,T) = S(t,T)\sigma\, dW.$$

This indicates that the forward price process is drift-free in a risk-neutral world.

7.1.3 Forward Positions

As we have seen, $S(t,T)$ is the delivery price of a forward contract agreed upon at time t with maturity T and is referred to as the forward price. This is *not* to be confused with the value of a forward contract (which upon conclusion of the contract is equal to zero, for example). The value at time t of a forward contract entered into at some earlier time $t_0 < t$ with a forward price $K = S(t_0,T)$, valid at that time, can be obtained by considering the following argument:

Entering into a second forward contract (in addition to the first forward contract as described above) at time t for the *sale* of the underlying at maturity T at the current forward price $S(t,T)$ has the effect that at maturity T the underlying is bought for the price K (on the basis of the first forward contract) and immediately sold at the price $S(t,T)$ (on the basis of the second contract). The cash flow at maturity is thus $S(t,T) - K$. This value discounted back to time t gives the value of the portfolio consisting of both forward contracts at time t. Since the value at time t of a forward contract with a delivery price $S(t,T)$ is zero, the value of the portfolio is identical to the value of the forward contract with a delivery price K. Upon consideration of Equation 7.1, we conclude that

$$f_S(t,T,K) = [S(t,T) - K]\, B_R(t,T) = \tilde{S}(t,T) - K\, B_R(t,T) \tag{7.4}$$

7.1.4 Future Positions and Basis Risk

The difference between a forward and a *future* is that futures are instruments which are traded on an exchange. The exchange, taking on the role of a clearing house, demands that margins be paid. The *variation margin* has the effect that the daily fluctuations in the value of a futures position (unrealized profit and loss) are credited or debited to a *margin account* which must be settled at the end of the trading day. Fluctuations in the forward prices therefore result in an *immediate* cash flow rather than one paid at maturity T. For this reason, the cash flow – in contrast to that of a forward contract – is *not* discounted in the valuation of a futures contract. The value of a futures position with the same data as the above defined forward contract is simply:

$$F_S(t,T,K) = S(t,T) - K = B_R(t,T)^{-1}\tilde{S}(t,T) - K \tag{7.5}$$

The forward price $S(t,T)$ is, strictly speaking, the delivery price of a forward contract concluded at time t with maturity T. It can be shown that delivery prices of futures are equal to those of forwards if interest rates are assumed to be deterministic. In this book, it will almost[1] always be assumed that this (i.e. Assumption 8 from Section 5)

[1] Except for the chapters dealing with stochastic interest rates, i.e. Chapters 14 and 15.

is the case when dealing with forwards and futures, i.e. the delivery price for both contracts will be given by Equation 7.1.

The difference between the spot and delivery price is called the *basis*. The basis $b(t,T)$, as can be seen from Equation 7.1, depends on the interest rate and the dividends of the underlying:

$$b(t,T) = S(t,T) - S(t) = \frac{\tilde{S}(t,T)}{B_R(t,T)} - S(t)$$

Basis trading is the attempt to make a profit from the fluctuations in the basis. The value of a portfolio consisting, for example, of a long futures contract with a delivery price K and a short position in the underlying itself is given by

$$F_S(t,T,K) - S(t) = S(t,T) - K - S(t) = b(t,T) - K$$

Since K is constant, the change in the value of this portfolio result solely from the change in the basis. The only risk involved in holding such a portfolio results directly from the basis alone and as such is referred to as the *basis risk*.

7.2 Options

7.2.1 Upper and Lower Bounds for Option Prices

A *plain vanilla European option* gives the holder the same rights as a forward contract but *without* the obligations of the forward contract. For a long forward position, for example, the underlying *must* be bought at the price K agreed to in the contract even if it is available at a lower spot price $S(T)$ at maturity T. The associated call option, however, will be exercised by the holder if doing so does not result in a loss, i.e. the option will be exercised only when $S(T) > K$. This is why the price of a plain vanilla European option is at least as high as that of the corresponding forward position. The price of an *American option*, owing to its additional right allowing the holder to exercise at any time up to maturity, is at least as high as its *intrinsic value* (proceeds obtained by exercising, i.e. payoff profile). Besides, the right to exercise the option early can only increase its value in comparison to its European counterpart. These relationships are summarized in Table 7.1, using Equation 7.4 for the value of a forward position.

Throughout this book we will denote the value at time t of a European call and put on an underlying S with strike price K and maturity T by $c_S(t,T,K)$ and $p_S(t,T,K)$, respectively, as in Table 7.1. For American options, capital letters will be used in the notation, i.e. $C_S(t,T,K)$and $P_S(t,T,K)$.

Any violation of the relations listed in Table 7.1 results in an arbitrage opportunity which can be realized by selling the over-valued instrument and buying the other instrument with the proceeds. If, for example, the price of a call were higher than that of the underlying (the last relation in the first line), an investor could sell the call and buy the underlying with the proceeds, keeping the difference. If the call is exercised, the investor delivers the underlying and receives (in addition to the money already earned from the difference) the strike price K in exchange. If the call is not exercised, the investor receives (again, in addition to the money already earned from

	European
Call	$\max\left\{0, \tilde{S}(t,T) - K\,B_R(t,T)\right\} \leq c_S(t,T,K) \leq C_S(t,T,K)$
Put	$\max\left\{0, K\,B_R(t,T) - \tilde{S}(t,T)\right\} \leq p_S(t,T,K) \leq P_S(t,T,K)$

	American
Call	$\max\{0, S(t) - K\} \leq C_S(t,T,K) \leq S(t)$
Put	$\max\{0, K - S(t)\} \leq P_S(t,T,K) \leq K$

Table 7.1: Boundaries for the values of plain vanilla options.

the difference) the amount earned from the sale of the underlying at the spot price at maturity.

Because of the last relation in the second row in Table 7.1, a put can be worth no more than the strike price K (this occurs when the value of the underlying is zero). This fact will prove to be important when deciding whether to exercise an American option before maturity.

7.2.2 Early Exercise of American Options

An option can be characterized as a contract consisting of two rights.

- Firstly, the right to exercise the option. For American options, this right can be exercised over the entire term of the option, for European options, in contrast, only at maturity.
- Secondly, the right to let the option expire without exercising it. This right can be exercised for both American and European options, but only at maturity.

The value of the right to let the option expire is the difference between the price of the option and the corresponding forward position, since the transaction agreed upon in the forward contract *must* be performed at maturity. This right never has a negative value, a fact which is evident from the inequalities in Table 7.1. Denoting this right by $Y^C(t,T,K)$ and $Y^P(t,T,K)$ for an American call and put, respectively, we can write (using Equation 7.4):

$$C_S(t,T,K) = Y^C(t,T,K) + f_S(t,T,K) = Y^C(t,T,K) + \tilde{S}(t,T) - B_R(t,T)K$$
$$P_S(t,T,K) = Y^P(t,T,K) - f_S(t,T,K) = Y^P(t,T,K) - \tilde{S}(t,T) + B_R(t,T)K$$

The criterion for the early exercise of American options is that the proceeds gained by exercising be greater than the value of the option, i.e. $S(t) - K > C_S(t,T,K)$ and $K - S(t) > P_S(t,T,K)$ for a call and put, respectively. Setting the price of an option equal to a forward contract plus the right not to exercise and making use of the spot price adjusted for dividends as given in Equation 7.2, the early exercise condition for a call, for example, is

$$\begin{aligned} S(t) - K &> Y^C(t,T,K) + \tilde{S}(t,T) - B_R(t,T)K \\ &= Y^C(t,T,K) + \underbrace{S(t) - d(t,T)}_{\tilde{S}(t,T)} - B_R(t,T)K \end{aligned}$$

The procedure is analogous for the put. The spot price on the left and right-hand side of the inequality cancel each other and the condition for early exercise is expressed in terms of the dividend of the underlying:

$$\begin{aligned} d(t,T) &> K\left[1 - B_R(t,T)\right] + Y^C(t,T,K) \quad \text{for Calls} \\ d(t,T) &< K\left[1 - B_R(t,T)\right] - Y^P(t,T,K) \quad \text{for Puts} \end{aligned} \tag{7.6}$$

The strike price K, invested in risk-free securities from t until maturity T, earns an interest yield of $K\left[B_R(t,T)^{-1} - 1\right]$. The present value of this amount at time t is given by

$$B_R(t,T)K\left[B_R(t,T)^{-1} - 1\right] = K\left[1 - B_R(t,T)\right]$$

In view of these considerations, Equation 7.6 can be intuitively interpreted as follows :

- American puts should be exercised prior to maturity if the present value $d(t,T)$ of the dividend (which one forgoes upon early exercise) is less than the interest yield earned on the strike price (obtained when exercising early) less the value of the right not to exercise (which is lost when the option is exercised).
- American calls should only be exercised if the present value of the dividends (which are gained by exercising early) is greater than the interest yield earned on the strike price (which is lost upon early exercise) plus the value of the right not to exercise the call (which is also lost upon early exercise). Since both the interest yield on the strike price and the right to forgo exercising the option are both always greater than or equal to zero, **calls on underlyings which pay no dividends should never be exercised before maturity**.

7.2.3 Relationships between Puts and Calls

As an example of the effect an arbitrage-free market has on options, consider the following two portfolios:

- Portfolio structure
  - Portfolio **A** consists of a European call (long) with strike price K and maturity T, and a European put (short) also with strike price K and maturity T, both on the same underlying S.
  - Portfolio **B** consists of a forward contract (long) to buy the underlying upon maturity T at the delivery price K.

- Value of the portfolio at time t
  - Portfolio **A** has at time t a value of $c_S(t,T,K)$ $-p_S(t,T,K)$ from the long call and the short put.
  - Portfolio **B** has at time t a value of $f_S(t,T,K)$ from the long forward.
- Portfolio value at time T
  - Portfolio **A** has at time T a value of $S(T)-K$ from exercising the call if $S(T)>K$ or $-(K-S(T))$ from exercising the put if $S(T)<K$. In either case the portfolio is worth $S(T)-K$.
  - Portfolio **B** has at time T a value of $S(T)-K$ from the maturity of the forward.

If both portfolios have the same value at time T, they must be worth the same at earlier times t as well to prevent arbitrage opportunities. Setting the values of the two portfolios equal yields the famous *put-call parity* for European options:

Call minus put equals forward

One interpretation of this relation is: a European call is the same as a forward contract on the underlying plus a guarantee (the put) that the underlying is worth no less than the agreed upon delivery price at maturity. If, for example, an investor is certain that the underlying price will rise and wishes to make the most out of this information, he or she should buy forward contracts and not calls since by doing the (in the investor's opnion unnecessary) insurance premium guarding against a falling price is saved.

If the strike price K is equal to the current delivery price $S(t,T)$, the value of the forward equals zero. From the put-call parity we can immediately conclude that in this case the price of the put must equal the price of the call.

A put-call parity for *American* options can be derived using similar arbitrage arguments. The possibility of exercising early has the consequence that it can be expressed only in terms of an inequality. Using Equation 7.4 for the value of a forward contract, the parities can be expressed as follows:

$$\begin{array}{ll} \text{European} & c_S(t,T,K)-p_S(t,T,K)=\tilde{S}(t,T)-K\,B(r,t,T)=f_S(t,T,K) \\ \text{American} & \tilde{S}(t,T)-K\le C_S(t,T,K)-P_S(t,T,K)\le S(t)-K\,B(r,t,T) \end{array} \tag{7.7}$$

A further relation between European puts and calls is known as the *call-put symmetry*:

$$c_S(t,T,K)=\frac{K}{S(T)}p_S(t,T,\frac{S(T)^2}{K})$$

As the put-call parity, this relation is easily obtained by observing that the payments at maturity on the left and right-hand side of the above equation are the same. This symmetry can in theory be used, for example, to hedge exotic options (barriers and look-backs). Note however, that determining the strike price of the put and the number of puts required in the portfolio in the above equation requires that the future *spot* price $S(T)$ of the underlying at maturity (not the forward price $S(t,T)$!) must be known. This makes the call-put symmetry difficult to apply in practice.

8

The Black-Scholes Differential Equation

Having used arbitrage considerations to derive various properties of derivatives, in particular of option prices (upper and lower bounds, parities, etc.), we now demonstrate how such arbitrage arguments, with the help of results from stochastic analysis, namely Ito's formula 3.19, can be used to derive the famous Black-Scholes equation. Along with the Assumptions 1, 2, 3, 4 and 5 from Section 5, the additional assumption that continuous trading is possible is essential to establishing the equation, i.e. in the following we assume that Assumption 6 from Section 5 holds. The Black-Scholes equation is a partial differential equation which must be satisfied by *every* price function of path-independent European derivatives on a single underlying[1]. Consequently, one method of pricing derivatives consists in solving this differential equation satisfying the boundary conditions corresponding to the situation being investigated. In fact, even quite a number of path-*dependent* options obey this differential equation. A prominent example is the barrier option. In general however, the price of path-dependent options cannot be represented as a solution to the Black-Scholes equation. It is possible to surmount these difficulties by imbedding the state space in a higher dimensional space defining one or several additional variables in an appropriate manner to represent the different paths. This method is demonstrated explicitly by Wilmott for Asian options with arithmetic means [166]. As we will see below, the valuation of American options can also be accomplished via the Black-Scholes equation (with free boundary conditions).

The Black-Scholes equation can be derived in several different ways. The derivation presented here is the most intuitive and, being based on arbitrage arguments, is somewhat less technical (with the exception of Ito's lemma) since the arguments are largely economical in nature. In addition to presenting its derivation, we will establish

[1]Path independence is necessary for the derivative's value V to be only a function of t and the current underlying price S. Path-dependent options prices V are functions of further variables which themselves are dependent on the history of the underlying. The total differential of V is then no longer given by Equation 3.19; additional terms must be included in the expression.

its relationship to an equation called the diffusion equation or heat equation and then illustrate a mathematically deeper derivation of the Black-Scholes equation using the backward equation 3.29. Another approach to the Black-Scholes equation which does not involve an application of Ito's lemma will later be presented in Section 13.2 for the derivation of Equation 13.5. For the sake of consistency with the situation considered there, we will here investigate derivatives on underlyings with a dividend yield q.

8.1 The Black-Scholes Equation from Arbitrage Arguments

We denote the value at time t of a derivative on an underlying S by $V(S,t)$. Here, as is usually the case in this book, the letter V is used as the general notation for the price (*value*) of a derivative without further specifying the derivative. The underlying is governed by a stochastic process satisfying Equation 3.16. Without loss of generality, we write the mean return and the volatility in terms of relative (with respect to S) parameters $\widetilde{\mu}$ and σ, i.e.

$$\begin{aligned} a(S,t) &= S\,\widetilde{\mu}(S,t) \\ b(S,t) &= S\,\sigma(S,t) \end{aligned} \tag{8.1}$$

so that the stochastic process has the form

$$dS = \widetilde{\mu}(S,t)S\,dt + \sigma(S,t)S\,dW \tag{8.2}$$

The well-known process given by Equation 3.21 is a special case of the one introduced here where the price yield (drift) $\widetilde{\mu}$ and volatility σ are constant. The process under investigation here is just a reformulation of the Ito Process Equation 3.16 and is thus just as general as 3.16.

8.1.1 The Black-Scholes Equation for European Options

The value V of a derivative on this underlying is a function of a stochastic variable and is therefore also stochastic. In accordance with Ito's lemma 3.19, V satisfies the following equation in which the term containing dS represents the stochastic component of the change in the value of the derivative:

$$dV\,(S,t) = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2}\right)dt + \frac{\partial V}{\partial S}dS \tag{8.3}$$

A portfolio can now be constructed consisting of the derivative $V(S,t)$ and underlyings where the number of underlyings in the portfolio is denoted by Δ. We denote the value of this portfolio by Π

$$\Pi = V(S,t) + \Delta S \tag{8.4}$$

The change in the underlying during a small time interval of length dt (analogous to Equation 13.3) results from the stochastic change dS and the dividend yield accumu-

lated at a rate q during this time interval:

$$\begin{aligned} d(\Delta S) &= \Delta dS + \Delta S \left[B_q(t, t+dt)^{-1} - 1 \right] \\ &\approx \Delta dS + \Delta S \, q(t) dt \end{aligned}$$

where in the last step, an approximation which holds for all compounding methods over a sufficiently short time span dt, namely

$$B_q(t, t+dt)^{-1} - 1 \approx q(t)dt$$

has been used. The total differential of the entire portfolio Π is thus

$$d\Pi = dV + d(\Delta S) = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2 \frac{\partial^2 V}{\partial S^2} + \Delta S q(t) \right) dt + \left(\frac{\partial V}{\partial S} + \Delta \right) dS \quad (8.5)$$

Choosing the number of underlyings in the portfolio such that

$$\Delta \stackrel{!}{=} -\partial V / \partial S \quad (8.6)$$

has the effect that the coefficient of the stochastic term dS vanishes in the above expression for the change in the portfolio process. Thus, a portfolio can be constructed in such a way that the change in its value is generated solely by a progression dt in time. This is called a *delta hedge*. The total differential of the corresponding portfolio is now given by

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2 \frac{\partial^2 V}{\partial S^2} - q(t) S \frac{\partial V}{\partial S} \right) dt$$

This no longer contains a stochastic term. It follows that Π is risk-free and as such increases at the same rate as given by the risk free interest rate $r(t)$, otherwise arbitrage would be possible. Thus,

$$d\Pi \stackrel{!}{=} \left[B_r(t, t+dt)^{-1} - 1 \right] \Pi \, dt \approx r(t) \Pi \, dt = r(t) \left[V(S,t) - \frac{\partial V}{\partial S} S \right] dt$$

where the definition of the portfolio's value, Equation 8.4, is substituted in the last step. Equating the two expressions for $d\Pi$ immediately yields the partial differential equation which must be satisfied by $V(S,t)$:

$$\underbrace{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2 \frac{\partial^2 V}{\partial S^2} - q(t) S \frac{\partial V}{\partial S}}_{\text{Yield from a } \Delta\text{-Hedged Portfolio}} = \underbrace{r(t) \left[V - S \frac{\partial V}{\partial S} \right]}_{\text{Yield from a Bank Account}} \quad (8.7)$$

This is the famous *Black-Scholes differential equation*. From the arbitrage argument used in its derivation, we immediately see the intuitive meaning of this differential equation, namely that the yield (per time) earned from a delta-hedged portfolio must be equal to the yield earned from a risk-free investment (*bank account*) in the amount of the value of the portfolio. This interpretation is very useful in understanding and

working with the Black-Scholes equation. In the literature, the equation is usually written in the following form:

$$\frac{\partial V}{\partial t} + [r(t) - q(t)]\, S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2} - r(t)V = 0 \tag{8.8}$$

Note that no reference has been made to the nature of the derivative in the above derivation (with the exception of path independence and the European mode of exercise). An assumption about the underlying, however, has been made, namely Equation 8.2. Consequently, the *all* European, path-independent derivatives on underlyings given by the general stochastic process 8.2 satisfy the above differential equation. The coefficients r, q and σ are functions of t or S or of both.

In general, differential equations have an infinite number of quite different solutions. Thus the prices of quite different derivatives can satisfy the same differential equation. Introducing the additional requirement that *boundary conditions* be satisfied as well as the differential equation, reduces the number of solutions from an infinite number to exactly one. Seen from this perspective, different derivatives (on the same underlying) are distinguished only by requiring different boundary conditions to be satisfied. As the name suggests, boundary conditions are conditions which the solution of the differential equation must satisfy at the boundary of the set on which it solves the differential equation. This set is a subset of the solution's domain of definition. In our case, the arguments of the solution are the variables S and t and the associated set is a subset of the plane spanned by S and t. For example, for a plain vanilla call or put, the set lies between $S = 0$ and $S = \infty$ and between $t =$today and $t = T$.

The boundary condition in the time coordinate is called the *initial condition*. The initial condition for an option is that the value $V(S,T)$ of the derivative at maturity T must be given by its payoff P (which is a function of the price of the underlying)[2]:

$$V(S, t = T) = P(S) \tag{8.9}$$

The payoff profile $P(S)$ is established in the derivatives contract.

The boundary conditions in the S-coordinate are often obtained from asymptotic considerations (for example, the value of a plain vanilla call is zero at $S = 0$). For some exotic options, boundary conditions in the S-coordinate are explicitly specified in the option contract. For example, a knock-out option contract specifies that the value of the option is zero if S attains a barrier H, in other words when $S = H$.

8.1.2 The Black-Scholes *In*equality for American Options

American options differ from European options in that the holder has the right to exercise the option *before* maturity. This additional right, under certain circumstances, makes the American option more valuable than its European counterpart. We will now

[2]The fact that the "initial comdition" in our case belongs really to the *end* of the time period under consideration is just a question of vocabulary and should not confuse the reader. Alternatively one can introduce a new time variable $\widetilde{t} := T - t$ (sometimes called *term to maturity*). Then the initial condition belongs to the lowest value of $\widetilde{t}$, namely $\widetilde{t} = 0$.

show the variant of the Black-Scholes differential equation that American options must satisfy.

The change in the option's price is again given by the total differential equation 8.3. The stochastic component is contained in the dS-term only. We again construct a portfolio Π consisting of the derivative $V(S,t)$ and Δ underlyings:

$$\Pi = V(S,t) + \Delta S$$

The total differential of this portfolio is again of the form 8.5 and the stochastic component can, as before, be eliminated by choosing the number of underlyings in the portfolio as $\Delta = -\partial V/\partial S$. The portfolio thus constructed is risk-free. From arbitrage considerations, we conclude that a risk-free portfolio can yield *no more* than the risk-free interest rate. But it is certainly possible that it may earn less than the risk-free rate if the holder of the option chooses to exercise the option at a less than optimal time, in other words if the option is exercised when it should have been held or held when it should have been exercised (the holder of a European option cannot make this mistake as European options do not entail the right to early exercise). Consequently, a risk-free portfolio containing American options can at best earn a yield equal to the risk-free rate if the option is optimally exercised, otherwise less is earned. The relation for American options corresponding to the Black-Scholes equation.8.7 is thus

$$\underbrace{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2} - q(t)S\frac{\partial V}{\partial S}}_{\text{Yield from a } \Delta\text{-Hedged portfolio}} \leq \underbrace{r(t)\left[V - S\frac{\partial V}{\partial S}\right]}_{\text{Yield from a Bank Account}}$$

Or written in the form corresponding to Equation 8.8:

$$\frac{\partial V}{\partial t} + [r(t) - q(t)]\, S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2} - r(t)V \leq 0 \tag{8.10}$$

Since the option can be exercised at any time, it is always worth at least the amount given by its payoff profile. In addition to the initial condition, Equation 8.9 (which also holds for American options),

$$V(S, t = T) = P(S) \tag{8.11}$$

the following inequality must therefore always hold

$$V(S,t) \geq P(S) \qquad \forall t \tag{8.12}$$

Intuitively, the two inequalities 8.10 and 8.12 imply the two possible cases:

- Either the value of the option is, on a certain parameter set of S and t, greater than the payoff profile (i.e. Equation 8.12 is a strict *in*equality) in which case the right to early exercise is worthless. Thus the option behaves as a European option on this parameter set and therefore Equation 8.10 holds as an *equality.*
- On the set of S and t where the option's value is not greater than its payoff profile (i.e. Equation 8.12 holds as an equality), it would not be reasonable to hold the

option. In this case, the holder should exercise early. Therefore the "risk-free" portfolio including the option would earn less than the risk-free rate since the risk-free rate can only be obtained by exercising *optimally*. Thus, in this case Equation 8.10 holds as a strict *in*equality.

In summary: there exist two different parameter sets (regions in the (S,t)-plane). In one region we have

$$\frac{\partial V}{\partial t} + [r(t) - q(t)]\, S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2} - r(t)V = 0 \quad \text{and} \quad V(S,t) > P(S) \tag{8.13}$$

and in the other region we have

$$\frac{\partial V}{\partial t} + [r(t) - q(t)]\, S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 V}{\partial S^2} - r(t)V < 0 \quad \text{and} \quad V(S,t) = P(S) \tag{8.14}$$

The boundary between these parameter sets is a curve denoted by $S^*(t)$, called the *optimal exercise boundary* on which the option is exercised optimally. For values of $S(t)$ on one side of this curve, the option should be held, on the other side, the option should have been exercised earlier. Optimal exercise occurs directly on the curve. On the set where Equation 8.12 holds as an equality, the value of the option need not be computed since it is simply given by the payoff profile. It remains to calculate its value on the set where its behavior is European. We must therefore find a solution to Equation 8.10 as an *equality* on a parameter set whose boundary is given by the optimal exercise boundary $S^*(t)$ and which in addition satisfies the initial condition given by Equation 8.11. The difficulty is that the optimal exercise boundary is not known a priori. This type of problem is known as a *free* boundary condition[3]. For values on the boundary $S^*(t)$, the value of the option (as a solution of 8.13) must be equal to the payoff profile. If it were greater, the option holder would not exercise (the boundary would thus lie elsewhere), whereas a smaller value would yield an arbitrage opportunity (buying the option and exercising immediately). Thus, the option price must be continuous on the boundary:

$$V(S = S^*, t) = P(S = S^*)\ \forall t$$

The boundary curve $S^*(t)$ is positioned to give the option the greatest possible value at the boundary for all t (this is precisely the meaning of *optimal* exercise). For most options (but not, for example, digitals), this means that the first derivative of the option price with respect to S is continuous on the boundary:

$$\left.\frac{\partial V(S,t)}{\partial S}\right|_{S=S^*} = \left.\frac{\partial P(S)}{\partial S}\right|_{S=S^*}$$

These two conditions (the continuity of the option price function and its first derivative) in general enable the determination of the free boundary $S^*(t)$.

[3]In contrast, for European options, the Black-Scholes differential equation is always solved on the set $S = 0$ to $S = \infty$. The boundary is thus always fixed and known for European options.

8.1.3 A First Contact with the Risk-Neutral World

The mean return (drift) $\widetilde{\mu}$ of the underlying does *not* appear in the general differential equation 8.8 for European derivatives nor in the corresponding (in)equalities 8.13 and 8.14 for American derivatives. If a parameter does not appear in the differential equation nor the boundary condition, it is completely irrelevant for the problem under consideration. The drift of the underlying therefore has absolutely no influence on the value of a derivative on the underlying concerned! It was somehow "lost" in the derivation. This occurred exactly at that point where the delta hedge, Equation 8.6, was introduced into the equation to eliminate the stochastic component and as $\widetilde{\mu}$ was a part of the stochastic component dS, it was eliminated as well. The volatility σ also appears in the stochastic component dS. Contrary to $\widetilde{\mu}$, the volatility σ remains a parameter in the hedged portfolio since, as a consequence of the Ito lemma, it has entered into the deterministic component of Equation 8.3, i.e. into the dt-term.

Apart from this purely technical explanation, from an economical point of view it is at first glance quite extraordinary that a parameter as important as the mean return of an underlying should have absolutely no effect on the value of derivatives on that underlying. In fact, most investors expect a mean rate of return as compensation for accepting the risk of investing at all. And they expect the compensation to be higher the greater the risk. The fact that the mean return has no influence on the price of a derivative (no matter how large the risk – represented by the volatility σ – is!) can be interpreted in the following way: the valuation occurs in a world in which investors are indifferent to the risk involved in investing. This is referred to as *risk neutrality* or a *risk-neutral valuation*. The valuation of derivatives takes place in a risk-neutral world with objective parameters. This, however, is not the real world. For example, the price of a risky stock must have a higher mean return (a higher drift) than a stock (paying the same dividends) which is less risky, in return for the additional risk. Investors in the real world are not risk-neutral. Thus, derivatives must be considered in two worlds: the risk-neutral world for pricing and the real world for trading.

8.2 The Black-Scholes Equation and the Backward Equation

Having completed a derivation of the Black-Scholes equation providing the advantage of an intuitive interpretation[4] in accordance with Equation 8.7, which can be simply extended to apply to American options, we now proceed from the fundamental principles of stochastic analysis introduced in Section 3.4 to establish a connection between the equations derived there (in particular the backward equation) and the Black-Scholes equation and thus the pricing of derivatives. The transition probabilities $p(S', t' \,|S, t\,)$ for the stochastic processes satisfying Equation 3.16 will prove to be very useful for this purpose.

As previously mentioned, the transition probabilities $p(S', t' \,|S, t\,)$ contain all available information about the stochastic process. All pertinent values can be established

[4]The Black-Scholes equation simply means that the yield from a delta-hedged portfolio must equal the yield from a risk-free investment.

when this function is known. A value of particular interest is the *expectation* of a function $f(S', t')$ at a future date $t' = T$ based on information available today[5]. We denote the expectation by the letter E, thus

$$\begin{aligned} E(S,t,T) &= \mathrm{E}\left[f(S', t' = T)\,|\text{at time } t \text{ the process } S' \text{ is equal to } S\right] \\ &= \mathrm{E}\left[f(S', T)\,|S'(t) = S\right] \end{aligned}$$

This expectation can be expressed in terms of the transition probabilities of the stochastic process starting today from the point (S, t) and traveling to the future point $(S', t' = T)$ as

$$E(S,t,T) = \int_{-\infty}^{\infty} f(S', T) p(S', T\,|S, t)\, dS'$$

Taking the derivative of this expression with respect to t, taking note of the fact that f depends on the time T, but not on t, we obtain an equation for the change in this expectation with respect to a change in time

$$\frac{\partial E(S,t,T)}{\partial t} = \int_{-\infty}^{\infty} f(S', T) \frac{\partial p(S', T\,|S, t)}{\partial t} dS'$$

We now immediately see that the change in this expectation is completely governed by the time derivative of the transition probability p. Since we wish to establish values at an *earlier* time point proceeding backwards from later information, we substitute the *backward* equation 3.29 for the time derivative of p in the above integral. This allows us to express the $\partial p/\partial t$ in terms of derivatives with respect to S:

$$\begin{aligned} \frac{\partial E(S,t,T)}{\partial t} &= \int_{-\infty}^{\infty} f(S', T) \left[-\frac{1}{2} b(S,t)^2 \frac{\partial^2 p}{\partial S^2} - a(S,t) \frac{\partial p}{\partial S}\right] dS' \\ &= -\frac{1}{2} b(S,t)^2 \int_{-\infty}^{\infty} f(S', T) \frac{\partial^2 p}{\partial S^2} dS' - a(S,t) \int_{-\infty}^{\infty} f(S', T) \frac{\partial p}{\partial S} dS' \\ &= -\frac{1}{2} b(S,t)^2 \frac{\partial^2}{\partial S^2} \underbrace{\int_{-\infty}^{\infty} f(S', T) p\, dS'}_{E(S,t,T)} - a(S,t) \frac{\partial}{\partial S} \underbrace{\int_{-\infty}^{\infty} f(S', T) p\, dS'}_{E(S,t,T)} \end{aligned}$$

In the last step, we have made use of the fact that $f(S', T)$ depends on S', but not on S and that the derivatives with respect to S thus have no effect on f. We conclude that the expectation E itself satisfies the backward equation

$$\frac{\partial E}{\partial t} + \frac{1}{2} b(S,t)^2 \frac{\partial^2 E}{\partial S^2} + a(S,t) \frac{\partial E}{\partial S} = 0 \tag{8.15}$$

with a somewhat different boundary condition than p, namely

$$E(S, t = T, T) = f(S, T)$$

[5] In applications, this function $f(S, T)$ is often the payoff profile of a derivative.

Writing the parameters a and b as relative parameters as in Equation 8.1, we see that the differential equation for E is quite similar to the Black-Scholes equation

$$\frac{\partial E}{\partial t} + \frac{1}{2}\,\sigma(S,t)^2 S^2 \frac{\partial^2 E}{\partial S^2} + \widetilde{\mu}(S,t) S \frac{\partial E}{\partial S} = 0$$

As in Equation 8.8 the terms $\sim \partial/\partial t$, $\sim S\,\partial/\partial S$ and $\sim S^2\partial^2/\partial S^2$ appear. Only the term not involving derivatives is missing, i.e. that corresponding to "$r(t)V$" in Equation 8.8.

We consider therefore the *discounted* future expectation (denoted by V below) of an arbitrary function $f(S',T)$

$$\begin{aligned} V(S,t,T) &= B_r(t,T) E(S,t,T) \\ &= B_r(t,T) \int_{-\infty}^{\infty} f(S',T) p(S',T\,|S,t)\, dS' \end{aligned} \tag{8.16}$$

The derivative of this expression with respect to time t is

$$\begin{aligned} \frac{\partial V}{\partial t} = \frac{\partial B_r(t,T)}{\partial t} & \underbrace{\int_{-\infty}^{\infty} f(S',T) p(S',T\,|S,t)\, dS'}_{B_r(t,T)^{-1} V} \\ & + B_r(t,T) \int_{-\infty}^{\infty} f(S',T) \frac{\partial p(S',T\,|S,t)}{\partial t} dS' \end{aligned}$$

Using an argument analogous to the one presented above, the substitution of the backward equation for $\partial p/\partial t$ yields[6]

$$\frac{\partial V}{\partial t} + \frac{1}{2} b(S,t)^2 \frac{\partial^2 V}{\partial S^2} + a(S,t) \frac{\partial V}{\partial S} = V \frac{\partial}{\partial t} \ln\left(B_r(t,T)\right) \tag{8.17}$$

with the same boundary condition as for E, namely

$$V(S, t=T, T) = f(S,T)$$

This corresponds exactly to Equation 8.15 with an addition "source" term appearing as a consequence of discounting. Again writing the parameters a and b as relative parameters as in Equation 8.1, the equation takes the form

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2(S,t) S^2 \frac{\partial^2 V}{\partial S^2} + \widetilde{\mu}(S,t) S \frac{\partial V}{\partial S} = V \frac{\partial}{\partial t} \ln\left(B_r(t,T)\right) \tag{8.18}$$

For continuous discounting (and also for *all* other compounding conventions over an infinitesimal time interval dt) we have

$$\frac{\partial}{\partial t} \ln\left(B_r(t,T)\right) = r(t)$$

[6] We have also used the following general property of the logarithm:

$$\frac{\partial}{\partial t} \ln\left(B_r\right) = \frac{1}{B_r} \frac{\partial B_r}{\partial t}$$

8.2.1 A Second Contact with the Risk-Neutral World

Thus, the differential equation 8.18 for the (until now) abstract value of a "discounted expectation" – including the boundary condition – has the same structure as the Black-Scholes differential equation for pricing derivatives, 8.8. In fact, we only need to choose the parameter $\widetilde{\mu}$ so that $\widetilde{\mu}(S,t) = r(t) - q(t)$ and the differential equation for the discounted expectation is *identical* to the Black-Scholes equation! This then has the following interpretation: the price of a derivative with a payoff profile $f(S,T)$ is the same as the discounted expectation of this payoff profile *if and only if* the drift of the underlying is given by[7]

$$\begin{aligned} &\widetilde{\mu} = r - q \quad \text{or equivalently} \\ &\mu = r - q - \frac{\sigma^2}{2} \end{aligned} \tag{8.19}$$

Here, we also see that the actual drift (mean return) of the underlying has absolutely no effect on the *valuation* of the derivative. We again encounter *risk neutrality*, as in Section 8.1.3. The approach via the backward equation however, provides a much clearer interpretation of risk neutrality:

- The price of a derivative can be expressed as the discounted expectation of its payoff profile.
- To do so, the drift of the underlying must be precisely specified as in Equation 8.19. This is called the *risk-neutral drift*.

As will be seen later, choosing a drift corresponds to choosing a particular probability measure with respect to which the expectation is to be taken. The choice in Equation 8.19 is therefore referred to as the choice of a *risk-neutral probability measure*.

8.3 The Relationship to the Heat Equation

The Black-Scholes equation is a *linear, parabolic partial differential equation of second order*. The simplest parabolic partial differential equation (PPDE) of second order is the *heat equation* or *diffusion equation.* This equation has the simple form

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial \tau} \tag{8.20}$$

with x being the spacial variable and τ the time variable. It is one of the most thoroughly investigated equations in physics and applied mathematics. The Black-Scholes equation for constant yields (interest rates and dividends) and constant volatility can, as will be shown below, be transformed into the heat equation with appropriate changes in variables. The Black-Scholes equation and the heat equation have several interesting properties in common. For example, the time derivative in both equations is of first

[7] In the second equation, Equation 3.23 was used in the conversion to the drift $\mu = \widetilde{\mu} - \sigma^2/2$ of the stochastic process for the returns $\ln(S)$.

order. The solution to the equation thus has a time orientation. This corresponds to the property that heat flow is not reversible: if, for example, a point on an otherwise cold metal rod is heated, the passage of time leads to a uniform distribution of the heat over the entire rod. The singularity, represented by the application of the heat at a single point, is quickly smoothed out over the entire piece of metal. This property applies to the Black-Scholes equation as well. The time orientation of the Black-Scholes equation, however, is such that time flows from the future into the past. Non-smooth payoff functions (the payoffs of plain vanilla puts and calls have a "corner" at the strike price) become smooth immediately as time begins to move backwards from T. The analogy to the heat equation is, however, not only very useful for obtaining an intuitive understanding of the Black-Scholes equation. The biggest advantage of comparing the two is that the heat equation is one of the best-understood partial differential equations and many results holding for it may be extended to the Black-Scholes equation. We have the results of nearly two hundred years of research at our disposal. For example, the general solution to Equation 8.20 with initial condition

$$u(x, \tau = 0) = u_0(x)$$

is given by

$$u(x,\tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4\tau}\ u_0(y)\ dy \tag{8.21}$$

Thus, it is worthwhile to transform the Black-Scholes equation into the heat equation in order to make use of the results already known for the heat equation and extend them to Equation 8.8. We will therefore now explicitly present the changes of variables required for this transformation.

First, we write V in a more convenient form in order to eliminate the term not involving derivatives, i.e. to eliminate the rV in Equation 8.8:

$$V =: e^{-r(T-t)}u \Longrightarrow \frac{\partial V}{\partial t} = rV + e^{-r(T-t)}\frac{\partial u}{\partial t}$$
$$\frac{\partial V}{\partial S} = e^{-r(T-t)}\frac{\partial u}{\partial S}\ , \quad \frac{\partial^2 V}{\partial S^2} = e^{-r(T-t)}\frac{\partial^2 u}{\partial S^2}$$

This transforms the equation and boundary condition into

$$\frac{\partial u}{\partial t} + (r-q)S\frac{\partial u}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} = 0\ , \quad u(S, t = T) = P(S)$$

We now replace the time t by a dimensionless variable which measures the derivative's time to maturity. This also effects the reversal of the time orientation in the differential equation[8]:

$$z := \frac{\sigma^2}{2}(T-t) \Longrightarrow \quad \frac{\partial}{\partial t} = \frac{\partial z}{\partial t}\frac{\partial}{\partial z} = -\frac{\sigma^2}{2}\frac{\partial}{\partial z}$$

At this stage we have made use of the fact that the volatility σ does not depend explicitly on the time t. This transforms the differential equation and the initial condition into

[8] Note that by definition the volatility has the dimension $1/\sqrt{\text{time}}$, i.e. σ^2 has the dimension $1/\text{time}$, thus making the variable z dimensionless.

$$-\frac{\sigma^2}{2}\frac{\partial u}{\partial z} + (r-q)S\frac{\partial u}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} = 0 \ , \quad u(S, z=0) = P(S)$$

In order to eliminate the S dependence of the coefficients, we replace S with the logarithm of S. The argument of the logarithm must, however, be dimensionless. We thus divide by a constant K which has the same dimension as S. A natural choice for this constant would be the strike price of the option which appears in the payoff profile.

$$y := \ln(\frac{S}{K}) \Longrightarrow S = Ke^y \ , \quad S\frac{\partial}{\partial S} = S\frac{\partial y}{\partial S}\frac{\partial}{\partial y} = S\frac{K}{S}\frac{1}{K}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}$$
$$\frac{\partial^2}{\partial y^2} = S\frac{\partial}{\partial S}(S\frac{\partial}{\partial S}) = S\underbrace{\frac{\partial S}{\partial S}}_{1}\frac{\partial}{\partial S} + S^2\frac{\partial^2}{\partial S^2} \Longrightarrow S^2\frac{\partial^2}{\partial S^2} = \frac{\partial^2}{\partial y^2} - \frac{\partial}{\partial y}$$

The resulting differential equation and boundary condition is then given by

$$-\frac{\partial u}{\partial z} + \left[\frac{2(r-q)}{\sigma^2} - 1\right]\frac{\partial u}{\partial y} + \frac{\partial^2 u}{\partial y^2} = 0 \ , \quad u(y, z=0) = P(Ke^y)$$

The term involving the first derivative with respect to y can also be eliminated. For this, a new variable x is needed which depends on both y and z. The change in variable will involve both "old" variables. In consequence, *two* "new" variables x and τ need to be introduced. The transformation of the derivatives is accomplished using the *Jacobian determinant* which can be found in any introductory textbook on analysis:

$$x := y + (\frac{2(r-q)}{\sigma^2} - 1)z \ , \quad \tau := z \Longrightarrow$$
$$\frac{\partial}{\partial y} = \frac{\partial x}{\partial y}\frac{\partial}{\partial x} + \frac{\partial \tau}{\partial y}\frac{\partial}{\partial \tau} = 1\frac{\partial}{\partial x} + 0$$
$$\frac{\partial^2}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\right) = \frac{\partial}{\partial y}\frac{\partial}{\partial x} = \frac{\partial^2}{\partial x^2}$$
$$\frac{\partial}{\partial z} = \frac{\partial x}{\partial z}\frac{\partial}{\partial x} + \frac{\partial \tau}{\partial z}\frac{\partial}{\partial \tau} = (\frac{2(r-q)}{\sigma^2} - 1)\frac{\partial}{\partial x} + 1\frac{\partial}{\partial \tau}$$

Here we have made use of the fact that r, q and σ do not depend on y (and thus on S) nor on z (thus t). This completes the transformation of the differential equation and initial condition to

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial \tau} \ , \quad u_0(x) \equiv u(x, \tau=0) = P(Ke^x) \tag{8.22}$$

This is simply the heat equation. The strategy for calculating option prices is therefore to find a solution u of the heat equation in terms of the variables x and τ and then to

substitute the transformed variables by the original variables as follows:

$$\begin{aligned} \tau = z &= \frac{\sigma^2}{2}(T-t) \\ x &= y + (\frac{2(r-q)}{\sigma^2} - 1)z = \ln(\frac{S}{K}) + (r - q - \frac{\sigma^2}{2})(T-t) \\ u &= e^{r(T-t)}V \\ u_0(x) &= P(Ke^x, K) \end{aligned} \tag{8.23}$$

thereby expressing the solution in terms of the original financial variables.

9

Integral Forms and Analytic Solutions in the Black-Scholes World

In addition to Assumptions 1, 2, 3, 4, 5 and 6 from Section 5 required to set up the differential equation in Section 8, we will now further simplify our model by assuming that the parameters involved (interest rates, dividend yields, volatility) are constant (Assumptions 9, 11 and thus 7 from Section 5) despite the fact that these assumptions are quite unrealistic. These were the assumptions for which Fischer Black and Myron Scholes first found an analytic expression for the price of a plain vanilla option, the famous Black-Scholes option pricing formula. For this reason, we often speak of the *Black-Scholes world* when working with these assumptions. In the Black-Scholes world, solutions of the Black-Scholes differential equation (i.e. option prices) for some payoff profiles (for example for plain vanilla calls and puts) can be given in closed form. We will now present two elegant methods to derive such closed form solutions.

9.1 Option Prices as Solutions of the Heat Equation

The first and perhaps most natural approach would be to take advantage of the constant parameter assumptions to transform the Black-Scholes equation into the heat equation 8.22 as presented above. Since the solution to the heat equation is known and given by Equation 8.21 we simply need to write the initial condition corresponding to the desired financial instrument in terms of the variables x and τ, and, in accordance with Equation 8.23, transform the solution u back in terms of the original financial variables. We now demonstrate this technique using a plain vanilla call as an example.

Expressing the payoff profile of the call in the variables of the heat equation gives

$$\begin{aligned} P(S) &= \max(S-K,0) \\ &\Longrightarrow \\ u_0(x) &= P(Ke^x) = \max(Ke^x - K, 0) = K\max(e^x - 1, 0) \end{aligned}$$

Substituting this into Equation 8.21 immediately yields the solution in integral form for this initial condition

$$u(x,\tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4\tau}\ u_0(y)\ dy \tag{9.1}$$

This *integral form* is not only valid for plain vanilla calls but for arbitrary payoff profiles and the resulting initial conditions u_0. The integral can be computed numerically using, for example, the Monte Carlo method. Decades of research on numerical methods for computing integrals can be taken advantage of here.

In the case of the plain vanilla call, however, it is in fact possible to obtain a closed *analytical* form of the solution of the above integral. Substituting the initial condition for the call into the above equation yields

$$\begin{aligned} u(x,\tau = \frac{K}{2\sqrt{\pi\tau}} &\int_{-\infty}^{\infty} e^{-(x-y)^2/4\tau} \max(e^y - 1, 0)\ dy \\ = \frac{K}{2\sqrt{\pi\tau}} &\int_{0}^{\infty} e^{-(x-y)^2/4\tau}(e^y - 1)\ dy \\ = \frac{K}{2\sqrt{\pi\tau}} &\int_{0}^{\infty} e^{y-(x-y)^2/4\tau}\ dy - \frac{K}{2\sqrt{\pi\tau}} \int_{0}^{\infty} e^{-(x-y)^2/4\tau}\ dy \end{aligned}$$

The second of the two integrals above can be computed after making the change in variable $\widetilde{y} \equiv (x-y)/\sqrt{2\tau}$ implying $dy = -\sqrt{2\tau}d\widetilde{y}$. The integral bounds must then be transformed as follows: $0 \to x/\sqrt{2\tau}$ and $\infty \to -\infty$. The integral has now become an integral over the density function of the standard normal distribution:

$$K\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\sqrt{2\tau}} e^{-\widetilde{y}^2/2}\ d\widetilde{y} = K\ \mathrm{N}\left(\frac{x}{\sqrt{2\tau}}\right)$$

The first of the above two integrals can be calculated by completing the squares as follows

$$y - \frac{(x-y)^2}{4\tau} = \tau + x - \frac{(y-x-2\tau)^2}{4\tau}$$

This transforms the integral into

$$\frac{K}{\sqrt{2\pi}\sqrt{2\tau}} e^{\tau+x} \int_{0}^{\infty} e^{-\frac{(y-x-2\tau)^2}{2\sqrt{2\tau}^2}}\ dy$$

The remaining integral can now again be expressed in terms of the standard normal distribution. The necessary change in variable is $\widetilde{y} \equiv -(y-x-2\tau)/\sqrt{2\tau}$. Combining

all these results the solution becomes

$$u(x,\tau) = K\ e^{x+\tau}\ \mathrm{N}\left(\frac{x}{\sqrt{2\tau}} + \sqrt{2\tau}\right) - K\ \mathrm{N}\left(\frac{x}{\sqrt{2\tau}}\right)$$

where $\mathrm{N}(x)$ – as always – denotes the cumulative standard normal distribution, Equation A.48. Substituting now for the original variables using Equation 8.23 gives

$$\begin{aligned} e^{r(T-t)}V(S,t) = K\ & e^{\ln(\frac{S}{K})+(r-q)(T-t)} \\ & \times \mathrm{N}\left(\frac{\ln(\frac{S}{K}) + (r-q-\frac{\sigma^2}{2})(T-t)}{\sqrt{\sigma^2(T-t)}} + \sqrt{\sigma^2(T-t)}\right) \\ & - K\ \mathrm{N}\left(\frac{\ln(\frac{S}{K}) + (r-q-\frac{\sigma^2}{2})(T-t)}{\sqrt{\sigma^2(T-t)}}\right) \end{aligned}$$

After multiplying by $e^{-r(T-t)}$, we finally obtain the famous Black-Scholes formula for the price of a European call:

$$V(S,t) = e^{-q(T-t)}S\,\mathrm{N}\,(x) - e^{-r(T-t)}K\ \mathrm{N}x - \sigma\sqrt{(T-t)}) \tag{9.2}$$

where

$$x \equiv \frac{\ln(\frac{S}{K}) + (r-q)(T-t)}{\sqrt{\sigma^2(T-t)}} + \frac{1}{2}\sigma\sqrt{(T-t)}$$

9.2 Option Prices and Transition Probabilities

We will now show how the foundations of stochastic analysis laid in Section 3.4 can be used to price options. In Section 8.2, we have seen that with the risk-neutral choice of drift, the prices of derivatives are given by the discounted expectation of the payoff profile, Equation 8.16. This expectation is determined using the transition probabilities $p(S',t'\,|S,t\,)$. *If* these are known, calculating the price of the option reduces to simply calculating the integral. In the Black-Scholes world, i.e. for the simple process 3.20, the transition probabilities are given explicitly by Equation 3.32 with $\widetilde{\mu} = \mu + \sigma^2/2$. Thus, Equation 8.16 becomes the integral form for the price of an arbitrary derivative with an associated payoff profile $f(S,T)$:

$$\begin{aligned} V(S,t,T) &= B_r(t,T)\int_{-\infty}^{\infty} f(S',T)p(S',T\,|S,t\,)dS' \\ &= \frac{B_r(t,T)}{\sqrt{2\pi\sigma^2(T-t)}}\int_{-\infty}^{\infty} f(S',T)\exp\left\{\frac{-\left[\ln(S'/S) - \mu(T-t)\right]^2}{2\sigma^2(T-t)}\right\}\frac{dS'}{S'} \end{aligned} \tag{9.3}$$

This integral can be computed numerically for arbitrary payoff profiles $f(S,T)$ and is equal to the price of the derivative for the risk-neutral choice of the drift as specified in Equation 8.19.

For some special payoff profiles, the integral can even be solved analytically, or reduced to an expression in terms of known functions. We demonstrate this using the concrete example of a plain vanilla call option with payoff profile $f(S',T) = \max(S'-K,0)$. For this payoff profile, the integral can be written as

$$V(S,t,T) = \frac{B_r(t,T)}{\sqrt{2\pi\sigma^2(T-t)}} \int_K^\infty (S'-K) \exp\left\{ \frac{-\left[\ln(S'/S) - \mu(T-t)\right]^2}{2\sigma^2(T-t)} \right\} \frac{dS'}{S'}$$

The substitution $u := \ln(S'/S)$ simplifies the integral to

$$\begin{aligned} V(S,t,T) &= \frac{B_r(t,T)}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln(K/S)}^\infty (Se^u - K) \exp\left\{ \frac{-\left[u - \mu(T-t)\right]^2}{2\sigma^2(T-t)} \right\} du \qquad (9.4) \\ &=: B_r(t,T) S I_1 - B_r(t,T) K I_2 \end{aligned}$$

Both integrals I_1 and I_2 can be easily calculated. In the first integral we complete the square in the argument of the exp-function:

$$\begin{aligned} u - \frac{\left[u-\mu(T-t)\right]^2}{2\sigma^2(T-t)} &= -\frac{u^2 - 2\mu(T-t)u + \mu^2(T-t)^2 - 2\sigma^2(T-t)u}{2\sigma^2(T-t)} \\ &= -\frac{\left[u - \left(\mu+\sigma^2\right)(T-t)\right]^2 - \left(\mu+\sigma^2\right)^2 (T-t)^2 + \mu^2(T-t)^2}{2\sigma^2(T-t)} \\ &= -\frac{\left[u - \left(\mu+\sigma^2\right)(T-t)\right]^2 - \left(\sigma^4 + 2\mu\sigma^2\right)(T-t)^2}{2\sigma^2(T-t)} \\ &= \frac{-\left[u - \left(\mu+\sigma^2\right)(T-t)\right]^2}{2\sigma^2(T-t)} + \left(\mu + \frac{\sigma^2}{2}\right)(T-t) \end{aligned}$$

Thus, the first integral becomes

$$\begin{aligned} I_1 &\equiv \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln(K/S)}^\infty \exp\left\{ u - \frac{\left[u-\mu(T-t)\right]^2}{2\sigma^2(T-t)} \right\} du \\ &= \frac{e^{\left(\mu+\sigma^2/2\right)(T-t)}}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln(K/S)}^\infty \exp\left\{ \frac{-\left[u - \left(\mu+\sigma^2\right)(T-t)\right]^2}{2\sigma^2(T-t)} \right\} du \end{aligned}$$

With the substitution

$$\begin{aligned} y &:= -\frac{u - \left(\mu+\sigma^2\right)(T-t)}{\sqrt{\sigma^2(T-t)}} \Rightarrow \\ \frac{dy}{du} &= -1/\sqrt{\sigma^2(T-t)} \Rightarrow du = -\sqrt{\sigma^2(T-t)}\,dy \end{aligned}$$

the upper and lower limits of integration become

$$\begin{aligned} y_{\text{upper}} &= -\frac{\infty - \left(\mu+\sigma^2\right)(T-t)}{\sqrt{\sigma^2(T-t)}} = -\infty \\ y_{\text{lower}} &= -\frac{\ln(K/S) - \left(\mu+\sigma^2\right)(T-t)}{\sqrt{\sigma^2(T-t)}} = \frac{\ln(S/K) + \left(\mu+\sigma^2\right)(T-t)}{\sqrt{\sigma^2(T-t)}} \end{aligned}$$

Exchanging the upper and lower limits results in a change in the sign of the integral. This is compensated for by the sign of du. Combining the above results, I_1 becomes after this substitution

$$\begin{aligned} I_1 &= e^{\left(\mu+\sigma^2/2\right)(T-t)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(S/K)+(\mu+\sigma^2)(T-t)}{\sqrt{\sigma^2(T-t)}}} \exp\left\{\frac{-y^2}{2}\right\} dy \\ &= e^{\left(\mu+\sigma^2/2\right)(T-t)} \mathrm{N}\left(\frac{\ln(S/K)+\left(\mu+\sigma^2\right)(T-t)}{\sqrt{\sigma^2(T-t)}}\right) \end{aligned}$$

where, as usual, N denotes the cumulative standard normal distribution.

The second integral can be computed after making the substitution $y := -(u - \mu(T-t))/\sqrt{\sigma^2(T-t)}$:

$$\begin{aligned} I_2 &\equiv \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln(K/S)}^{\infty} \exp\left\{\frac{1}{2}\frac{-\left[u-\mu(T-t)\right]^2}{\sigma^2(T-t)}\right\} du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(S/K)+\mu(T-t)}{\sqrt{\sigma^2(T-t)}}} \exp\left\{-\frac{y^2}{2}\right\} dy \\ &= \mathrm{N}\left(\frac{\ln(S/K)+\mu(T-t)}{\sqrt{\sigma^2(T-t)}}\right) \end{aligned}$$

The generalization of Equation 8.19 for the risk-neutral choice of drift in arbitrary compounding methods is (see Equation 11.25)

$$\mu(T-t) := \ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right) - \frac{\sigma^2}{2}(T-t)$$

This simplifies the integrals further to

$$I_1 = \frac{B_q(t,T)}{B_r(t,T)} \mathrm{N}\left(x\right) \quad , \qquad I_2 = \mathrm{N}\left(x - \sigma\sqrt{T-t}\right)$$

where x is, as usual, given by

$$x = \frac{\ln(\frac{B_q(t,T)S}{B_r(t,T)K}) + \frac{1}{2}\sigma^2(T-t)}{\sqrt{\sigma^2(T-t)}} \tag{9.5}$$

Collecting these results, we obtain the price of a plain vanilla call as

$$\begin{aligned} V(S,t,T) &= B_r(t,T)SI_1 - B_r(t,T)KI_2 \\ &= B_q(t,T)S(t)\,\mathrm{N}\left(x\right) - B_r(t,T)K\,\mathrm{N}\left(x-\sigma\sqrt{T-t}\right) \end{aligned} \tag{9.6}$$

Again, this is the famous *Black-Scholes option pricing formula* and corresponds exactly to Equation 9.2 for continuous compounding.

9.3 Compilation of Black-Scholes Option Prices for Different Underlyings

9.3.1 Options on the Spot Price

The above derivation holds for a call on an underlying with a dividend *yield.* The Black-Scholes equations, however, continue to hold when the underlying pays dividends in discrete time intervals. In all cases, we can proceed as if the underlying were not paying any dividend and had a spot price is given by the dividend-adjusted spot price in Equation 3.5. The payoff profile and the Black-Scholes value are summarized here for puts and calls. The payoff profiles at time T are:

$$\begin{aligned} c_S(T,T,K) &= \max\{0, S(T) - K\} \\ p_S(T,T,K) &= \max\{0, K - S(T)\} \end{aligned}$$

The Black-Scholes option prices at time t are:

$$\begin{aligned} c_S(t,T,K) &= \quad \tilde{S}(t,T)\mathrm{N}(\ \ x) - K\,B_r(t,T)\mathrm{N}(\ \ x - \sigma\sqrt{T-t}) \\ p_S(t,T,K) &= -\tilde{S}(t,T)\mathrm{N}(-x) + K\,B_r(t,T)\mathrm{N}(-x + \sigma\sqrt{T-t}) \end{aligned} \tag{9.7}$$

where

$$x = \frac{\ln\left(\frac{\tilde{S}(t,T)}{K\,B_r(t,T)}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t} = \frac{\ln\left(\frac{S(t,T)}{K}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}$$

The application of this famous *Black-Scholes option pricing formula* to option portfolios is demonstrated in detail in the Excel workbook BlackScholesModel.xls enclosed. In this workbook, the derivatives of the option price with respect to its parameters, called the *Greeks*, are also computed. The workbook can be used as a complete option calculator (the fields colored yellow are the input fields).

9.3.2 Options on the Forward Price

The underlying is now not the spot price $S(t)$ but the forward price $S(t,T')$ for a time $T' \geq T$, where T is the maturity date of the option. Options on the forward price refer to either futures or forwards. In both cases, exercising the option results in payment of the value of the underlying forward contract and – in addition – the exerciser of the option goes long (in the case of a call) or short (in the case of a put) in the forward contract concerned[1]. Since the value of a futures and forward position according to Equations 7.5 and 7.4, respectively, are different, the payoff profile and thus the value of options on these contracts are different as well.

[1] But of course with the then valid forward price as the delivery price so that the forward contract – as always – has zero value when entered into.

Options on Futures

Upon maturity at time T of the option, the value of the future with a maturity date $T' \geq T$ is paid if this value is positive. The payoff profiles are

$$c_F(T,T,K) = \max\{0, F_S(T,T',K)\} = \max\{0, S(T,T') - K\}$$
$$p_F(T,T,K) = \max\{0, -F_S(T,T',K)\} = \max\{0, K - S(T,T')\}$$

A method often used to find the Black-Scholes price is to transform the payoff profile into a payoff profile of a known option. For this reason, we write the payoff profile of the call as

$$c_F(T,T,K) = \frac{B_q(T,T')}{B_r(T,T')} \max\left\{0, S(T) - \frac{B_r(T,T')}{B_q(T,T')}K\right\}$$

where we have used Equation 7.1 for the forward price at time T (for the case of a dividend yield q). Thus, the price of a call on a future with strike price K can be written as the price of B_q/B_r calls on the spot price with strike price KB_r/B_q. The argument for the put is completely analogous. A call on a *future* with strike price K thus has the same payoff profile as B_q/B_r calls on the *spot* with strike price KB_r/B_q. Thanks to of Equation 9.7, the price of an option on the spot price is known. Substituting KB_r/B_q for the strike price and using Equation 3.4 for B_r and B_q yields:

$$\begin{aligned} c_S(t,T,K) &= \frac{B_q(T,T')}{B_r(T,T')} \left[B_q(t,T)S(t)\mathrm{N}(x')\right. \\ &\quad - \left. B_r(t,T)K\frac{B_r(T,T')}{B_q(T,T')}\mathrm{N}(x' - \sigma\sqrt{T-t})\right] \\ &= \frac{B_q(t,T')}{B_r(T,T')}S(t)\mathrm{N}(x') - B_r(t,T)K\mathrm{N}(x' - \sigma\sqrt{T-t}) \end{aligned}$$

x' corresponds to the x in Equation 9.7 with the modified strike price:

$$\begin{aligned} x' &= \frac{\ln\left(\frac{B_q(t,T)S(t)}{B_r(t,T)[K\,B_r(T,T')/B_q(T,T')]}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t} \\ &= \frac{\ln\left(\frac{B_q(t,T')S(t)}{B_r(t,T')K}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t} \end{aligned}$$

Now, again with the help of Equation 7.1, the spot price $S(t)$ is written in terms of the actual underlying, namely the forward price. Using the expression for B_r in Equation 3.4 finally gives the Black-Scholes price for options on futures:

$$\begin{aligned} c_F(t,T,K) &= B_r(t,T)\left[S(t,T')\mathrm{N}(x') - K\,\mathrm{N}(x' - \sigma\sqrt{T-t})\right] \\ p_F(t,T,K) &= B_r(t,T)\left[-S(t,T')\mathrm{N}(-x') + K\,\mathrm{N}(-x' + \sigma\sqrt{T-t})\right] \end{aligned} \tag{9.8}$$

where

$$x' = \frac{\ln\left(\frac{S(t,T')}{K}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}$$

Comparing this with the corresponding prices for options on the spot prices, Equation. 9.7, for the special case of an underlying whose dividend yield is exactly equal to the risk-free interest rate, in other words where (from Equation 3.5) $\tilde{S}(t,T) = B_r(t,T)S(t)$, we obtain the following "cookbook" recipe:

Plain vanilla option on futures can be priced like options on the spot price of an (artificial) underlying whose spot price is equal to $S(t,T')$ and whose dividend yield is equal to the risk free rate r.

If the future matures at the same date as the option, i.e. if $T' = T$, then there is (because of Equation 7.1) no difference in either the payoff profile or the price of the option on the future and the option on the spot price. In this case Equation 18.4.2 (i.e. pricing options using the forward price, even if it is a option on the spot price) is referred to as the *Black-76 model*.

In summary, if either $T = T'$ or $q = r$, there is no difference in the option on a future and the option on a spot.

Options on Forwards

On the maturity date T of the option, the value of the forward maturing on $T' \geq T$ will be paid to the holder of a call if this value is positive. This is different from the value of the future since in the case of a forward, the value is discounted from maturity T' back to T (see Equation. 7.4). The payoff profiles are thus:

$$c_f(T,T,K) = \max\{0, f_S(T,T',K)\} = B_r(T,T')\max\{0, S(T,T') - K\}$$
$$p_f(T,T,K) = \max\{0, -f_S(T,T',K)\} = B_r(T,T')\max\{0, K - S(T,T')\}$$

Comparing this with the payoff profiles for options on futures shows that an option on a forward corresponds to B_r options on the future. Therefore the Black-Scholes prices can be immediately obtained from Equation 18.4.2

$$c_f(t,T,K) = B_r(t,T')\left[S(t,T')\mathrm{N}(x') - K\,\mathrm{N}(x' - \sigma\sqrt{T-t})\right]$$
$$p_f(t,T,K) = B_r(t,T')\left[-S(t,T')\mathrm{N}(-x') + K\,\mathrm{N}(-x' + \sigma\sqrt{T-t})\right] \qquad (9.9)$$

with $x\prime$ defined as in Equation 18.4.2. The difference between this and Equation 18.4.2 is that for options on *forwards* the discounting is done from the maturity of the *forward* contract T', whereas for options on *futures* the discounting is done from the maturity of the *option* T.

If the forward matures on the same date as the option, $T' = T$, there is no difference in the payoff profile or in the Black-Scholes price between an option on the forward and an option on the spot price. Only in this case are the prices of an option on the spot, an option on a future and an option on a forward all equal.

9.3.3 Options on Interest Rates

Forward Volatilities

In the derivation of the Black-Scholes equation for options on the forward price, it has been assumed that volatility remained constant throughout. Although only the

forward price appears in Equations 18.4.2 and 9.9, the volatility refers strictly speaking to the volatility of the spot price. Despite this fact, the volatility of the forward price is generally used in the market, this model being commonly referred to as the *Black-76 model*. In the context of the assumptions made for the Black-Scholes equation, in particular that the volatility is assumed to be constant, this, theoretically, makes no difference in the valuation.

The Black-76 model is commonly used when the underlying S is an interest rate or an interest rate instrument (like a bond, for instance). It can be shown that the Black-76 model holds even when the Black-Scholes assumptions are weakened somewhat. The underlying process must not necessarily be a random walk with constant volatility. It is sufficient that the logarithm of the underlying $S(T)$ at option maturity is normally distributed. The variance of the distribution of $\ln(S(T))$ will be written as

$$\text{var}\left[\ln S(T)\right] = \sigma(T)^2 T$$

The parameter $\sigma(T)$ is called the *forward volatility*. It is the volatility of the underlying price $S(T)$ at maturity T. Because the Black-76 model "lives" in the Black-Scholes world, interest rates are assumed to be non-stochastic. Therefore the forward price $S(t,T)$ and the future price $S(T)$ are equal if the underlying S is an interest rate or an interest rate instrument[2]. At an earlier time $t < T$ we can thus use the forward price $S(t,T)$ for $S(T)$ and the current volatility *of this forward price* for the forward volatility. This argument justifies the use of the Black-76 model with the volatility of the forward price when pricing interest rate options in the Black-Scholes world.

Normal versus Lognormal Distribution for Changes in Interest Rates

The Black-76 model is applied to options on interest rate instruments such as bonds as well as to options which depend directly on interest rates, such as caps and floors. For $S(t,T)$ we take either the forward price of the underlying instrument (for example, of the bond) or the forward rate of the reference interest rate (3-month LIBOR rate in 6 months, for example). In both cases the application of the Black-76 model implicitly assumes that each underlying (bond price or interest rate) is lognormally distributed. This leads to a contradiction since interest rates are defined to be a bond's yield. Yields, however, are the logarithm of relative price changes. If in the case of a *bond* option, it is assumed that these logarithms are normally distributed, the interest rates (=yields) cannot simultaneously be *log*normally distributed. This is, however, done in practice: the market prices bond options under the assumption that bond prices are lognormally distributed, i.e. that interest rates are normally distributed. On the other hand, caps, floors and collars are priced under the assumption that interest rates are lognormally distributed (similarly, swaptions are priced under the assumption of lognormal swap rates). These two points of view can be compensated for by taking different volatilities (price volatilities and interest rate volatilities, respectively) as input parameters. As will be shown in Section 28.3.3, price volatility and yield volatility are related to one another through the modified duration, see Equation 28.18.

[2]This will be shown explicitly in Section 15.3.

The assumption of normally distributed changes in interest rates is consistent with the pricing of bond options but conceals an additional problem: negative interest rates could arise. We thus have the choice between two less than optimal solutions.

- Assuming that changes in interest rates are normally distributed allows the pricing of bond options using the Black-Scholes model but admits the possibility of negative interest rates.
- Assuming that changes in interest rates are lognormally distributed makes it impossible for interest rates to take on negative values but bond options cannot be priced using the Black-Scholes model.

Traders choose the lesser of the two evils depending on the market in which they are trading. In the cap market for example, the disadvantage of not being able to price bond options is irrelevant. Therefore, the trader assumes that the interest rates are lognormally distributed (as long as he or she is comfortable with the assumptions of the Black-Scholes world in the first place). For the bond and swap markets, on the other hand, an assumption making the valuation of a bond option impossible is completely unacceptable. In this case, it is assumed that interest rates are normally distributed (as long as the trader is comfortable with the assumptions of the Black-Scholes world in the first place).

Black-76 Model for Interest Rate Options

With the above interpretations of the input parameters and under the assumptions described above, Equation 18.4.2 yields the Black-76 model for interest rate options. Explicitly:

$$\begin{aligned} c(t,T,K) &= B_r(t,T)\left[\quad S(t,T)\mathrm{N}(x) - K\,\mathrm{N}(\quad x - \sigma\sqrt{T-t})\right] \\ p(t,T,K) &= B_r(t,T)\left[-S(t,T)\mathrm{N}(-x) + K\,\mathrm{N}(-x + \sigma\sqrt{T-t})\right] \end{aligned} \tag{9.10}$$

where

$$x = \frac{\ln\left(\frac{S(t,T)}{K}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}$$

Here, $S(t,T)$ is either the forward rate of the underlying interest rate or forward price of an underlying interest rate *instrument* (like a bond, for example). This equation forms the basis for pricing interest rate option *in the Black-Scholes world*. Formally, the difference between this and Equation 18.4.2 is that the forward price with respect to the option's maturity is used, i.e. $T = T'$. As mentioned after Equation 18.4.2, there is no difference in this case between an option on a forward price and an option on a spot price. We could just as well work with Equation 9.7. The only subtlety involved is that the *forward volatility* or the volatility of the forward rate should be used.

10

Numerical Solutions Using Finite Differences

Among the many numerical procedures available to solve partial differential equations the *finite difference method* stands out as probably the most widely used method in mathematical finance. Finite difference methods are very powerful and flexible as well. They can be applied to a wide variety of different derivatives with either European or American payoff modes. In this section, we will provide a very detailed discussion of these important methods in a generality far exceeding that which is usually presented in comparable books.

Finite difference methods approximate the partial derivatives appearing in partial differential equations like the Black-Scholes equation 8.8 using finite difference quotients. The equation is then solved on a *grid* spanned by the linearly independent variables (for example, time t and price S of the underlying) appearing in the PDE[1]. Doing so, we obtain a solution surface which represents the price on each of the grid points (S, t). In general, a PDE has an unbounded number of solutions. Usually we are only interested in a solution which satisfies specific boundary and/or initial conditions. The finite difference method requires the specification of both boundary and initial conditions.

Since finite differences can be applied in quite general settings, we will require only the assumptions needed for arbitrage free trading, i.e. Assumptions 1, 2, 3, 4, 5 from Section 5, with the additional assumption 6. Assumption 6 ensures that the variables in the problem are continuous. This is necessary if we want to obtain differentiable solutions. In order to provide a manageable overview of finite differences, we will restrict the treatment given here to random walks (Assumption 7) with non-stochastic interest rates and volatilities (Assumptions 8 and 10, respectively).

[1] In the following we will often use the abbreviation *PDE* for "partial differential equation", as is common practice in the related literature.

10.1 Discretizing the Black-Scholes Equation

Below, the Black-Scholes Differential Equation 8.8 will be solved numerically with the help of finite difference methods. In the literature, Equation 8.8 is often transformed into an equation based on a new variable given by $Z = \ln(S)$. Such an equation has the advantage that the coefficients no longer depend explicitly on S. On occasion, it is claimed that a further advantage of this change of variable is that a uniform grid in Z is numerically more efficient than a uniform grid in S. However, the differences are usually negligible and this argument does not hold for barrier options. Furthermore, it is often preferable to use a non-uniform grid. We could distribute the grid points logarithmically, for example so that the non-uniform grid in S corresponds to a uniform grid in Z. In view of these considerations, we will continue to use the Black-Scholes equation in the form given by 8.8. But we will present the finite difference method for general *non-uniform grids*.

The finite difference method now consists in determining the value V of a financial instrument on a grid with coordinates S and t by approximating the partial derivatives with *finite* differences. We will restrict the discussion here to a rectangular grid allowing, however, the distance between grid points to be non-uniform. Such a grid is completely determined by the grid points in the S and t directions denoted by:

$$\begin{aligned} &S_i \; ; \; i = 0, 1, 2, \ldots, M \\ &t_j \; ; \; j = 0, 1, 2, \ldots, N \end{aligned} \tag{10.1}$$

We introduce the notation $W_{i,j} = W(S_i, t_j)$ for the solution's approximation in order to distinguish it from the exact solution $V(S_i, t_j)$ evaluated at points on the grid.

For the sake of clarity we will at several instances start with a grid whose time steps as well as the steps in the direction of the underlying price are uniformly spaced. For such *uniform grids* the spacing between grid points is simply

$$\begin{aligned} S_i - S_{i-1} = \delta S \; \forall i \implies S_i = S_0 + i\,\delta S \\ t_j - t_{j-1} = \delta t \; \forall j \implies t_i = t_0 + i\,\delta t \end{aligned} \tag{10.2}$$

We will then, however, always generalize our discussion to non-uniform grids.

The fundamental idea behind the method presented here is to determine the value of the derivative from the values at neighboring time points. Since the value of the derivative (as a function of the underlying's price) is usually known at maturity (it is given a priori by the payoff profile $P(S)$ as a function of S), we proceed using the strategy of starting at maturity $t_N = T$ and calculating *backwards* to time t_{N-1}, from there calculating back to t_{N-2} and so on. To accomplish this, we express the value of the instrument in terms of its Taylor series with respect to time, and express the time derivatives appearing in the Taylor expansion in terms of the derivatives with respect to the underlying price by using the Black-Scholes PDE.

10.1.1 The Explicit Method

In the explicit method, the Taylor series expansion is used to calculate values at an *earlier* time $t - \delta t$ from the values at time t:

$$V(S, t - \delta t) = V(S, t) - \delta t \frac{\partial V(S, t)}{\partial t} + O(\delta t^2)$$

The last term on the right-hand side of the equation states that no terms will be considered which are of order two or greater with respect to the time difference δt. If the time difference is small enough, we can assume that the time dependence of $V(S, t)$ can be adequately described if we simply neglect terms of order $O(\delta t^2)$. Note that on the left-hand side of the above equation, the derivative of the value evaluated at time $t - \delta t$ appears, whereas the right-hand side consists of terms evaluated at time t. The trick is now *not* to express the right-hand side in terms of a difference quotient in time (we would not have accomplished anything by doing so since it would involve introducing another time point $t - \delta t$ or $t + \delta t$) but rather to express the time derivative in terms of partial derivatives with respect to S obtained from the Black-Scholes equation 8.8. These partial derivatives with respect to S are evaluated at time t so that the value of the derivative at an earlier time $t - \delta t$ can in fact be recovered solely from information available at time t. Here, we *explicitly* calculate earlier values from those (known) values from a later time. This method is thus referred to as the *explicit method.*

10.1.2 The Implicit Method

In the implicit method, we use the Taylor series expansion in the time variable to obtain an expression for values at a *later* time $t + \delta t$ from the values at time t:

$$V(S, t + \delta t) = V(S, t) + \delta t \frac{\partial V(S, t)}{\partial t} + O(\delta t^2) \tag{10.3}$$

In this case, later (known) values are expanded in terms of earlier (unknown) ones. This expansion can only be used *implicitly* to calculate the unknown values from the known ones, hence the name *implicit method.*

In both cases, the partial derivatives of V with respect to time are replaced with partial derivatives with respect to S via the Black-Scholes equation 8.8.

$$\frac{\partial V(S, t)}{\partial t} = r(t)V(S, t) \; - [r(t) - q(t)]\, S \frac{\partial V(S, t)}{\partial S} - \frac{1}{2}\sigma^2(S, t)S^2 \frac{\partial^2 V(S, t)}{\partial S^2} \tag{10.4}$$

The S-derivatives appearing here will subsequently be approximated by finite difference quotients, as will be shown below.

10.1.3 Combinations of Explicit and Implicit Methods (Crank-Nicolson)

The two methods described above can be combined by taking a linear combination of the two respective Taylor series expansions. To avoid the appearance of three different

time points in the resulting expression ($t-\delta t, t$ and $t+\delta t$) a change in variable in one of the Taylor series should be made. For example, the transformation $t \to t+\delta t$ in the Taylor series expansion for the explicit method (and dividing by δt) yields an equation of the form:

$$\frac{V(S,t+\delta t)-V(S,t)}{\delta t} = \frac{\partial V(S,t+\delta t)}{\partial t} + O(\delta t) \tag{10.5}$$

Here, we have rearranged the Taylor series so that the left-hand side is written in terms of a *difference quotient* in t, while on the right-hand side a *differential* quotient appears (which will later be replaced by a difference quotient with respect to the underlying price). Note that dividing through by δt has the effect of reducing the order of the error term to a linear order in δt. An analogous procedure for the implicit method allows its respective Taylor series to be rearranged as well (here, a variable transformation is unnecessary since the series is already expressed in terms of t and $t+\delta t$):

$$\frac{V(S,t+\delta t)-V(S,t)}{\delta t} = \frac{\partial V(S,t)}{\partial t} + O(\delta t) \tag{10.6}$$

The only thing now distinguishing the two expressions is that the differential quotient on the right-hand side (and thus, the difference quotients with respect to S yet to be determined) is written in terms of the time t, whereas in the expression derived from the explicit method, the differential quotient is expressed in terms of the later time point $t+\delta t$. Naturally, the equality holds if we take any linear combination of the two equations:

$$\frac{V(S,t+\delta t)-V(S,t)}{\delta t} = (1-\theta)\frac{\partial V(S,t+\delta t)}{\partial t} + \theta\frac{\partial V(S,t)}{\partial t} + O(\delta t)\,,\ 0 \le \theta \le 1$$

Note the following correspondence[2] between the notation above and the the discrete notation introduced in Equation 10.1:

$$\begin{aligned} S \mathrel{\widehat{=}} S_i \quad , \quad t \mathrel{\widehat{=}} t_j \quad , \quad t+\delta t \mathrel{\widehat{=}} t_{j+1} \\ W_{i,j} = W(S_i,t_j) \mathrel{\widehat{=}} V(S,t) \\ W_{i,j+1} = W(S_i,t_{j+1}) \mathrel{\widehat{=}} V(S,t+\delta t) \end{aligned}$$

In the discrete notation the above equation reads

$$\frac{W_{i,j+1}-W_{i,j}}{t_{j+1}-t_j} = (1-\theta)\frac{\partial V(S_i,t_{j+1})}{\partial t} + \theta\frac{\partial V(S_i,t_j)}{\partial t} + O\left(t_{j+1}-t_j\right)\ ,\ 0 \le \theta \le 1 \tag{10.7}$$

In this notation, the equation holds for *non*-uniform grids, as well. Setting $\theta = 1$, we obtain the implicit method, with $\theta = 0$ the explicit method. The particular choice of $\theta = 1/2$ has a special name. It is known as the *Crank-Nicolson method*.

Equations 10.5 and 10.6 can also be interpreted as follows: the differen*ce* quotients ("finite differences") on the left-hand side of the equations are approximations of the differen*tial* quotients with respect to time found on the right-hand side of the equations

[2]The sign $\mathrel{\widehat{=}}$ means "corresponds to".

(with this interpretation we unfortunately lose the intuitive interpretation of expanding earlier values in terms of later ones).

The above approximations are exact up to *linear* terms in δt. There are several methods available for approximating partial derivatives using finite differences. A greater accuracy can be obtained if, for instance, all three time points t_{j-1}, t_j and t_{j+1} are included in the finite differences approximating the time derivative. Then the approximation is exact up to second order. This three-time procedure requires that the time derivative be approximated by a carefully selected convex combination of forward, backward and symmetric finite differences. Another possibility is to use symmetric finite differences. This also gives an approximation exact up to second order, however, does not lead to a stable procedure for solving the differential equation. In this book we will not pursue such more precise approximations for the time derivative.

10.1.4 Symmetric Finite Differences of the Underlying Price

For the sake of consistency, the difference quotients with respect to the underlying price S should be exact up to order $O(\delta S^2)$ since it follows from the random walk assumption that $dS \sim \sqrt{dt}$ and thus

$$\delta t \sim \delta S^2$$

This means that in order to attain the same degree of accuracy as in the time direction, the approximation in the S direction must be exact up to order $O(\delta S^2)$. To achieve this we will use *symmetric differences* to approximate the derivatives of first and second order with respect to the underlying needed for Equation 10.4.

In order to demonstrate the concepts without making the notation unnecessarily complicated, we begin by assuming that the distance between grid points is constant as in Equation 10.2 and then generalize to non-uniform grids. As above, we expand the value function in its Taylor series, this time in the S dimension:

$$\begin{aligned}
&V(S-\delta S,t)\\
&= V(S,t) - \delta S\frac{\partial V(S,t)}{\partial S} + \frac{1}{2}\delta S^2\frac{\partial^2 V(S,t)}{\partial S^2} - \frac{1}{6}\delta S^3\frac{\partial^3 V(S,t)}{\partial S^3} + O(\delta S^4)\\
&V(S+\delta S,t)\\
&= V(S,t) + \delta S\frac{\partial V(S,t)}{\partial S} + \frac{1}{2}\delta S^2\frac{\partial^2 V(S,t)}{\partial S^2} + \frac{1}{6}\delta S^3\frac{\partial^3 V(S,t)}{\partial S^3} + O(\delta S^4)
\end{aligned}$$

Subtracting the first equation from the second and subsequently dividing by δS yields an approximation of the first derivative which is exact up to second order $O(\delta S^2)$. Adding the two equations and then dividing by δS^2 yields an approximation of the second derivative which is also exact up to second order $O(\delta S^2)$:

$$\begin{aligned}
\frac{\partial V(S,t)}{\partial S} &= \frac{V(S+\delta S,t) - V(S-\delta S,t)}{2\,\delta S} + O(\delta S^2)\\
\frac{\partial^2 V(S,t)}{\partial S^2} &= \frac{V(S+\delta S,t) - 2\;V(S,t) + V(S-\delta S,t)}{\delta S^2} + O(\delta S^2)
\end{aligned} \tag{10.8}$$

For general, *non*-uniforms grids, the above expressions are somewhat more complicated and their derivation is a bit more technical. The principle, however, remains the same: the partial derivatives of V with respect to the underlying evaluated at point S_i can be approximated up to order two with symmetric differences. To this end, V evaluated at $S_{i-1}$and S_{i+1}, is first expanded in its Taylor series about the points S_i:

$$\begin{aligned} V(S_{i-1},t) &= V(S_i,t) - \frac{\partial V(S_i,t)}{\partial S}(S_i - S_{i-1}) \\ &\quad + \frac{1}{2}\frac{\partial^2 V(S_i,t)}{\partial S^2}(S_i - S_{i-1})^2 + O\left((S_i - S_{i-1})^3\right) \\ V(S_{i+1},t) &= V(S_i,t) + \frac{\partial V(S_i,t)}{\partial S}(S_{i+1} - S_i) \\ &\quad + \frac{1}{2}\frac{\partial^2 V(S_i,t)}{\partial S^2}(S_{i+1} - S_i)^2 + O\left((S_{i+1} - S_i)^3\right) \end{aligned}$$

We need this approximation for the partial derivatives to be exact up to second order. We thus neglect all terms of order $O((\delta S_{\max})^3)$ where $\delta S_{\max}$ denotes the greatest distance between two neighboring nodes in the S-grid. Unlike the uniform grid case, mere addition and subtraction of the two equations does not isolate the desired differential quotients since is possible that $(S_i - S_{i-1}) \neq (S_{i+1} - S_i)$. To overcome this inconvenience, we attempt to express the differential quotients as a linear combination of the function evaluated at the points S_i, $S_{i-1}$and S_{i+1}. We begin by assuming that the first derivative can be written as

$$\frac{\partial V(S_i,t)}{\partial S} = aV(S_{i-1},t) + bV(S_i,t) + cV(S_{i+1},t)$$

Substituting the above Taylor series for $V(S_{i+1},t)$ and $V(S_{i-1},t)$ into this equation and using the linear independence of $V(S_i,t)$, $\frac{\partial V(S_i,t)}{\partial S}$ and $\frac{\partial^2 V(S_i,t)}{\partial S^2}$ to compare their coefficients leads to the following system of equations for the unknown coefficients a, b and c

$$\begin{aligned} a + b + c &= 0 \\ -(S_i - S_{i-1})a + (S_{i+1} - S_i)c &= 1 \\ (S_i - S_{i-1})^2 a + (S_{i+1} - S_i)^2 c &= 0 \end{aligned}$$

which has the solution

$$\begin{aligned} a &= -\frac{S_{i+1} - S_i}{(S_i - S_{i-1})(S_{i+1} - S_{i-1})} \\ b &= \left(\frac{S_{i+1} - S_i}{S_i - S_{i-1}} - \frac{S_i - S_{i-1}}{S_{i+1} - S_i}\right)\frac{1}{S_{i+1} - S_{i-1}} \\ c &= +\frac{S_i - S_{i-1}}{(S_{i+1} - S_i)(S_{i+1} - S_{i-1})} \end{aligned}$$

The desired approximation of the first partial derivative with respect to S is thus

$$\begin{aligned}\frac{\partial V(S_i,t)}{\partial S} = \frac{1}{S_{i+1}-S_{i-1}}[&-\frac{S_{i+1}-S_i}{S_i-S_{i-1}}V(S_{i-1},t) \\ &+\left(\frac{S_{i+1}-S_i}{S_i-S_{i-1}}-\frac{S_i-S_{i-1}}{S_{i+1}-S_i}\right)V(S_i,t)+\frac{S_i-S_{i-1}}{S_{i+1}-S_i}V(S_{i+1},t)] \\ &+O\left((\delta S_{\max})^2\right)\end{aligned} \tag{10.9}$$

For the special case of a uniform S-grid with $S_{i+1}-S_i = S_i - S_{i-1} = \delta S$, this reduces to the expression in Equation 10.8.

The same approach can be taken to isolate the second derivative. This time, we assume that the second derivative can be represented as a linear combination of the value function evaluated at the points S_{i-1}, S_i and S_{i+1}:

$$\frac{\partial^2 V(S_i,t)}{\partial S^2} = aV(S_{i-1},t)+bV(S_i,t)+cV(S_{i+1},t)$$

As was done above, the Taylor expansions for $V(S_{i+1},t)$ and $V(S_{i-1},t)$ are substituted into this equation. Comparing the coefficients of $V(S_i,t)$, $\frac{\partial V(S_i,t)}{\partial S}$ and $\frac{\partial^2 V(S_i,t)}{\partial S^2}$ again leads to a system of equations for the unknown coefficients a, b and c, now given by

$$\begin{aligned}a+b+c &= 0 \\ -(S_i-S_{i-1})a+(S_{i+1}-S_i)c &= 0 \\ (S_i-S_{i-1})^2a+(S_{i+1}-S_i)^2c &= 2\end{aligned}$$

which has the solution

$$\begin{aligned}a &= \frac{2}{(S_i-S_{i-1})(S_{i+1}-S_{i-1})} \\ b &= -\frac{2}{S_{i+1}-S_{i-1}}\left(\frac{1}{S_i-S_{i-1}}+\frac{1}{S_{i+1}-S_i}\right) \\ c &= \frac{2}{(S_{i+1}-S_i)(S_{i+1}-S_{i-1})}\end{aligned}$$

The approximation of the second derivative of V with respect to S is now given by

$$\begin{aligned}\frac{\partial^2 V(S_i,t)}{\partial S^2} = \frac{2}{S_{i+1}-S_{i-1}}[&\frac{V(S_{i-1},t)}{S_i-S_{i-1}} \\ &-\left(\frac{1}{S_i-S_{i-1}}+\frac{1}{S_{i+1}-S_i}\right)V(S_i,t)+\frac{V(S_{i+1},t)}{S_{i+1}-S_i}]+O(\delta S_{\max})\end{aligned} \tag{10.10}$$

For a uniform S-grid this reduces to the expression in Equation 10.8.

10.2 Difference Schemes

The approximations for the derivatives with respect to S can now be substituted into Equation 10.4. For the sake of simplicity, we again consider first the case of a uniform grid as in Equation 10.2, i.e. we will use the approximation given by Equation

10.8. Under this assumption $S_i = S_0 + i\,\delta S$ holds and the approximation 10.4 for the differential quotient with respect to t evaluated at the point $t = t_j$ takes the form

$$\begin{aligned}\frac{\partial V(S_i,t_j)}{\partial t} &\approx r_j W_{i,j} \;-\; (r_j - q_j)\, S_i\, \frac{W_{i+1,j} - W_{i-1,j}}{2\delta S} \\ &\quad - \frac{1}{2}\,\sigma_{i,j}^2\, S_i^2\, \frac{W_{i+1,j} - 2\,W_{i,j} + W_{i-1,j}}{\delta\, S^2} \\ &= A_{i,j} W_{i-1,j} + B_{i,j} W_{i,j} + C_{i,j} W_{i+1,j}\end{aligned}$$

where

$$\begin{aligned}A_{i,j} &= \frac{S_i}{2\delta S}(r_j - q_j - \frac{S_i}{\delta S}\,\sigma_{i,j}^2) \\ B_{i,j} &= r_j + \left(\frac{S_i}{\delta S}\right)^2 \sigma_{i,j}^2 \\ C_{i,j} &= -\frac{S_i}{2\delta S}(r_j - q_j + \frac{S_i}{\delta S}\sigma_{i,j}^2)\end{aligned} \tag{10.11}$$

On the right-hand side of the equation, we have again denoted the approximation of the exact solution $V(S_i, t_j)$ evaluated at the grid points by $W_{i,j} = W(S_i, t_j)$. At this stage, we allow the interest rate and the volatility to depend on time. This is the only reason why the coefficients A, B and C have been equipped with the index j. For time-independent interest rates and volatilities, the index j on the coefficients is superfluous.

The corresponding expression for the *non*-uniform grid has exactly the same structure with somewhat more complicated coefficients A, B and C. Replacing the partial derivatives with respect to S with their approximations 10.9 and 10.10 for the non-uniform grid in 10.4 yields the following expression for the differential quotient with respect to t evaluated at $t = t_j$

$$\begin{aligned}\frac{\partial V(S_i,t_j)}{\partial t} &\approx r_j\, W_{i,j} - (r_j - q_j)\, S_i \{\frac{1}{S_{i+1} - S_{i-1}}[-\frac{S_{i+1} - S_i}{S_i - S_{i-1}} W_{i-1,j} \\ &+ \left(\frac{S_{i+1} - S_i}{S_i - S_{i-1}} - \frac{S_i - S_{i-1}}{S_{i+1} - S_i}\right) W_{i,j} + \frac{S_i - S_{i-1}}{S_{i+1} - S_i} W_{i+1,j}]\} \\ &- \frac{1}{2}\sigma_{i,j}^2\, S_i^2 \{\frac{2}{S_{i+1} - S_{i-1}}[\frac{W_{i-1,j}}{S_i - S_{i-1}} \\ &- \left(\frac{1}{S_i - S_{i-1}} + \frac{1}{S_{i+1} - S_i}\right) W_{i,j} + \frac{W_{i+1,j}}{S_{i+1} - S_i}]\} \\ &= \frac{S_i}{(S_{i+1} - S_{i-1})(S_i - S_{i-1})} \left\{(r_j - q_j)(S_{i+1} - S_i) - \sigma_{i,j}^2\, S_i\right\} W_{i-1,j} \\ &+ \{r_j - \frac{(r_j - q_j)\, S_i}{S_{i+1} - S_{i-1}} \left(\frac{S_{i+1} - S_i}{S_i - S_{i-1}} - \frac{S_i - S_{i-1}}{S_{i+1} - S_i}\right) \\ &+ \frac{\sigma_{i,j}^2\, S_i^2}{S_{i+1} - S_{i-1}} \left(\frac{1}{S_i - S_{i-1}} + \frac{1}{S_{i+1} - S_i}\right)\} W_{i,j} \\ &- \frac{S_i}{(S_{i+1} - S_{i-1})(S_{i+1} - S_i)} \left\{(r_j - q_j)(S_i - S_{i-1}) + \sigma_{i,j}^2\, S_i\right\} W_{i+1,j}\end{aligned}$$

Thus, as was the case for the uniform grid

$$\frac{\partial V(S_i,t_j)}{\partial t} \approx A_{i,j}\,W_{i-1,j} + B_{i,j}\,W_{i,j} + C_{i,j}\,W_{i+1,j} \tag{10.12}$$

with the slightly more complicated coefficients

$$\begin{aligned}
A_{i,j} &= \frac{S_i}{(S_{i+1}-S_{i-1})(S_i-S_{i-1})}\left\{(r_j-q_j)(S_{i+1}-S_i) - \sigma_{i,j}^2\,S_i\right\} \qquad (10.13)\\
B_{i,j} &= r_j + \frac{S_i}{(S_{i+1}-S_{i-1})}\{-(r_j-q_j)\ \left(\frac{S_{i+1}-S_i}{S_i-S_{i-1}} - \frac{S_i-S_{i-1}}{S_{i+1}-S_i}\right)\\
&\quad + \sigma_{i,j}^2\,S_i\left(\frac{1}{S_i-S_{i-1}} + \frac{1}{S_{i+1}-S_i}\right)\}\\
C_{i,j} &= -\frac{S_i}{(S_{i+1}-S_{i-1})(S_{i+1}-S_i)}\left\{(r_j-q_j)(S_i-S_{i-1}) + \sigma_{i,j}^2\,S_i\right\}
\end{aligned}$$

For non-uniform grids as well, the time dependence of the coefficients A, B and C is solely a consequence of the time dependence of r and σ. For time-independent interest rates and volatilities, the coefficients need not have an index j. For uniform grids (at least in the S-direction), in other words, for $S_i = S_0 + i\,\delta S$, the above expression for the coefficients reduces to 10.11.

Substituting this approximation 10.12 for the differential quotient with respect to t in the general equation 10.7 finally gives the general form of the finite difference scheme:

$$\begin{aligned}
&\theta A_{i,j}\,W_{i-1,j} + (\theta B_{i,j} + \frac{1}{t_{j+1}-t_j})\ W_{i,j} + \theta C_{i,j}\,W_{i+1,j} \qquad (10.14)\\
&\approx \frac{W_{i,j+1}}{(t_{j+1}-t_j)} - (1-\theta)\,[A_{i,j+1}\,W_{i-1,j+1} + B_{i,j+1}\ W_{i,j+1} + C_{i,j+1}\,W_{i+1,j+1}]
\end{aligned}$$

This form holds for both uniform and non-uniform grids, only the coefficients A, B and C are different in the two cases as shown in Equations 10.11 and 10.13.

As in 10.7, θ can take on arbitrary values between 0 and 1 giving the *implicit* method for $\theta = 1$, the *explicit* method for $\theta = 0$ and the *Crank-Nicolson* method for $\theta = 1/2$. In order to illustrate the importance of this difference equation, note that the values to be calculated on the left-hand side of the equation consist of terms evaluated at time point t_j whereas those on the right-hand side concern only values at time point t_{j+1}, and as such are known, having been calculated in the previous step. Thus, three option values evaluated at time t_j are calculated from three option values evaluated at time t_{j+1}.

The range of the time index j here is always $j = 0, \ldots N-1$ since $j = N$ is already given by the initial condition, i.e. the payoff profile at maturity $T = t_N$. The range of the S-index i (see Equation 10.1) is at least $i = 1, \ldots, M-1$, but can take on the values including $i = 0$ and/or $i = M$ (depending on the boundary conditions to be satisfied, see below). The difference equation for $i = 0$ and $i = M$ seems problematic at first glance since values of W are required at points not defined in the grid, for example " $W_{0-1,j}$" or "$W_{M+1,j}$". As will be shown below, these problems can in fact be overcome by consideration of the boundary conditions themselves.

The difference scheme can be written in matrix form (here, for the case where the range of the index i is given by $i = 1, \ldots, M-1$; other cases can be expressed analogously):

$$\underbrace{\begin{pmatrix} a_1 & b_1 & c_1 & 0 & \cdots & \cdots & \ldots & 0 \\ 0 & a_2 & b_2 & c_2 & 0 & & & \vdots \\ \vdots & 0 & a_3 & b_3 & c_3 & 0 & & \vdots \\ \vdots & & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & a_{M-1} & b_{M-1} & c_{M-1} \end{pmatrix}}_{M+1 \text{ Columns}} \underbrace{\begin{pmatrix} W_{0,j} \\ W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ W_{M,j} \end{pmatrix}}_{M+1 \text{ Rows}} = \underbrace{\begin{pmatrix} D_{1,j} \\ D_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ D_{M-1,j} \end{pmatrix}}_{M-1 \text{ Rows}} \tag{10.15}$$

where

$$a_i = \theta A_{i,j} \ , \ \ b_i = \theta B_{i,j} + \frac{1}{t_{j+1} - t_j} \ , \ \ c_i = \theta C_{i,j}$$

$$D_{i,j} = \frac{W_{i,j+1}}{(t_{j+1} - t_j)} - (1-\theta)\left[A_{i,\ j+1}\, W_{i-1,j+1} + B_{i,j+1}\, W_{i,j+1} + C_{i,j+1}\, W_{i+1,j+1}\right]$$

The $D_{i,j}$ on the right-hand side depend only on the values at time t_{j+1} which have already been calculated in the previous step and are hence completely determined. For each time step j, we have $(M-1)$ equations for $(M+1)$ unknown $W_{i,j.}$ Other ranges for the index i give the same result: the system of equations is underdetermined; there are two equations fewer than there are unknowns. The two additional equations needed to solve the above system are provided by two boundary conditions.

10.2.1 Initial Conditions

Besides the specification of boundary conditions with respect to the S variable, we also need a boundary condition with respect to the t variable (called *initial condition*) to be satisfied in order to obtain a unique solution to the above defined difference equation. The initial condition is specified by the *payoff profile* $P(S)$ of the derivative concerned, i.e. by the value $V(S, t_N = T)$ given by

$$V(S,T) = P(S)$$

or in the discrete "grid notation"

$$W_{i,N} = P_i$$

For example, for the payoff profile $P(S)$ for a European call option with strike price K, the initial condition is given by

$$V(S,T) = P(S) \ = \ \max(S-K, 0) \quad \Longrightarrow \quad W_{i,N} = \max(S_i - K, 0)$$

10.2.2 Dirichlet Boundary Conditions

If either the terms of the option contract (for example, barrier options) or some other information allow us to specify directly the *value* of the option for certain values of S, these values can be used as boundary conditions for the S-grid. Such boundary conditions where the option value itself is given at the boundary are called *Dirichlet boundary conditions*. Let R^U denote these given option values at the *upper boundary* S_M, and R^L denote the given option values at the *lower boundary* S_0, i.e.:

$$V(S_M, t) = R^U(t), \quad V(S_0, t) = R^L(t)$$

or in the discrete "grid notation"

$$W_{M,j} = R_j^U \ , \quad W_{0,j} = R_j^L \ .$$

It is often the case that only approximations for R_j^U and R_j^L are known. If this is the case, the difference scheme provides an approximation which is at best as good as this approximation for the boundary conditions. If, for example, the Crank-Nicolson scheme is applied but the boundary condition can only be approximated in the first order of S, then the solution procedure as a whole is exact up to first order even though the Crank-Nicolson scheme provides an approximation which is exact up to second order.

Two of the option values to be calculated for each time j are thus specified directly if Dirichlet boundary conditions are given. The dimension of the problem (in the sense of dimension equals number of unknowns) is therefore only $(M-1)$. This is exactly the number of equations in 10.15. Our goal now is to transform Equation 10.15 into a system consisting of a square matrix, i.e. to reduce the dimension "in each direction" to $(M-1)$. To keep the boundary conditions separate, we write the vector of option values in the form

$$\begin{pmatrix} W_{0,j} \\ W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ W_{M,j} \end{pmatrix} = \begin{pmatrix} 0 \\ W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,j} \\ 0 \end{pmatrix} + \begin{pmatrix} W_{0,j} \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ W_{M,j} \end{pmatrix}$$

In the first vector, only the (yet to be determined) option values away from the boundary of the grid appear. We call this the *unknown vector*. The second vector contains only the (already specified) values of the option on the boundary. We refer to this vector as the *known vector*. The matrix in Equation 10.15 acts on both of these vectors. Let us first consider the unknown vector: the only element in the first column of the matrix, namely a_1, acts only on the first row of the unknown vector. This, however, equals zero. We obtain from the matrix multiplication of the unknown vector the same

result as if the first column in the matrix and the first row in the vector were removed. The situation is the same for the last column of the matrix and the last row of the unknown vector. Thus we can simply remove the last column of the matrix and the last row of the unknown vector without changing the result of the matrix multiplication.

Matrix multiplication of the full matrix in Equation 10.15 with the *known* vector is explicitly performed. Combining everything the system of equations in Equation 10.15 can be equivalently written using an $(M-1)\times(M-1)$ matrix:

$$\underbrace{\begin{pmatrix} b_1 & c_1 & 0 & \cdots & \cdots & 0 \\ a_2 & b_2 & c_2 & 0 & & \cdots \\ 0 & a_3 & b_3 & c_3 & \ddots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & c_{M-2} \\ 0 & \cdots & \cdots & 0 & a_{M-1} & b_{M-1} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,j} \end{pmatrix}}_{\mathbf{W}_j} + \begin{pmatrix} a_1\, W_{0,j} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ c_{M-1}\, W_{M,j} \end{pmatrix} = \begin{pmatrix} D_{1,j} \\ D_{2,j} \\ \vdots \\ \vdots \\ D_{M-2,j} \\ D_{M-1,j} \end{pmatrix}$$

$$\Longleftrightarrow$$

$$\underbrace{\begin{pmatrix} b_1 & c_1 & 0 & \cdots & \cdots & 0 \\ a_2 & b_2 & c_2 & 0 & & \cdots \\ 0 & a_3 & b_3 & c_3 & \ddots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & c_{M-2} \\ 0 & \cdots & \cdots & 0 & a_{M-1} & b_{M-1} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,j} \end{pmatrix}}_{\mathbf{W}_j} = \begin{pmatrix} d_{1,j} \\ d_{2,j} \\ \vdots \\ \vdots \\ d_{M-2,j} \\ d_{M-1,j} \end{pmatrix} \tag{10.16}$$

with the coefficients given by

$$\begin{aligned} a_i &= \theta A_{i,j}\ ,\ \ b_i = \theta B_{i,j} + \frac{1}{t_{j+1}-t_j}\ ,\ \ c_i = \theta C_{i,j} \quad \text{for } i = 1,\ldots,M-1 \\ d_{i,j} &= \frac{W_{i,j+1}}{(t_{j+1}-t_j)} \qquad\qquad \text{for } i = 2,\ldots,M-2 \\ &\quad - (1-\theta)\left[A_{i,\ j+1}\, W_{i-1,j+1} + B_{i,j+1}\, W_{i,j+1} + C_{i,j+1}\, W_{i+1,j+1}\right] \\ d_{1,j} &= \frac{W_{1,j+1}}{(t_{j+1}-t_j)} - (1-\theta)\left[B_{1,\ j+1}\, W_{1,j+1} + C_{1,j+1}\, W_{2,j+1}\right] \\ &\quad - \left[\theta A_{1,j}\ R_j^L + (1-\theta) A_{1,j+1}\, R_{j+1}^L\right] \\ d_{M-1,j} &= \frac{W_{M-1,j+1}}{(t_{j+1}-t_j)} - (1-\theta)\left[A_{M-1,j+1}\, W_{M-2,j+1} + B_{M-1,j+1}\, W_{M-1,j+1}\right] \\ &\quad - \left[\theta C_{M-1,j}\, R_j^U + (1-\theta) C_{M-1,j+1}\, R_{j+1}^U\right] \end{aligned} \tag{10.17}$$

Along the way, the result of the matrix multiplication with the known vector (i.e. with the boundary condition) has been brought over to the right-hand side. This right-hand side (the $d_{i,j}$) only depends on the values at the time points t_{j+1} (which have already

been calculated in the previous step) and on the (given) boundary conditions. The $d_{i,j}$ are thus completely determined. The only unknowns in this system are the $(M-1)$ elements of the vector W_j. These can now be calculated *if* it is possible to invert the matrix $\mathbf{A}$. Because of the special form of the matrix (the only non-zero elements are in the diagonal and the two off-diagonals), a calculation-intensive matrix inversion can be avoided. Instead, we employ a very fast procedure, known as the *L-U decomposition*. To this end, we decompose the matrix $\mathbf{A}$ into the product of a matrix $\mathbf{L}$, which contains non-zero elements only in the diagonal and the "lower" off-diagonal and a matrix $\mathbf{U}$ whose only non-zero elements are found in the diagonal and the "upper" off-diagonal:

$$\mathbf{A} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ l_2 & 1 & 0 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & l_{M-2} & 1 & 0 \\ 0 & \cdots & \cdots & 0 & l_{M-1} & 1 \end{pmatrix}}_{\mathbf{L}} \underbrace{\begin{pmatrix} h_1 & u_1 & 0 & \ldots & \ldots & 0 \\ 0 & h_2 & u_2 & 0 & \ldots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & \ddots & u_{M-2} \\ 0 & \cdots & \cdots & \cdots & 0 & h_{M-1} \end{pmatrix}}_{\mathbf{U}}$$

From this *ansatz* the elements of the matrices $\mathbf{L}$ and $\mathbf{U}$ are determined to be

$$\begin{aligned} u_i &= c_i && \text{für } \; i = 1, \ldots M-2 \\ h_1 &= b_1 \\ l_i &= \frac{a_i}{h_{i-1}} && \text{für } \; i = 2, \ldots M-1 \\ h_i &= b_i - l_i\, u_{i-1} && \text{für } \; i = 2, \ldots M-1 \end{aligned}$$

Calculating the elements in the order indicated above, we can easily compute all the elements appearing in both matrices. Inverting $\mathbf{L}$ and $\mathbf{U}$ is quite simple. We begin by writing

$$\mathbf{A}\,\mathbf{W}_j = \mathbf{L}\,\underbrace{\mathbf{U}\,\mathbf{W}_j}_{\mathbf{x}} =: \mathbf{L}\mathbf{x}$$

and solve the system for our newly defined vector $\mathbf{x}$:

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ l_2 & 1 & 0 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & l_{M-2} & 1 & 0 \\ 0 & \cdots & \cdots & 0 & l_{M-1} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_{M-1} \end{pmatrix} = \begin{pmatrix} d_{1,j} \\ d_{2,j} \\ d_{3,j} \\ \vdots \\ \vdots \\ d_{M-2,j} \\ d_{M-1,j} \end{pmatrix} .$$

To solve this system for $\mathbf{x}$, we start with the first row of the matrix which contains

only one non-zero element and proceed from top to bottom to obtain

$$\begin{aligned} x_1 &= d_{1,j} \\ x_i &= d_{i,j} - l_i\, x_{i-1} \quad \text{für} \quad i = 2 \dots M-1 \end{aligned}$$

Now that the vector $\mathbf{x}$ is known, the option values $W_{i,j}$ can simply be calculated from the above definition of $\mathbf{x}$ by solving the equation $\mathbf{U}\, W_j = \mathbf{x}$:

$$\begin{pmatrix} h_1 & u_1 & 0 & \dots & \dots & 0 \\ 0 & h_2 & u_2 & 0 & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & \ddots & u_{M-2} \\ 0 & \dots & \dots & \dots & 0 & h_{M-1} \end{pmatrix} \begin{pmatrix} W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,j} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_{M-1} \end{pmatrix}.$$

To "invert" the matrix $\mathbf{U}$, we begin this time with the last row of the matrix and work our way up to obtain

$$\begin{aligned} W_{M-1,j} &= \frac{x_{M-1}}{h_{M-1}} \\ W_{i,j} &= \frac{x_i - u_i W_{i+1,j}}{h_i} \quad \text{for} \quad i = 1 \dots M-2 \end{aligned} \tag{10.18}$$

Thus, we have found an explicit solution for the general finite difference scheme 10.14, expressing the option values at time t_j in terms of the values at a later time point t_{j+1} under consideration of the given values at the boundaries.

We will now consider the plain vanilla put and call as well as the barrier option as examples for the application of this scheme with Dirichlet boundary conditions. While, in principle, we must find the values of the plain vanilla options on the entire set from $S = 0$ to $S = \infty$, the value of the barrier option is specified directly in the option contract on one side of the barrier (and on the barrier itself). The area to be covered by its grid is thus smaller than that for plain vanilla options.

Let us, however, begin by considering plain vanilla options. For these options, the boundary conditions must be specified at "$S = 0$" and "$S = \infty$". We choose a lower bound S_0 so small[3], that the value of the *call* for such values of the underlying price is negligible (the option is well out of the money). Furthermore, we choose an upper bound S_M so large that the value of a *put* is negligible for such values of the underlying price (the option is well out of the money). The put-call parity is then used to establish the other boundary values for each of the two options. Recall that for European options on an underlying paying a dividend yield q the put-call parity, Equation 7.7, is given by:

$$\text{Price(Call)} - \text{Price(Put)} = \text{Price(Forward)} = Se^{-q(T-t)} - Ke^{-r(T-t)}$$

[3]We could also directly define $S_0 = 0$. The boundary value of both calls and puts for this case are given exactly by 10.22.

The right-hand side is the value of a forward contract[4] with strike price K. Since the put at the upper boundary is worthless, the call option has the same value as the forward contract (the option is so far in the money that it will be exercised with certainty). Conversely, the call at the lower boundary is worthless, therefore the put has the same value as a short forward contract (the option is so far in the money that it will be exercised with certainty). Summarizing, we have the following Dirichlet boundary conditions for the plain vanilla options:

	$S_0 \approx 0$	$S_M \approx \infty$
Call	$R_j^L = 0$	$R_j^U = S_M e^{-q(T-t_j)} - Ke^{-r(T-t_j)}$
Put	$R_j^L = -S_0 e^{-q(T-t_j)} + Ke^{-r(T-t_j)}$	$R_j^U = 0$

We now take an up-and-out barrier call option as an example of a somewhat exotic and path dependent option and specify its boundary condition. The initial condition and the boundary condition for $S = 0$ is exactly the same as that of a plain vanilla call but as soon as S attains the upper barrier level H, the option becomes worthless. In general, the option's holder still receives a payment if the barrier is reached, called the *rebate*. Let us assume that the rebate R is due at precisely the knock-out time point (when S touches the barrier). Then the option's value is given by the value of the rebate, namely R. It is thus convenient to select the upper boundary of the grid as $S_M = H$. The boundary condition for such a barrier option is then simply $W_{M,\,j} = R$.

10.2.3 Neumann Boundary Condition

It often occurs that the first derivative of the option price S is specified at the boundary of the grid rather than the option price itself,

$$\left.\frac{\partial V(S,t)}{\partial S}\right|_{S=S_M} = R^U(t), \quad \left.\frac{\partial V(S,t)}{\partial S}\right|_{S=S_0} = R^L(t)$$

Boundary conditions of this type are called *Neumann boundary conditions*[5]. Since these boundary conditions do not directly specify the *values* of the solution (the option values) on the boundary, all $(M+1)$ values must be calculated in each time slice. The dimension (the number of values to be calculated) of the problem is thus $(M+1)$. In order to obtain the same number of equations, the index i in Equation 10.14 must range from 0 to M. As a result, two "additional unknowns", namely "$W_{-1,j}$" and "$W_{M+1,j}$" appear in the system of equations. The two Neumann boundary conditions will be used to eliminate these additional unknowns. In comparison to the Dirichlet conditions, we will then have system of equations consisting of two more equations but exactly as many equations as unknowns.

Since we use symmetric differences in the S-direction (see for example Equation 10.9) the grid must be extended on the boundary by one grid point in the S-direction

[4]For the sake of simplicity, the price of the forward contract is given for the case of a flat interest rate term structure, a flat dividend yield curve and no discrete dividend payments. However, this put-call parity also holds for interest rates and dividend yields which are time-dependent.

[5]As already pointed out in the section on Dirichlet boundary conditions, the solution is at best as exact as the given boundary conditions. This is important in all cases where there are only approximations to the boundary conditions $R^U(t)$ and $R^L(t)$ available.

for each index j. In other words, we add two additional grid points S_{-1} and S_{M+1} at each time point t_j. This means that the index i in the difference equation 10.14 takes on the values $i = 0, \ldots, M$. In the finite difference approximation 10.9 of the first derivative with respect to S, we can substitute the respective Neumann conditions for the cases $i = 0$ and $i = M$ on the left-hand side of the equation and then rearrange the terms so that the option values outside the grid are expressed in terms of those defined within the grid and the boundary conditions. This yields

$$W_{M+1,j} \approx \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2 W_{M-1,j} + \left[1 - \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2\right] W_{M,j} + \frac{S_{M+1} - S_M}{S_M - S_{M-1}}(S_{M+1} - S_{M-1}) R_j^U$$

at the upper boundary, i.e. for $i = M$. Likewise for $i = 0$ on the lower boundary we have

$$W_{-1,j} \approx \left[1 - \left(\frac{S_0 - S_{-1}}{S_1 - S_0}\right)^2\right] W_{0,j} + \left(\frac{S_0 - S_{-1}}{S_1 - S_0}\right)^2 W_{1,j} - \frac{S_0 - S_{-1}}{S_1 - S_0}(S_1 - S_{-1}) R_j^L$$

Substituting these expressions into Equation 10.12, we obtain the approximation for the differential quotient with respect to *time* at the boundary of the S-grid. For the upper boundary, this gives:

$$\begin{aligned}
\frac{\partial V(S_M, t_j)}{\partial t} &\approx A_{M,j}\, W_{M-1,\ j} + B_{M,j}\, W_{M,j} \\
&\quad + C_{M,j} \left\{ \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2 W_{M-1,j} + \left[1 - \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2\right] W_{M,j} \right. \\
&\quad \left. + \frac{S_{M+1} - S_M}{S_M - S_{M-1}}(S_{M+1} - S_{M-1}) R_j^U \right\} \\
&= A^*_{M,j}\, W_{M-1,j} + B^*_{M,j}\, W_{M,j} \\
&\quad + \frac{S_{M+1} - S_M}{S_M - S_{M-1}}(S_{M+1} - S_{M-1}) R_j^U\, C_{M,j}
\end{aligned}$$

where

$$A^*_{M,j} = A_{M,j} + \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2 C_{M,j}$$

$$B^*_{M,j} = B_{M,j} + \left[1 - \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}}\right)^2\right] C_{M,j}$$

Likewise for the lower boundary:

$$\begin{aligned}\frac{\partial V(S_0,t_j)}{\partial t} &\approx A_{0,j}\,\{\left[1-\left(\frac{S_0-S_{-1}}{S_1-S_0}\right)^2\right]W_{0,\,j}+\left(\frac{S_0-S_{-1}}{S_1-S_0}\right)^2 W_{1,j}\\ &\quad-\frac{S_0-S_{-1}}{S_1-S_0}(S_1-S_{-1})R_j^L\}+B_{0,j}\,W_{0,\,j}+C_{0,j}\,W_{1,j}\\ &= B^*_{0,j}\,W_{0,j}+C^*_{0,j}\,W_{1,j}-\frac{S_0-S_{-1}}{S_1-S_0}(S_1-S_{-1})R_j^L A_{0,j}\end{aligned}$$

where

$$\begin{aligned}B^*_{0,j} &= B_{0,j}+\left[1-\left(\frac{S_0-S_{-1}}{S_1-S_0}\right)^2\right]A_{0,j}\\ C^*_{0,j} &= C_{0,j}+\left(\frac{S_0-S_{-1}}{S_1-S_0}\right)^2 A_{0,j}\end{aligned}$$

All preparations have now been made for specifying the difference scheme 10.14 at the boundaries. Replacing the differential quotient with respect to *time* in Equation 10.7 with the above approximation gives

$$\begin{aligned}&\theta A^*_{M,j}\,W_{M-1,j}+(\theta B^*_{M,j}+\frac{1}{t_{j+1}-t_j})\,W_{M,j}\\ &\approx\frac{W_{M,j+1}}{(t_{j+1}-t_j)}-(1-\theta)\left[A^*_{M,j+1}\,W_{M-1,j+1}+B^*_{M,j+1}\,W_{M,j+1}\right]\\ &-\frac{S_{M+1}-S_M}{S_M-S_{M-1}}(S_{M+1}-S_{M-1})\left[\theta R_j^U\,C_{M,j}+(1-\theta)R^U_{j+1}\,C_{M,j+1}\right]\end{aligned}$$

for the upper boundary. Likewise, the difference scheme for the lower boundary is obtained as

$$\begin{aligned}&(\theta B^*_{0,j}+\frac{1}{t_{j+1}-t_j})\,W_{0,j}+\theta C^*_{0,\,j}\,W_{1,j}\\ &\approx\frac{W_{0,j+1}}{(t_{j+1}-t_j)}-(1-\theta)\left[B^*_{0,j+1}\,W_{0,j+1}+C^*_{0,j+1}\,W_{1,j+1}\right]\\ &+\frac{S_0-S_{-1}}{S_1-S_0}(S_1-S_{-1})\left[\theta R_j^L A_{0,j}+(1-\theta)R^L_{j+1}A_{0,j+1}\right]\end{aligned}$$

Combining the above results, the finite difference scheme 10.14 for Neumann boundary

conditions can be written in the following matrix form:

$$\underbrace{\begin{pmatrix} b_0 & c_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ a_1 & b_1 & c_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & a_2 & b_2 & c_2 & 0 & 0 & \dots & 0 \\ 0 & 0 & a_3 & b_3 & c_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \ddots & & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & 0 & a_{M-1} & b_{M-1} & c_{M-1} \\ 0 & \dots & \dots & \dots & \dots & 0 & a_M & b_M \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} W_{0,j} \\ W_{1,j} \\ W_{2,j} \\ \\ \\ \\ \\ W_{M,j} \end{pmatrix}}_{\mathbf{W}_j} = \underbrace{\begin{pmatrix} d_{0,j} \\ d_{1,j} \\ d_{2,j} \\ \\ \\ \\ \\ d_{M,j} \end{pmatrix}}_{\mathbf{d}_j} \tag{10.19}$$

where

$$\begin{aligned}
a_i &= \theta A_{i,j} \ , \quad b_i = \theta B_{i,j} + \frac{1}{t_{j+1} - t_j} \ , \quad c_i = \theta C_{i,j} \quad \text{for} \quad i = 1, \dots, M-1 \\
d_{i,j} &= \frac{W_{i,j+1}}{(t_{j+1} - t_j)} \qquad\qquad\qquad\qquad\qquad\qquad \text{for} \quad i = 1, \dots, M-1 \\
&\quad - (1-\theta) \left[A_{i,\ j+1} W_{i-1,j+1} + B_{i,j+1} W_{i,j+1} + C_{i,j+1} W_{i+1,j+1} \right] \\
a_M &= \theta A^*_{M,j} \ , \quad b_M = \theta B^*_{M,j} + \frac{1}{t_{j+1} - t_j}, \\
d_{M,j} &= \frac{W_{M,j+1}}{(t_{j+1} - t_j)} - (1-\theta) \left[A^*_{M,\ j+1} W_{M-1,j+1} + B^*_{M,j+1} W_{M,j+1} \right] \\
&\quad - \frac{S_{M+1} - S_M}{S_M - S_{M-1}} (S_{M+1} - S_{M-1}) \left[\theta R^U_j \, C_{M,j} + (1-\theta) R^U_{j+1} \, C_{M,j+1} \right] \\
b_0 &= \theta B^*_{0,j} + \frac{1}{t_{j+1} - t_j} \ , \quad c_0 = \theta C^*_{0,j} \\
d_{0,j} &= \frac{W_{0,j+1}}{(t_{j+1} - t_j)} - (1-\theta) \left[B^*_{0,\ j+1} W_{0,j+1} + C^*_{0,j+1} W_{1,j+1} \right] \\
&\quad + \frac{S_0 - S_{-1}}{S_1 - S_0} (S_1 - S_{-1}) \left[\theta R^L_j A_{0,j} + (1-\theta) R^L_{j+1} A_{0,j+1} \right]
\end{aligned} \tag{10.20}$$

This system of equations must be solved for each time step taking the boundary conditions into consideration. The right-hand side of Equation 10.19 consists only of values evaluated at time point t_{j+1} (which have already been computed in the previous step) and of the (given) boundary conditions. The $d_{i,j}$ are thus completely determined. The system of equations has the same structure as in the corresponding system 10.16 for the Dirichlet problem except that it is two dimensions larger. Here, $\mathbf{A}$ is an $(M+1) \times (M+1)$ matrix. This is explained by the fact that the grid has been extended at the upper and lower boundary in the S-direction. The simple structure of the matrix $\mathbf{A}$ again allows the performance of a L-U decomposition. As before, we begin by setting $\mathbf{A}\, W_j = \mathbf{L}\, \mathbf{U}\, W_j =: \mathbf{L}\mathbf{x}$ and, exactly as was the case for Dirichlet

boundary conditions, we obtain

$$\begin{aligned}
u_i &= c_i && \text{for } i = 0, \dots M-1 \\
h_0 &= b_0 \\
l_i &= \frac{a_i}{h_{i-1}} && \text{for } i = 1, \dots M \\
h_i &= b_i - l_i\, u_{i-1} && \text{for } i = 1, \dots M \\
x_0 &= d_{0,j} \\
x_i &= d_{i,j} - l_i\, x_{i-1} && \text{for } i = 1 \dots M \\
W_{M,j} &= \frac{x_M}{h_M} \\
W_{i,j} &= \frac{x_i - u_i W_{i+1,j}}{h_i} && \text{for } i = 0 \dots M-1
\end{aligned} \tag{10.21}$$

The only difference is that the range of the index i has been extended and that a_i, b_i, c_i, d_i are now given by Equation 10.20.

As an example of the Neumann problem, we consider the plain vanilla put and call. The Neumann boundary conditions are usually easier to determine than the Dirichlet conditions. For S_0 the call is well out of the money and the change in the price with respect to S can be assumed to be negligible. The corresponding boundary condition is thus $\frac{\partial V(S,t)}{\partial S}\Big|_{S=S_0} = 0$. For extremely large S, the call is way in the money and thus its behavior is approximated by the payoff profile. The derivative in the S direction is thus 1 with the corresponding boundary condition given by $\frac{\partial V(S,t)}{\partial S}\Big|_{S=S_0} = 1$. The behavior of the put is the exact opposite. Summarizing, the Neumann boundary conditions for the plain vanilla put and call are given by :

	$S_0 \approx 0$	$S_M \approx \infty$	$t_N = T$
Call	$R_j^L = 0$	$R_j^U = 1$	$P_i = \max(S_i - K, 0)$
Put	$R_j^L = -1$	$R_j^U = 0$	$P_i = \max(K - S_i, 0)$

For the sake of completeness, we have also included the initial conditions in the table.

The type of boundary condition selected for the analysis depends on the type of option being priced. For plain vanilla options, Neumann boundary conditions are attractive because they can be easily calculated. The pricing of a forward contract at each time point (which must be done to establish the Dirichlet boundary conditions), on the other hand, can be tedious, particularly when the term structure is not flat and discrete dividends must be taken into consideration.

For Knock-out barrier options, however, the exact value of the option at the barrier is known (0 or the rebate), while the change of its value at the boundary is unknown. In this case, it makes sense to use the Dirichlet boundary conditions. In addition, a combination of Dirichlet and Neumann boundary conditions can be applied to one and the same problem as needed.

10.2.4 Unspecified Boundary Conditions

In many cases, finite difference methods can be used to calculate the price of derivatives even *without* information on the S-boundary conditions, i.e. even when only an initial condition is available. In such cases, the grid must be constructed in such a way that S takes on values from $S_0 = 0$ to $S_M \to \infty$. "Infinity" should be understood as "as large as possible". For these special boundary values S, the following general boundary conditions can be given which hold for many options. Since no option values have been specified at the boundaries (yet), the dimension of the problem is at present $(M + 1)$, the range of i thus $i = 0, \ldots, M$.

First consider the lower boundary $i = 0$. Equation 10.13 with $S_0 = 0$ yields immediately for the coefficients

$$A_{0,j} = 0,\ B_{0,j} = r_j,\ C_{0,j} = 0$$

Since $A_{0,j} = 0$, the values "$W_{0-1,j}$" defined outside the grid do not appear in the finite difference scheme of Equation 10.14. Substituting these very simple coefficients immediately gives the difference scheme at the boundary $i = 0$ as

$$W_{0,j} \approx \frac{1 - (t_{j+1} - t_j)(1 - \theta)r_{j+1}}{1 + (t_{j+1} - t_j)\theta r_j} W_{0,j+1}$$

This means that at $S = 0$ an explicit option value can always be established, even when it is not specified as a boundary condition! The earlier value of W is simply the later value discounted, where linear compounding is used for the explicit method ($\theta = 0$), while simple compounding is used for the implicit method ($\theta = 1$). The general case is a mixed form of these two compounding conventions. Iterating up to $j + 1 = N$ we obtain at $S = 0$ (for all times t_j) the explicit value of the option as the discounted payoff profile:

$$W_{0,j} \approx P_0 \prod_{k=j}^{N-1} \frac{1 - (t_{k+1} - t_k)(1 - \theta)r_{k+1}}{1 + (t_{k+1} - t_k)\theta r_k} \quad \forall j = 0, \ldots, N - 1 \tag{10.22}$$

where $P_0 = P(S = 0)$ denotes the value of the payoff profile for $S = 0$ at $t = T$.

This explicit option price at the boundary can now be inserted into the difference scheme just as was done for the Dirichlet boundary conditions. The vector $W_{i,j}$ is decomposed into a known vector and an unknown vector. The known vector has only one non-zero element namely the boundary condition at $S = 0$. Thus, we have reduced the dimension of the problem to M and the range of the index to $i = 1, \ldots, M$.

For $i = M$, i.e. for very large values of S we can assume that the option under investigation is either well in or well out of the money. The option price thus behaves, in good approximation as a function of S, like its payoff profile $P(S)$ which is a known function. In particular, the second derivative $\frac{\partial^2 V}{\partial S^2}$ is a known function, being the second derivative of the payoff profile. For all (reasonable) traded options it is a safe assumption that the second derivative of the payoff profile with respect to S is zero[6]

[6]Only for exotic power options with a payoff profile $\sim S^n$ with $n > 1$ is this theoretically not the case. But in reality, all *traded* power options have a cap.

for large S. This information can be utilized to express the value $W_{M+1,j}$ in terms of values defined on the grid through Equation 10.10. This procedure is analogous to that for Neumann boundary conditions:

$$\begin{aligned}
0 &= \frac{\partial^2 V(S_M, t_j)}{\partial S^2} \\
&= \frac{2}{S_{M+1} - S_{M-1}} [\frac{W_{M-1,j}}{S_M - S_{M-1}} \\
&- \left(\frac{1}{S_M - S_{M-1}} + \frac{1}{S_{M+1} - S_M} \right) W_{M,j} + \frac{W_{M+1,j}}{S_{M+1} - S_M}] \\
&\Longrightarrow \\
W_{M+1,j} &= -\frac{S_{M+1} - S_M}{S_M - S_{M-1}} W_{M-1,j} + \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}} + 1 \right) W_{M,j}
\end{aligned}$$

Using this in the approximation 10.12 for the differential quotient with respect to t yields

$$\begin{aligned}
\frac{\partial V(S_M, t_j)}{\partial t} &\approx A_{M,j} W_{M-1,\ j} + B_{M,j} W_{M,j} \\
&+ C_{M,j} \left[-\frac{S_{M+1} - S_M}{S_M - S_{M-1}} W_{M-1,j} + \left(\frac{S_{M+1} - S_M}{S_M - S_{M-1}} + 1 \right) W_{M,j} \right] \\
&= A^*_{M,j} W_{M-1,j} + B^*_{M,j} W_{M,j}
\end{aligned}$$

where

$$\begin{aligned}
A^*_{M,j} &= A_{M,j} - \frac{S_{M+1} - S_M}{S_M - S_{M-1}} C_{M,j} \\
B^*_{M,j} &= B_{M,j} + \left(1 + \frac{S_{M+1} - S_M}{S_M - S_{M-1}} \right) C_{M,j}
\end{aligned}$$

The difference scheme 10.14 on the boundary $i = M$ thus becomes

$$\begin{aligned}
&\theta A^*_{M,j} W_{M-1,j} + (\theta B^*_{M,j} + \frac{1}{t_{j+1} - t_j}) W_{M,j} \\
&\approx \frac{W_{M,j+1}}{(t_{j+1} - t_j)} - (1 - \theta) \left[A^*_{M,j+1} W_{M-1,j+1} + B^*_{M,j+1}\ W_{M,j+1} \right]
\end{aligned}$$

Including all this information allows the finite difference scheme to be represented as a system of equations which, at $i = 1$, has the same form as the corresponding system 10.16 for the Dirichlet boundary condition and at $i = M$ is similar to Equation 10.19

for the Neumann boundary condition.

$$\underbrace{\begin{pmatrix} b_1 & c_1 & 0 & 0 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & \dots & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & a_{M-1} & b_{M-1} & c_{M-1} \\ 0 & \dots & \dots & \dots & 0 & a_M & b_M \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ W_{M,j} \end{pmatrix}}_{\mathbf{W}_j} = \underbrace{\begin{pmatrix} d_{1,j} \\ d_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ d_{M,j} \end{pmatrix}}_{\mathbf{d}_j}$$

The coefficients are now given by:

$$\begin{aligned}
a_i &= \theta A_{i,j} \ , \quad b_i = \theta B_{i,j} + \frac{1}{t_{j+1} - t_j} \ , \quad c_i = \theta C_{i,j} \qquad \text{für} \ \ i = 1, \dots, M-1 \\
d_{i,j} &= \frac{W_{i,j+1}}{(t_{j+1} - t_j)} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{für} \ \ i = 2, \dots, M-1 \\
&\quad - (1-\theta) \left[A_{i,\ j+1} W_{i-1,j+1} + B_{i,j+1} W_{i,j+1} + C_{i,j+1} W_{i+1,j+1} \right] \\
a_M &= \theta A^*_{M,j} \ , \quad b_M = \theta B^*_{M,j} + \frac{1}{t_{j+1} - t_j} \\
d_{1,j} &= \frac{W_{1,j+1}}{(t_{j+1} - t_j)} \\
&\quad - (1-\theta) \left[A_{1,\ j+1} W_{0,j+1} + B_{1,j+1} W_{1,j+1} + C_{1,j+1} W_{2,j+1} \right] \ - \ a_{1,j} W_{0,j} \\
&= \frac{W_{1,j+1}}{(t_{j+1} - t_j)} - (1-\theta) \left[B_{1,j+1} W_{1,\ j+1} + C_{1,j+1} W_{2,j+1} \right] \\
&\quad - \left[(1-\theta) A_{1,j+1} + \theta A_{1,j} \frac{1 - (t_{j+1} - t_j)(1-\theta) r_{j+1}}{1 + (t_{j+1} - t_j) \theta r_j} \right] W_{0,j+1} \\
d_{M,j} &= \frac{W_{M,j+1}}{(t_{j+1} - t_j)} - (1-\theta) \left[A^*_{M,\ j+1} W_{M-1,j+1} + B^*_{M,j+1} W_{M,j+1} \right] \qquad (10.23)
\end{aligned}$$

The right-hand side depends only on values at time point t_{j+1} which have been computed in the previous step. Thus the $d_{i,j}$ are completely determined. Again, the simple structure of the matrix $\mathbf{A}$ allows an L-U decomposition. We begin by setting

$\mathbf{A}\,W_j = \mathbf{L}\,\mathbf{U}\,W_j =: \mathbf{Lx}$ and proceed as already shown above to obtain

$$\begin{aligned}
u_i &= c_i && \text{für } i = 1, \dots M-1 \\
h_1 &= b_1 \\
l_i &= \frac{a_i}{h_{i-1}} && \text{für } i = 2, \dots M \\
h_i &= b_i - l_i\, u_{i-1} && \text{für } i = 2, \dots M \\
x_1 &= d_{1,j} \\
x_i &= d_{i,j} - l_i\, x_{i-1} && \text{für } i = 2 \dots M \\
W_{M,j} &= \frac{x_M}{h_M} \\
W_{i,j} &= \frac{x_i - u_i W_{i+1,j}}{h_i} && \text{für } i = 1 \dots M-1
\end{aligned} \tag{10.24}$$

10.2.5 Free Boundary Conditions for American Options

Thus far our discussion has been restricted to European options. Finite difference methods can also be extended to American options. The right to early exercise must, of course, be taken into account in pricing the option. The Black-Scholes equation holds only on a restricted set and must be replaced by the two (in)equalities given by Equations 8.13 and 8.14. Equivalently expressed, the Black-Scholes equati*on* 8.13 holds on a parameter set with a free boundary given by $S^*(t)$. This line $S^*(t)$ separates the parameter set where the Black-Scholes *equation* holds from the parameter set where the value of the option is simply given by the payoff profile 8.14. Directly on this curve $S^*(t)$ we also have

$$V(S^*(t), t) = P(S^*(t))$$

This is a Dirichlet boundary condition. The boundary condition is called *free* because $S^*(t)$ is not known a priori.

A simple solution for this problem can be found using the *explicit* finite difference procedure. As for the European option, we start by determining a solution vector $W_{i,j}$, $i = 0, 1, \dots, M$ for a time step j. An intermediate calculation is performed to determine the solution vector $\widetilde{W}_{i,j}$ for the American option. In this calculation, the European price is determined and compared to the intrinsic value $P(S)$. The value of the American option is then defined as the larger of the two, i.e.

$$\widetilde{W}_{i,j} = \max\left[W_{i,j}, P(S_i)\right]$$

The procedure continues with the next time step using this solution vector.

If an implicit (or mixed) finite difference procedure is utilized, this will *not* be exact since the vector $W_{i,j}$ will already have been used implicitly in its calculation. Therefore the procedure can in principle not be split into the above two steps. It can be shown, however, that the error in doing so remains small if the grid in t is fine enough. In particular, if the distance between two time points is one day or less, the error is often negligible. An additional trick can be used for call options. It is known that the exercise of a call option can only be optimal (if at all) immediately before

the payment of a discrete dividend. We can thus introduce extra time points on and immediately proceeding the due dates of the dividend payments with a time difference of, for example, half a day, and thus minimize the error.

In any case, in addition to the free boundary conditions, values for the upper and lower boundaries of the S-grid must also be specified. This is already clear from the explicit procedure described above. We can obviously only compare the solution vector for a European option with the payoff profile if this vector has already been determined. To determine this solution vector for the European option, we need two boundary conditions. They can be either Neumann, Dirichlet, or more general boundary conditions (such as the second derivative equals zero). For an American option all boundary conditions involving derivatives are generally the same as for the corresponding European option. The maximum of the payoff profile and the boundary condition of the corresponding European option is a good candidate for the Dirichlet boundary condition. Thus, for American plain vanilla put and call options:

	Call
$S_0 \approx 0$	$R_j^L = 0$
$S_M \approx \infty$	$R_j^U = \max\left[S_M e^{-q(T-t_j)} - Ke^{-r(T-t_j)}, S_M - K\right]$
$t_N = T$	$P_i = \max(S_i - K, 0)$

	Put
$S_0 \approx 0$	$R_j^L = \max\left[-S_0 e^{-q(T-t_j)} + Ke^{-r(T-t_j)}, -S_0 + K\right]$
$S_M \approx \infty$	$R_j^U = 0$
$t_N = T$	$P_i = \max(K - S_i, 0)$

Note that for American options this is not enough. We also need to specify the conditions on the free boundary $S^*(t)$. And before we can do this, we need to determine $S^*(t)$ in the first place.

The Lamberton and Lapeyre Procedure

In many cases, the free boundary lies on only one side of the grid. For this case, a suitable procedure for evaluating the American put options has been suggested by *Lamberton* and *Lapeyre* [113]. This procedure does not necessitate any appreciable additional computational effort. It simply varies the order of operations needed to solve the system of equations for one time step. We will now present this procedure.

The Dirichlet boundary condition for the plain vanilla option as given above will be used as the specified boundary condition (other types of boundary conditions could also have been used). The difference scheme then has the form given in Equation 10.16 with coefficients given by Equation 10.17. Before introducing the procedure, we consider the following limiting cases: the exercise region for a put option must lie in regions where the value of S is small since for large S (for $S > K$) the intrinsic value of a put is zero; the option only has a time value. For calls (if at all), the exercise region must lie in regions where S is large since for small S (for $S < K$), the intrinsic value of the call is zero.

We now consider the iteration scheme 10.18. In this scheme, we start by determining the value of $W_{M-1,j}$. This is an option value for large S. From the above consideration

we know that an optimal early exercise of an American *call* option cannot happen in the small S region, but *could* happen in the large S region. Therefore we have to allow for this possibility at the grid point (S_{M-1}, t_j). Should it indeed be optimal to exercise in this time step then the value $W_{M-1,j}$ must be replaced by the intrinsic value. This fact can be incorporated into our scheme using the expression

$$\widetilde{W}_{M-1,j} = \max\left[\frac{x_{M-1}}{h_{M-1}}, S_{M-1} - K\right]$$

for our calculation of $\widetilde{W}_{M-1,j}$ instead of 10.18. We thus obtain the correct boundary value $\widetilde{W}_{M-1,j}$ for the American call option. In the next step we use

$$\widetilde{W}_{M-2,j} = \max\left[\frac{x_{M-2} - u_{M-2}\widetilde{W}_{M-1,j}}{h_{M-2}}, S_{M-2} - K\right]$$

instead of the corresponding expression 10.18 for the European option. Likewise, this leads to the correct result since the value $\widetilde{W}_{M-1,j}$ for an American option was calculated correctly before. This means, however, that an American call option is correctly computed even with implicit or mixed methods if we simply substitute

$$\widetilde{W}_{i,j} = \max\left[\frac{x_i - u_i\widetilde{W}_{i+1,j}}{h_i}, S_i - K\right]$$

for $i = M-2, M-3, \ldots, 2, 1$ at every time step. The procedure just described started with large S (Index $i = M-1$) and computes step-wise to smaller values of S. This makes it only suitable for *call* options.

The procedure must be modified for *puts*. Instead of an **LU** decomposition we now carry out a **UL** decomposition, for which the upper and lower triangular matrices appear in reverse order. The decomposition of the coefficient matrix $\mathbf{A}$ is then given by:

$$\mathbf{A} = \underbrace{\begin{pmatrix} h_1 & u_1 & 0 & \ldots & \ldots & 0 \\ 0 & h_2 & u_2 & 0 & \ldots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & \ddots & u_{M-2} \\ 0 & \cdots & \cdots & \cdots & 0 & h_{M-1} \end{pmatrix}}_{\mathbf{U}} \underbrace{\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ l_2 & 1 & 0 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & l_{M-2} & 1 & 0 \\ 0 & \cdots & \cdots & 0 & l_{M-1} & 1 \end{pmatrix}}_{\mathbf{L}}.$$

The matrix elements differ only slightly for those given by the **LU** decomposition:

$$\begin{aligned} u_i &= c_i && \text{for } \; i = 1, \ldots M-2 \\ h_{M-1} &= b_{M-1} && \\ l_i &= \frac{a_i}{h_i} && \text{for } \; i = 2, \ldots M-1 \\ h_i &= b_i - l_{i+1}\, u_i && \text{for } \; i = 1, \ldots M-2 \end{aligned}$$

The decisive difference is that h_i is iteratively determined from large values down to small ones (instead of small to large). Proceeding in this way, the matrix elements of $\mathbf{U}$ and $\mathbf{L}$ can be easily calculated. Analogously to the case of the $\mathbf{LU}$ decomposition we set

$$\mathbf{A}\,\mathbf{W}_j = \mathbf{U}\underbrace{\mathbf{L}\,\mathbf{W}_j}_{\mathbf{x}} =: \mathbf{U}\mathbf{x}$$

and first solve the system of equations for the new vector $\mathbf{x}$:

$$\begin{pmatrix} h_1 & u_1 & 0 & \dots & \dots & 0 \\ 0 & h_2 & u_2 & 0 & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & 0 & \ddots & u_{M-2} \\ 0 & \dots & \dots & \dots & 0 & h_{M-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_{M-1} \end{pmatrix} = \begin{pmatrix} d_{1,j} \\ d_{2,j} \\ d_{3,j} \\ \vdots \\ \vdots \\ d_{M-2,j} \\ d_{M-1,j} \end{pmatrix} .$$

To compute $\mathbf{x}$ we start with the row in the matrix which contains only one non-zero element, i.e. the last row, and proceed from bottom to top to obtain

$$\begin{aligned} x_{M-1} &= \frac{d_{M-1,j}}{h_{M-1}} \\ x_i &= \frac{d_{i,j} - u_i\, x_{i+1}}{h_i} \quad \text{für} \quad i = 1 \dots M-2 \,. \end{aligned}$$

Now that $\mathbf{x}$ is known, the desired option values $W_{i,j}$ can be calculated from the definition of $\mathbf{x}$, solving the equation $\mathbf{L}\,W_j = \mathbf{x}$:

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ l_2 & 1 & 0 & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & l_{M-2} & 1 & 0 \\ 0 & \cdots & \cdots & 0 & l_{M-1} & 1 \end{pmatrix} \begin{pmatrix} W_{1,j} \\ W_{2,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,j} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_{M-1} \end{pmatrix} .$$

To "invert" the matrix $\mathbf{L}$, we start with the row containing only one non-zero element, working from top to bottom to obtain

$$\begin{aligned} W_{1,j} &= x_1 \\ W_{i,j} &= x_i - l_i W_{i-1,j} \quad \text{für} \quad i = 2 \dots M-1 \,. \end{aligned}$$

Now, the iteration begins with the small index values, i.e. with small S-values. This is what we need since from the limit considerations above we know that optimal early exercise of an American put option can only happen in the small S region. Therefore

Scheme	θ	Numerical Effort	Convergence	Stability
Implicit	1.0	Large	Slow	Unlimited
Explicit	0.0	Minor	Slow	Limited
Crank-Nicolson	0.5	Large	Fast	Unlimited

Table 10.1: Properties of the three most commonly used finite difference methods

we have to start in this region to "capture" the early exercise possibilities. This is completely analogous to the case of an American call presented above, where we also started in the region where early exercise could be possible (it was the large S region in that case).

Consequently, we now obtain correct values of American put options for implicit and mixed methods by setting:

$$\widetilde{W}_{1,j} = \max\left[x_1, S_1 - K\right]$$
$$\widetilde{W}_{i,j} = \max\left[x_i - l_i W_{i-1,j}\,,\, S_i - K\right] \quad \text{for} \quad i = 2 \ldots M-1\,.$$

10.3 Convergence Criteria

As for all numerical methods, the essential question to be answered before implementing finite differences is whether the procedure is stable, i.e. whether the numerical solution in fact converges towards the actual solution. To motivate this complex subject, we first consider as an example the special case $\theta = 0$, i.e. the *explicit* method. As can be seen from Equations 10.17, 10.20 and 10.23, the off-diagonal terms are zero regardless of the type of boundary conditions used. This implies that the system of equations is completely uncoupled. For example, if no explicit boundary conditions were specified, the system of equations reduces to

$$\begin{aligned} W_{i,j} &= \left(1 - (t_{j+1} - t_j) B_{i,j+1}\right) W_{i,\ j+1} \qquad \text{for} \ \ i = 1 \ldots M-1 \\ &\quad - (t_{j+1} - t_j) A_{i,j+1}\, W_{i-1,j+1} - (t_{j+1} - t_j) C_{i,j+1}\ W_{i+1,j+1} \\ W_{M,j} &= \left(1 - (t_{j+1} - t_j) B^*_{M,j+1}\right) W_{M,\ j+1} - (t_{j+1} - t_j) A^*_{M,j+1}\, W_{M-1,j+1} \end{aligned} \tag{10.25}$$

This means that we obtain the unknown values at time point t_j directly (without carrying out any matrix inversion!) in terms of the values already calculated for time point t_{j+1}. This result can also be obtained much simpler and more directly by starting with Equation 10.5 for the explicit method rather than from the more general expression 10.7.

At this point the reader could ask why we have taken the trouble of introducing the generalized expression involving θ when the explicit method is so simple. For this we have to understand how the parameter θ influences the properties of the finite difference scheme. An important criterion is stability. Table 10.1 contains an overview of the three most commonly used values of θ.

A difference scheme is called *stable* if small deviations from the correct solution do *not* grow arbitrarily fast (no faster than exponentially) with time[7]. If a difference scheme is not stable, errors in the approximation arising in an iteration step are amplified by calculating backwards in the time-grid from one iteration step to the next. Obviously, a difference scheme is only useful if it is stable, since deviations from the correct solution would otherwise grow faster than exponentially when calculating backwards through the grid, finally yielding a result with an arbitrarily large error. It can be shown that for $\theta \geq 0.5$, the finite difference scheme is always stable, independent of the choice of grid.

A difference scheme *converges* towards the correct solution if it is both *stable* and *consistent*. *Consistency* is satisfied if the difference scheme applied to a smooth function at a fixed time point gives an arbitrarily good approximation of the solution to the differential equation by making the grid fine enough. Consistency thus means "convergence for one time step". If the system is consistent and stable, i.e. the growth of the error over one iteration step is bounded, the scheme converges *entirely*, i.e. over all time steps.

There exists a criterion for the stability of an *explicit* difference scheme ($\theta = 0$) for the well-known heat equation 8.22 which, as we have seen, is closely related to the Black-Scholes equation. This criterion is (expressed in the variables from Equation 8.22):

$$\frac{\delta\tau}{\delta x^2} \leq \frac{1}{2}$$

For the Black-Scholes equation, the stability condition is considerably more complicated [154]:

$$\frac{1}{2}\left[(r(t)-q(t))S\frac{\delta t}{\delta S}\right]^2 \leq \frac{1}{2}\sigma(S,t)^2S^2\frac{\delta t}{(\delta S)^2} \leq \frac{1}{2} \tag{10.26}$$

We are, in fact, now dealing with two inequalities. The inequality appearing on the left in the above expression is generally satisfied for the usual values for r, q and σ. The second inequality however, poses a strong restriction in the application of the explicit method since it requires the number of time steps to increase quadratically with the number of the S-steps (i.e. as δS becomes smaller). A consequence of the S-dependence of this criterion is that the difference scheme can be stable on one part of the grid and unstable on the other. Such local instabilities often go unnoticed, but lead to incorrect results; for example, negative prices could arise. Such an example can be found in Hull [88]. The criterion 10.26 for stability can be generalized further: the generalized difference scheme for non-uniform grids remains stable for $\theta < 0.5$ when the following holds for the second inequality:

$$\frac{(t_{j+1}-t_j)}{(S_{i+1}-S_i)^2}\sigma_{i,j}^2S_i^2 \leq \frac{1}{2(1-2\theta)}\,, \quad \theta < \frac{1}{2}$$

The Crank-Nicolson method $\theta = 0.5$ has an advantage over the implicit and explicit methods in that it converges considerably faster. Because of the averaging of the finite differences at times t_j and t_{j+1}, this method is exact up to second order although

[7]Independent of the number of time steps.

the time derivative corresponds to only a first order approximation. Achieving the same degree of accuracy using the implicit method would require considerably more time steps. The Crank-Nicolson method loses some of its efficient convergence when the initial and/or boundary conditions become less smooth. In many applications, for instance, the first derivative of the initial condition is not continuous. This is already the case for the plain vanilla option (at the strike). However, the discontinuity at that point leads to a negligible inaccuracy in the option price. A significant error, however, can be found for the value of the delta, while the values for gamma at the point of the discontinuity can be so extreme that the error may exceed 100%. As a result, it may be the case that the gamma for the plain vanilla option cannot be computed with sufficient accuracy when the spot price equals the strike price (the first derivative of the payoff is discontinuous at this point). In such situations the implicit method is often the only suitable method for calculating gamma with sufficient accuracy. Another possibility would be to use a three-time-level approximation of the time derivative which also has the effect of smoothing the oscillations. A detailed description of this and other procedures can be found, for example, in the works of Smith [151] and Willmott [164].

10.3.1 Improving the Convergence Properties

If only few time steps are chosen, we observe that the results oscillate strongly with the number of time steps used [151]. Nonetheless, as few as 20 time steps are already sufficient to obtain stable results for a plain vanilla option with a time to maturity of six months. But it is usually worthwhile to increase the number of time steps, depending on the accuracy required and the available computation time. Note however, that the improvement in the approximation resulting from increasing the number of time steps is limited. Improving the approximation requires a refinement of both the time-grid as well as the S-grid. The proportion of the number of t-points to the number of S-points in the grid depends on the lifetime of the option, the required width of the S-grid (see below) and the type of difference scheme applied. The Crank-Nicolson scheme, for example, requires a significantly coarser t-grid to attain comparable accuracy to that obtained from using the implicit method with the otherwise same parameters. Experimenting with these parameters is in any case recommended.

In general, it makes sense to start with a *uniform* grid. An analysis of the solution surface, i.e. the price function at each point of the grid quickly reveals information about the time regions at which new grid points should be introduced. Figure 10.1 shows such a solution surface for an American call option with discrete dividend payments (see also the next section) approximately half way through its lifetime. Additional time points have been introduced at and immediately before the due date of the dividend payment. Doing so exactly incorporates the dividend payment into the calculation. We can clearly see where exercising shortly before the dividend payment is optimal (the jump occurs as a result). Likewise, it makes sense to introduce additional grid points in the time region shortly before the maturity of the option since there the curvature of the solution surface is quite large.

At every time point where an external shock such as a dividend payment occurs, the addition of grid points prevents the oscillation of the solution dependent on the number of t-grid points. This also holds for the S-grid at the points where the payoff

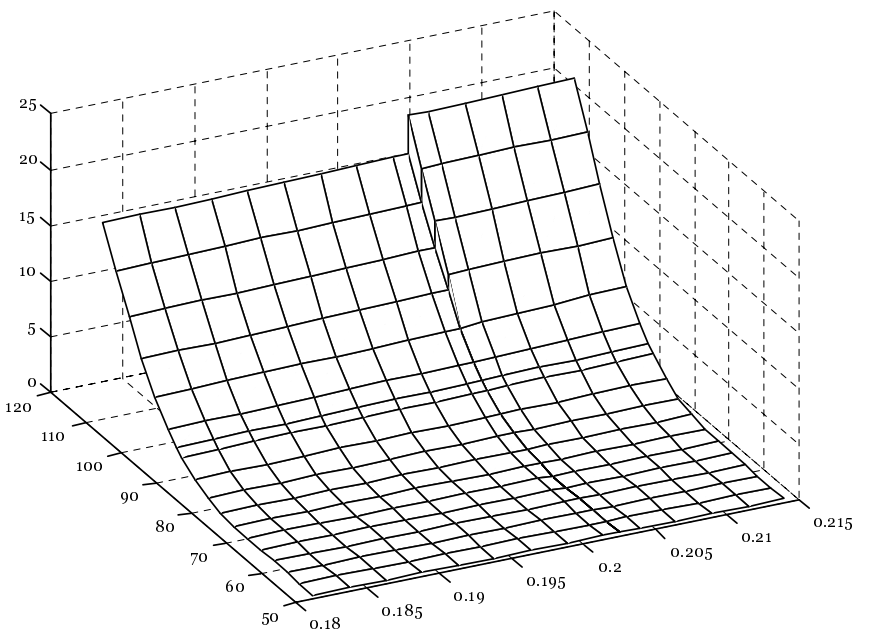


Figure 10.1: Part of the solution surface of an American plain vanilla call on an underlying with a discrete dividend payment.

function is discontinuous. For example, it makes sense to choose the strike price as a grid point, and also the underlying value S for which the option price is to be determined. Likewise the barrier of a barrier option should ideally be directly on the grid. However, a non-uniform grid can *reduce* accuracy, since, for example, the Crank-Nicolson scheme is exact up to second order only for a *uniform* grid in t. In practice serious consideration should thus be giving as to where and whether additional grid points should be introduced.

A good choice of whole S-region, i.e. of S_0 and S_{M+1} can substantially increase the accuracy attained. A rule of thumb is commonly used for choosing S_0 and S_{M+1}, such as

$$S_0 = \frac{1}{4}\min(S, K)$$
$$S_{M+1} = 4\max(S, K)$$

where here, S is the spot price of the underlying at time $t = 0$ and K the strike price of the option. The S-region thus obtained, however, (depending on the values of other parameters such as the volatility and the interest rate term structure) seldom leads to an optimally accuracy if the t–grid and the number of S–steps are assumed as given.

Alternatively, we could attempt to find the minimum (and maximum) value the underlying can attain in its lifetime *with a given probability*. This can be done if the underlying's cumulative probability distribution P has already been established. In the simplest cases, even an analytic formula for this probability can be derived (see Section A.4). In order to obtain minimal (maximal) S-values for a given probability (confidence) c, the expression $P(x \leq a) = c$ must be solved for a, the *percentile* of the underlying's distribution corresponding to the confidence level c. It is often worth the trouble to make this single calculation to establish the S-region since it can result in a more precise solution with substantially fewer S-steps.

10.4 Discrete Dividends

Consideration of discrete dividends represents a further complication. The particular difficulty is that the different methods available rely on different assumptions. One commonly used method is based on the separation of the stochastic and deterministic components of the stochastic process S. The dividend payments correspond to the deterministic component. One assumption is thus that the exact value of the dividend payment is known a priori. The present value D_t of the dividends which are to be paid until maturity of the option is subtracted from the spot rate S of the underlying:

$$\widetilde{S} = S - D_t$$

Instead of S we now use $\widetilde{S}$ as the new variable in the Black-Scholes equation. D_t is the nominal value of the dividends discounted back to t (where t denotes the time step currently under consideration). This process is consistent with the Black-Scholes formula for the pricing of a plain vanilla European option. There, discrete dividends are generally taken into consideration by subtracting the present value of the dividends from the spot price of the underlying. Consequently, the same volatility holds for both cases. Note however, that the *intrinsic* value of the option is is still given by

$$\max(S - X, 0) = \max(\widetilde{S} + D_t - X, 0)$$

This value is required to determine the free boundary condition for an early exercise of the option.

In the world of finite differences, we have an additional and fundamentally different method at our disposal. We first rewrite Equation 8.8 replacing the dividend term $q(t)S$ resulting from the dividend rate $q(t)$ with $D(S,t)$:

$$\frac{\partial V}{\partial t} + (r(t)S - D(S,t))\frac{\partial V}{\partial S} + \frac{1}{2}\sigma(S,t)S^2\frac{\partial^2 V}{\partial S^2} = r(t)V$$

The new term $D(S,t)$ represents the nominal value of the dividend payments at time t. If no dividends are paid at time t, the value of this function $D(S,t)$ is equal to zero. $D(S,t)$ is thus highly discontinuous. At time τ of a dividend payment, the option price must remain continuous, although the spot price of S is reduced by the amount given by the dividend payment D:

$$\lim_{\epsilon \to 0} V(S(\tau - \epsilon), \tau - \epsilon) = \lim_{\epsilon \to 0} V(S(\tau + \epsilon), \tau + \epsilon) \quad \epsilon > 0$$

where

$$\lim_{\epsilon \to 0} S(t-\epsilon) = S$$
$$\lim_{\epsilon \to 0} S(t+\epsilon) = S - D$$

This jump in S can be simulated with finite difference methods by first determining the solution vector at time τ. Subsequently, the vector is translated by D so that the continuity condition is satisfied. If necessary, missing intermediate values must be determined by interpolation, or in the case of boundary values, by extrapolation.

10.5 Example

A complete, executable example program written in Visual Basic can be found on the CD-ROM included with this book (see FiniteDifferenceMethod.xls). The program was not written for optimal performance, but was structured in such a way as to incorporate all the concepts introduced in this section. For example, the S-grid is recalculated for every new time step (which, of course, need not be done if the grid is uniform). The volatility, interest rates and payoff are separated into individual functions to clearly organize the points at which such structures could be loaded externally via interfaces. If, for example, the interest rate or volatility is constant, this effort is superfluous. The program is structured to be, for the most part, self-explanatory with helpful comments incorporated into the code. It can serve as a starting point for generating optimal-performance variants suitable for specific problems. This Excel workbook also demonstrates the calculation of option sensitivities, known as *Greeks*, which will be introduced in detail later in Chapter 13. Thus, this workbook can serve as a complete little option calculator (the yellow fields are input fields, a standard which applies to all workbooks on the CD-ROM).

In addition to the L-U decomposition introduced above, the example includes a further method for solving a system of equations, which is widely used in practice. The method is a very old procedure referred to as *Gaussian elimination*. Gaussian elimination is just as fast as the L-U decomposition and can be applied to both American and European options. The idea is even simpler than that of the L-U decomposition. Instead of decomposing the matrix in Equation 10.16 into the matrix product of an upper and lower triangular matrix, it is transformed into a single triangular matrix with non-zero terms appearing only in the diagonal and a single off-diagonal. There are two possible ways of doing this: the matrix can be transformed into an upper or lower triangular matrix by eliminating the lower off-diagonal or upper off-diagonal terms respectively.

The first element in the lower off-diagonal, a_2, is eliminated by multiplying the first row of the matrix by a_2/b_1, and subtracting the result from the second row. After doing so, the second row contains only two elements, namely $b_2' = b_1 - c_1 a_2/b_1$ in diagonal and c_2 in the upper off-diagonal. This procedure is repeated analogously by multiplying the second row by a_3/b_2' and subtracting from the third, and so on. In terms of the system of equations, this means that certain equations are multiplied by constants and subtracted from one another. These are the usual operations implemented when

solving a system of equations. In performing such operations, it is clear (in contrast to the L-U decomposition) that the right-hand side of the equation also changes. Carrying out this procedure for the entire matrix yields the following system of equations:

$$\begin{pmatrix} b'_1 & c_1 & 0 & \cdots & \cdots & 0 \\ 0 & b'_2 & c_2 & 0 & & \cdots \\ \vdots & 0 & b'_3 & c_3 & \ddots & \cdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \ddots & c_{M-2} \\ 0 & \cdots & \cdots & \cdots & 0 & b'_{M-1} \end{pmatrix} \begin{pmatrix} W_{1,\,j} \\ W_{2,\,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,\,j} \end{pmatrix} = \begin{pmatrix} d_{1,\,j}' \\ d_{2,\,j}' \\ \vdots \\ \vdots \\ d_{M-2,\,j}' \\ d_{M-1,\,j}' \end{pmatrix}$$

where

$$b'_1 = b_1 \quad , \quad d_{1,\,j}{}' = d_{1,\,j}$$
$$b'_i = b_i - c_{i-1}\frac{a_i}{b'_{i-1}} \quad , \quad d_{i,\,j}{}' = d_{i,\,j} - d_{i-1,\,j}{}'\frac{a_i}{b'_{i-1}} \quad , \quad \forall i = 2, \ldots, M-1$$

This system can be solved quite easily by starting with the row containing only one non-zero term, in this case the last row, and working upwards from the bottom. We obtain the option values

$$W_{M-1,\,j} = \frac{d'_{M-1,\,j}}{b'_{M-1}}$$
$$W_{i,\,j} = \frac{d_{i,\,j}{}' - c_i \;\; W_{i+1,\,j}}{b'_i} \qquad \forall i = 1, \ldots, M-2$$

To this extent, this method offers an alternative to the L-U decomposition for the valuation of European options. The free boundary condition for American options is dealt with, as in the method of Lamberton and Lapeyre, by calculating the value of the option by Gaussian elimination as if it were European and, at each time step, comparing it with the intrinsic value of the option $P(S)$, i.e. the payoff profile. The value of the American option at this time point is then defined as the greater of the two. The components of the solution vector $\widetilde{W}_{i,j}$ are thus given by

$$\widetilde{W}_{M-1,\,j} = \max\left[\frac{d'_{M-1,\,j}}{b'_{M-1}}, P(S_{M-1})\right]$$
$$\widetilde{W}_{i,\,j} = \max\left[\frac{d_{i,\,j}{}' - c_i \;\widetilde{W}_{i+1,\,j}}{b'_i}, P(S_i)\right] \qquad \forall i = 1, \ldots, M-2$$

As was already stressed earlier when we presented the method of *Lamberton and Lapeyre*, this procedure functions only when we calculates along the S-grid points starting from the exercise region, where the value of the option is given by its intrinsic value (the payoff profile), toward the region where the Black-Scholes equation holds. Otherwise we would not "capture" the early exercise possibilities. The Gaussian elimination procedure described above starts from large S (index $i = M-1$) and calculates

backwards step by step towards smaller values of S. This works fine for call options since early exercise of a call can only (if at all) be optimal for large S because for small S (for $S < K$) the intrinsic value of the call is of course zero.

For put options, however, the exercise region can only lie in areas where the value of S is small, since for large S (for $S > K$), the intrinsic value of the put is zero, the option itself having only a time value. Therefore the second possibility for performing a Gaussian elimination has to be used for puts, namely eliminating the upper off-diagonal. This results in a procedure which begins with small S proceeding step by step towards larger S-values. The upper off-diagonal is eliminated by multiplying the last row of the matrix by c_{M-2}/b_{M-1}, subtracting the result from the next to last row. Proceeding analogously, the newly created next to last row is multiplied by c_{M-3}/b'_{M-2} and subtracted from the row lying immediately above it, and so on. Performing this procedure for the entire matrix yields the following system of equations:

$$\begin{pmatrix} b'_1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_2 & b'_2 & 0 & & & \vdots \\ 0 & a_3 & b'_3 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & a_{M-2} & b'_{M-2} & 0 \\ 0 & \cdots & \cdots & 0 & a_{M-1} & b'_{M-1} \end{pmatrix} \begin{pmatrix} W_{1,\,j} \\ W_{2,\,j} \\ \vdots \\ \vdots \\ \vdots \\ W_{M-1,\,j} \end{pmatrix} = \begin{pmatrix} d'_{1,\,j} \\ d'_{2,\,j} \\ \vdots \\ \vdots \\ d'_{M-2,\,j} \\ d'_{M-1,\,j} \end{pmatrix}$$

where

$$b'_{M-1} = b_{M-1} \quad , \quad d'_{M-1,\,j} = d_{M-1,\,j}$$
$$b'_i = b_i - a_{i+1}\frac{c_i}{b'_{i+1}} \ , \quad d'_{i,\,j} = d_{i,\,j} - d'_{i+1,\,j}\frac{c_i}{b'_{i+1}} \quad , \quad \forall i = 1,\ldots, M-2$$

This system can be solved quite easily by starting with the row containing only one non-zero element, in this case the first row, and proceeding from top to bottom. We thus obtain the option values

$$W_{1,\,j} = \frac{d'_{1,\,j}}{b'_1}$$
$$W_{i,\,j} = \frac{d'_{i,\,j} - a_i\, W_{i-1,\,j}}{b'_i} \qquad \forall i = 2, \ldots, M-1$$

The components of the solution vector $\widetilde{W}_{i,j}$ for an American option with payoff $P(S)$ is given by

$$\widetilde{W}_{1,\,j} = \max\left[\frac{d'_{1,\,j}}{b'_1}, P(S_1)\right]$$
$$\widetilde{W}_{i,\,j} = \max\left[\frac{d'_{i,\,j} - a_i\, \widetilde{W}_{i-1,\,j}}{b'_i}, P(S_i)\right] \qquad \forall i = 2, \ldots, M-1$$

11

Binomial and Trinomial Trees

Binomial and *trinomial trees* are very popular tools commonly used in practice to calculate prices and sensitivity parameters of derivatives while avoiding direct reference to the fundamental differential equations governing the price of the instrument. These methods provide a useful alternative to those (numerical or analytical) methods presented in the previous sections for solving differential equations. In principle (ignoring for the moment the potential computing-time problems), binomial and trinomial trees can also be used in pricing path-dependent derivatives.

In addition to the usual assumptions when excluding arbitrage opportunities (Assumptions 1, 2, 3, 4 and 5 of Section 5), non-stochastic interest rates (Assumption 8) will generally be assumed in the subsequent sections. These assumptions allow a general theory of binomial trees to be presented. In order to actually calculate option prices, the underlying must be assumed to behave according to a model. Thus, from Section 11.3 onwards, it will be assumed that the underlying can be modeled as a random walk with non-stochastic volatility, i.e. the additional assumptions 7 and 10 from Section 5 will be made. As will be discussed in Section 11.3, obtaining recombining binomial trees requires the volatility to be constant over time, i.e. Assumption 11 must hold. Furthermore, we will assume that the underlying earns a dividend *yield* in accordance with Equation 3.5 rather than discrete dividend payments.

11.1 General Trees

11.1.1 Evolution of the Underlying and the Replicating Portfolio

Generally, in a tree procedure, the time span in question (the lifetime of the derivative) between t and T is divided into n time intervals of equal length dt:

$$T - t = n\,dt \tag{11.1}$$

In each such time interval the underlying price $S(t)$ may increase in value to $u\,S$ (with $u > 1$) with a probability p', or it may decrease in value to dS (with $d < 1$) with a

probability $(1-p')$

$$S(t) \to \left\langle \begin{array}{l} S_u(t+dt) = u(t)S(t) \text{ with probability } p' \\ S_d(t+dt) = d(t)S(t) \text{ with probability } 1-p' \end{array} \right. \tag{11.2}$$

After three steps, for example, the price can take on $2^3 = 8$ possible values:

$$S(t) \to \left\langle \begin{array}{l} u(t)S(t) \left\langle \begin{array}{l} u(t+dt)u(t)S(t) \left\langle \begin{array}{l} u(t+2dt)u(t+dt)u(t)S(t) \\ u(t+2dt)u(t+dt)u(t)S(t) \end{array} \right. \\ d(t+dt)u(t)S(t) \left\langle \begin{array}{l} u(t+2dt)d(t+dt)u(t)S(t) \\ u(t+2dt)d(t+dt)u(t)S(t) \end{array} \right. \end{array} \right. \\ d(t)S(t) \left\langle \begin{array}{l} u(t+dt)d(t)S(t) \left\langle \begin{array}{l} u(t+2dt)u(t+dt)d(t)S(t) \\ d(t+2dt)u(t+dt)d(t)S(t) \end{array} \right. \\ d(t+dt)d(t)S(t) \left\langle \begin{array}{l} u(t+2dt)d(t+dt)d(t)S(t) \\ d(t+2dt)d(t+dt)d(t)S(t) \end{array} \right. \end{array} \right. \end{array} \right.$$

Consider now a portfolio consisting of Δ underlyings and g monetary units in cash. If the dividend yield q earned in the time interval dt is paid, the cash is compounded at a risk-free rate r and the price of the underlying behaves as described above, the value of the portfolio after dt is given by

$$\Delta(t)S(t) + g(t) \to \left\langle \begin{array}{l} \Delta\,(t)u(t)S(t)B_q(t)^{-1} + g(t)B_r(t)^{-1} \\ \Delta\,(t)d(t)S(t)B_q(t)^{-1} + g(t)B_r(t)^{-1} \end{array} \right. \tag{11.3}$$

Here, for the sake of simplifying the notation for the discount factors over a small time interval dt we have defined

$$B_r(t) := B_r(t, t+dt) \quad , \quad B_q(t) := B_q(t, t+dt) \tag{11.4}$$

In the following sections, Δ and g will be chosen so that the price of this portfolio behaves exactly as does the value of the derivative we wish to price. This value of the portfolio will then replicate the value of the derivative at each time point and is thus referred to as the *replicating portfolio.*

11.1.2 Evolution of the Derivative

If the underlying moves in accordance with Equation 11.2 in the time interval dt, the price V of a derivative on this underlying evolves in accordance with

$$V(S,t) \to \left\langle \begin{array}{l} V(S_u, t+dt) \\ V(S_d, t+dt) \end{array} \right.$$

where V^u and V^d represent the value of a derivative[1] whose underlying has a price of S_u and S_d, respectively. Setting the value of the portfolio (Equation 11.3) equal to the value of the derivative at time t as well as after the next binomial step $t+dt$, we obtain the following three equations:

$$V(S,t) = \Delta(t)\,S(t) + g(t) \tag{11.5}$$

[1] In order to emphasize that this method is valid for all kinds of derivatives, we will continue to denote the value of the derivative with the letter V.

$$V(S_u, t+dt) = \Delta(t)\, u(t) S(t) B_q(t)^{-1} + g(t) B_r(t)^{-1} \tag{11.6}$$

$$V(S_d, t+dt) = \Delta(t)\, d(t) S(t) B_q(t)^{-1} + g(t) B_r(t)^{-1} \tag{11.7}$$

These can be easily rearranged[2] to express the number of underlyings, the cash amount and the value of the derivative at time t in terms of values at time $t + dt$:

$$\Delta(t) = \frac{V(S_u, t+dt) - V(S_d, t+dt)}{[u(t) - d(t)] S(t)/B_q(t)}$$

$$g(t) = \frac{u(t)\, V(S_d, t+dt) - d(t)\, V(S_u, t+dt)}{[u(t) - d(t)]/B_r(t)}$$

$$V(S,t) = B_r(t)\, [p(t) V(S_u, t+dt) + (1-p(t)) V(S_d, t+dt)] \tag{11.8}$$

where

$$p(t) = \frac{B_q(t)/B_r(t) - d(t)}{u(t) - d(t)} \tag{11.9}$$

In Equation 11.8, we have succeeded in expressing the unknown value of the derivative at time t in terms of quantities known at time t, namely $B_q(t)$, $B_r(t)$, $u(t)$ and $d(t)$, (in Section 14, we will show how the values of u and d are determined) and the (likewise unknown) derivative values at time $t + dt$. The reader might ask what this has accomplished. This expression will in deed prove to be useful if the value of the derivative at a future time is known. Such a future time is, for example, the maturity date T of the derivative. At this time, the value of the derivative as a function of the underlying price is given explicitly by its payoff profile, and as such, is known. The strategy is thus to repeat the procedure described above until reaching a time at which the value of the option is known (in most cases, maturity T). This procedure will be demonstrated below.

Equation 11.8 holds for European derivatives since it is implicitly assumed that the option still exists after a time step has been taken. In order to account for the possibility of exercising early as in the case of derivatives with American features, the derivative's value as given in Equation 11.8 is compared with its intrinsic value at each node in the tree. Then the larger of the two values is taken as the derivative price at that node. For instance for American calls and puts with payoff profiles $S(t) - K$ and $K - S(t)$, respectively, Equation 11.8 would be replaced by

$$\begin{aligned} &C_S(t) \\ &= \max\left\{ B_r(t) \left[p(t) C_S^u(t+dt) + (1-p(t)) C_S^d(t+dt) \right], S(t) - K \right\} \\ &P_S(t) \\ &= \max\left\{ B_r(t) \left[p(t) P_S^u(t+dt) + (1-p(t)) P_S^d(t+dt) \right], K - S(t) \right\} \end{aligned}$$

[2]Subtracting Equation 11.7 from Equation 11.6 yields

$$V_S^u - V_S^d = \Delta(t)\, (u-d) S(t) B_q(t)^{-1}$$

This allows us to isolate the $\Delta(t)$ term easily. Multiplying Equation 11.7 by u, and Equation 11.6 by d and subtracting the results yields

$$u\, V_S^d - d\, V_S^u = (u-d) g(t) B_r(t)^{-1}$$

This can be readily solved for $g(t)$. Substituting the expressions thus obtained for $\Delta(t)$ and $g(t)$ into Equation 11.5 yields the value of the derivative $V_S(t, T, K)$.

11.1.3 Forward Contracts

The evolution of the replicating portfolio consisting of underlyings and cash in a bank account is described by Equation 11.3. According to Equation 7.5, the evolution of a futures position is given by

$$V(S,t) \rightarrow \left\langle \begin{array}{l} V(S_u, t+dt) = S_u(t+dt,T) - S(t,T) \\ V(S_d, t+dt) = S_d(t+dt,T) - S(t,T) \end{array} \right.$$

Here, it is not the value of the future at time t which is unknown, (this is equal to zero since $K = S(t,T)$), but but the forward price of the underlying $S(t,T)$. Setting the portfolio equal to the future at both time t and at the next time in the binomial tree $t + dt$ yields three equations:

$$0 = V(S,t) = \Delta(t)\, S(t) + g(t) \tag{11.10}$$

$$S_u(t+dt,T) - S(t,T) = V(S_u, t+dt) = \Delta(t)\, u(t) S(t) B_q(t)^{-1} + g(t) B_r(t)^{-1} \tag{11.11}$$

$$S_d(t+dt,T) - S(t,T) = V(S_d, t+dt) = \Delta(t)\, d(t) S(t) B_q(t)^{-1} + g(t) B_r(t)^{-1} \tag{11.12}$$

With the help of these three equations, the number of underlyings, the money in the bank account and the forward price at time t can (making use of the expression p defined in Equation 11.9) be expressed[3] in terms of the forward price at time $t + dt$:

$$\begin{aligned} \Delta(t) &= \frac{S_u(t+dt,T) - S_d(t+dt,T)}{[u(t) - d(t)] S(t) B_q(t)} \\ g(t) &= -\frac{S_u(t+dt,T) - S_d(t+dt,T)}{[u(t) - d(t)] B_q(t)} \\ S(t,T) &= p(t) S_u(t+dt,T) + (1 - p(t)) S_d(t+dt,T) \end{aligned} \tag{11.13}$$

11.2 Recombinant Trees

11.2.1 The Underlying

If the parameters u and d are independent of time

$$u(t + j\,dt) = u(t) \equiv u \;\; \forall j \quad , \quad d(t + j\,dt) = d(t) \equiv d \;\; \forall j$$

then obviously $udS(t) = duS(t)$ holds, i.e. an upward move followed by a downward move results in the same underlying price as a downward move followed by an upward move. Thus the tree is forced to *recombine*. This significantly reduces the number of possible nodes, making the computation much more efficient. Such a *recombining binomial tree* has the form depicted in Figure 11.1. The probability for a *single* path

[3] Subtracting Equation 11.11 from Equation 11.12 yields

$$S^u(t+dt,T) - S^d(t+dt,T) = \Delta(t)\,(u-d) S(t) B_q(t)^{-1}$$

This allows us to easily isolate $\Delta(t)$. Because of Equation 11.10, $g(t) = -\Delta(t)S(t)$ holds, which immediately yields $g(t)$ if $\Delta(t)$ is known. Substituting the expressions for $\Delta(t)$ and $g(t)$ into Equation 11.11 or Equation 11.12 yields, after a simple calculation, the forward price $S(t,T)$.

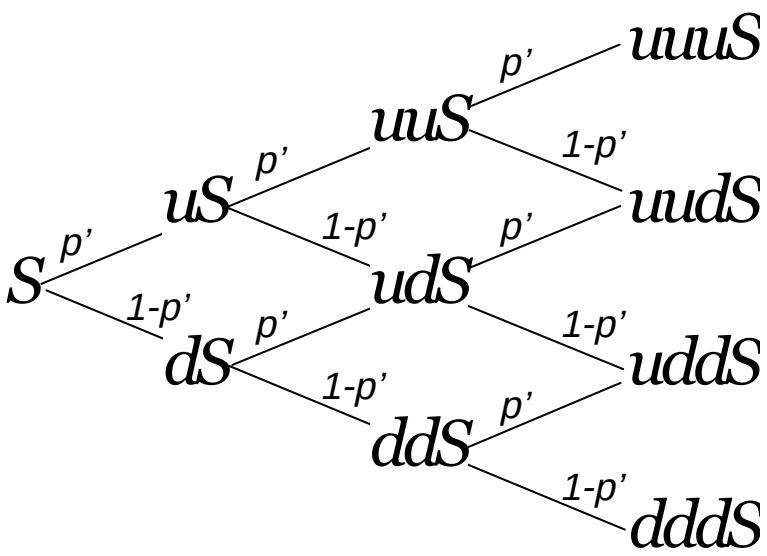


Figure 11.1: The first steps in a recombining binomial tree.

ending at $S(T) = u^j d^{n-j} S(t)$ is

$$p'^j (1-p')^{n-j}$$

The number of all paths ending at $S(T) = u^j d^{n-j} S(t)$ can be deduced from permutation laws and is given by the *binomial coefficient*

$$\binom{n}{j} \equiv \frac{n!}{j!(n-j)!}$$

The probability in the above tree of arriving at a value $S(T) = u^j d^{n-j} S(t)$ regardless of the path taken to get there is equal to the number of such paths multiplied by the probability of realizing such a path.

$$P\left[S(T) = u^j d^{n-j} S(t)\right] = \binom{n}{j} p'^j (1-p')^{n-j} = B_{n,p'}(j) \tag{11.14}$$

This is the definition of the probability density function of the *binomial distribution*, see Section A.4.2. This is how the binomial distribution enters into the binomial trees. As can be seen from the above derivation, a necessary condition for arriving at this distribution is that the parameters u and d be constant over time.

11.2.2 The Binomial Distribution for European Derivatives

In addition to the assumptions made at the beginning of Chapter 11 and the one just made, namely that the parameters u and d are constant over time, we will henceforth

assume that the yields (interest rates and dividends) are constant over time as well, i.e. that Assumptions 9 and 12 from Section 5 hold[4]:

$$B_r(t + j\,dt) = B_r(t)\ \forall j \quad , \quad B_q(t + j\,dt) = B_q(t)\ \forall j \tag{11.15}$$

In consequence, the parameter p defined in Equation 11.9 is time independent as well.

$$p(t + j\,dt) = p(t) = p\ \forall j$$

Equation 11.8 holds not only at time t but for other times as well, for example at time $t + dt$. This is true for both V_S^u and V_S^d:

$$\begin{aligned} V(S_u, t + dt) &= B_r(t)\,[p\,V(S_{uu}, t + 2dt) + (1 - p)\,V(S_{dd}, t + 2dt)] \\ V(S_d, t + dt) &= B_r(t)\,[p\,V(S_{ud}, t + 2dt) + (1 - p)\,V(S_{dd}, t + 2dt)] \end{aligned}$$

Substitution into Equation 11.8 leads to an expression for $V(S, t)$ as a function of the derivative price at time $t + 2dt$. Analogous expressions can be obtained for V_S^{uu}, V_S^{ud}, etc. This recursive procedure performed iteratively for $n = (T - t)/dt$ binomial steps gives

$$\begin{aligned} V(S, t) &= B_r(t, T) \sum_{j=0}^{n} \binom{n}{j} p^j (1 - p)^{n-j} V\left(u^j d^{n-j} S(t), T\right) \\ &= B_r(t, T) \sum_{j=0}^{n} B_{n,p}(j) V\left(u^j d^{n-j} S(t), T\right) \end{aligned} \tag{11.16}$$

where the second line was obtained by observing that $\binom{n}{j} p^j (1 - p)^{n-j}$ corresponds to a binomial probability density $B_{n,p}(j)$ but with parameters n and p (*not* with p' as in Equation 11.14).

Thus the value of the derivative at time t has been expressed as a sum over its values at a later time T. If this time T is chosen to be the maturity then the value of the derivative at time t is written in terms of its payoff profile. This reads explicitly for European calls and puts

$$\begin{aligned} c_s(t) &= B_r(t, T) \sum_{j=0}^{n} \binom{n}{j} p^j (1 - p)^{n-j} \max\left\{0, u^j d^{n-j} S(t) - K\right\} \\ p_s(t) &= B_r(t, T) \sum_{j=0}^{n} \binom{n}{j} p^j (1 - p)^{n-j} \max\left\{0, K - u^j d^{n-j} S(t)\right\} \end{aligned}$$

Because of the maximum function appearing in the summand, the sum for the call is effectively taken over the values of j for which $u^j d^{n-j} S(t)$ is larger than K. This condition can be written as $(u/d)^j > d^{-n} K/S$. Taking the logarithm of both sides yields the equivalent condition

$$j > \ln\left(\frac{K}{S(T) d^n}\right) / \ln\left(\frac{u}{d}\right)$$

[4]On both sides we use here again the short notation defined in Equation 11.4.

The sum is taken over whole numbers j. The smallest whole number greater than the right-hand side in the above inequality is

$$y = 1 + \text{Trunc}\left(\ln(\frac{K}{S(t)d^n})/\ln(\frac{u}{d})\right) \tag{11.17}$$

where the function "Trunc" is defined as the greatest whole number smaller than the argument (decimal values are simply truncated and not rounded). The number y defined in Equation 11.17 is thus the lower limit in the sum for the call (correspondingly, the sum for the put is taken over the whole numbers j ranging from 0 to the upper limit $y-1$). The value of a call is thus

$$\begin{aligned} c_s(t) &= S(t)B_r(t,T)\sum_{j=y}^{n}\binom{n}{j}p^j(1-p)^{n-j}u^j d^{n-j} \\ &- KB_r(t,T)\underbrace{\sum_{j=y}^{n}\binom{n}{j}p^j(1-p)^{n-j}}_{B_{n,p}(j\geq y)} \end{aligned}$$

According to Equation A.41, the last sum is the probability that a binomially distributed random variable (where $B_{n,p}$ denotes the binomial distribution with parameters n and p) is greater than or equal to y. Under Assumption 11.15 that yields are constant, i.e. $B_r(t) = B_r$ independent of t, we can write

$$B_r(t,T) = \prod_{k=0}^{n-1} B_r(t+kdt) = B_r^n \quad , \quad B_q(t,T) = B_q^n$$

Now the first sum can be represented as a binomial probability as well:

$$\begin{aligned} c_s(t) &= S(t)B_q(t,T)\sum_{j=y}^{n}\binom{n}{j}p^j u^j \frac{B_r^n}{B_q^n}(1-p)^{n-j}d^{n-j} \\ &- KB_r(t,T)\sum_{j=y}^{n}\binom{n}{j}p^j(1-p)^{n-j} \\ &= S(t)B_q(t,T)\underbrace{\sum_{j=y}^{n}\binom{n}{j}\hat{p}^j(1-\hat{p})^{n-j}}_{B_{n,\hat{p}}(j\geq y)} \\ &- KB_r(t,T)\underbrace{\sum_{j=y}^{n}\binom{n}{j}p^j(1-p)^{n-j}}_{B_{n,p}(j\geq y)} \end{aligned} \tag{11.18}$$

where[5]

$$\hat{p} = u\frac{B_r}{B_q}p \Rightarrow 1-\hat{p} = d\frac{B_r}{B_q}(1-p)$$

[5]Writing $S^u = uS$ and $S^d = dS$ in the third equation in 11.13 and using Equation 7.1 for $S(t,T)$

The value of a European put can be determined analogously with the help of the binomial distribution. So the prices of European options expressed in terms of binomial distributions are

$$\begin{aligned} c_s(t) &= B_q(t,T)\,S(t)B_{n,\hat{p}}(j \geq y) - B_r(t,T)KB_{n,p}(j \geq y) \\ p_s(t) &= -B_q(t,T)\,S(t)\,[1 - B_{n,\hat{p}}(j \geq y)] + B_r(t,T)\,K\,[1 - B_{n,p}(j \geq y)] \end{aligned} \tag{11.19}$$

Note the similarity to the famous Black-Scholes equation (see, for example Equation 9.6 or Equation 9.7). The difference is that the *binomial* distribution appears in the above expression in place of the *normal* distribution. In Section 11.4, we will see that the binomial distribution for infinitesimally small intervals dt converges towards a normal distribution and thus the binomial model approaches the Black-Scholes model in the limit $dt \to 0$.

As another example of the above procedure we demonstrate how the forward price can be determined by iterating Equation 11.13 for $n = (T-t)/dt$ binomial steps

$$S(t,T) = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} u^j d^{n-j} S(T)$$

where the spot price $S(T)$ for the forward contract after n steps is substituted into the equation, since the maturity date T will have been reached after this time. Under the assumption of constant interest rates and dividend yields, we obtain

$$\begin{aligned} S(t,T) &= \frac{B_q(t,T)}{B_r(t,T)} S(t) \sum_{j=0}^{n} \binom{n}{j} p^j u^j \frac{B_r^n}{B_q^n} (1-p)^{n-j} d^{n-j} \\ &= \frac{B_q(t,T)}{B_r(t,T)} S(t) \sum_{j=0}^{n} \binom{n}{j} \hat{p}^j (1-\hat{p})^{n-j} \\ &= \frac{B_q(t,T)}{B_r(t,T)} S(t) \underbrace{B_{n,\hat{p}}(j \geq 0)}_{1} \end{aligned}$$

corresponding to the result in Equation 7.1 obtained solely on the basis of arbitrage considerations.

---

yields

$$\frac{B_q}{B_r} S(t) = p\,u\,S(t) + (1-p)\,d\,S(t)$$

Dividing by the left-hand side gives

$$1 = p\frac{B_r}{B_q} u + (1-p)\frac{B_r}{B_q} d$$

It then follows immediately that

$$1 - \hat{p} \equiv 1 - u\frac{B_r}{B_q} p = d\frac{B_r}{B_q}(1-p)$$

11.2.3 A Third Contact with the Risk-Neutral World

Neither for derivatives (see Equation 11.8) nor for forward prices (see Equation 11.13) does the probability p' for the underlying S to increase to S_u (see, for example Equation 11.2) enter into any equation. The *valuation* of derivatives (or forward prices) does not depend on the probability that the underlying rises or falls! Instead, it depends on the value p as defined in Equation 11.9. If we could interpret p as a kind of "artificial probability", and if for pricing purposes we could put ourselves into an "artificial world" in which, after one step, the price of the underlying is given by S_u with this "probability" p (and not with the probability p'as in the real world), then the probability for the underlying to decrease to S_d over one time step would be $(1-p)$ in this "artificial world". The expression in brackets in Equation 11.8 would then just be the *expectation* of the derivative price *one* time step later with respect to the probability p in the artificial world. This holds for *many* time steps as well, since the binomial density $B_{n,p}(j)$ appearing in Equation 11.16 is the probability in the *artificial* world for the underlying to arrive at the value $S(T) = u^j d^{n-j} S(t)$; just as Equation 11.14 was this probability in the *real* world. The sum in Equation 11.16 over all the derivative values is then the *expectation* of the derivative value at the future time T in the artificial world.

We can summarize our observations in the following way: In an artificial world, where the probability of an up-move is p (and not p' as in the real world), **today's price of a derivative is the expectation of its future price discounted back to today**.

$$V(S,t) = B_r(t,T)\,\mathrm{E}[V(S,T)] \quad \text{Derivative on } S \tag{11.20}$$

The notation "$\mathrm{E}_p[]$" here means: "expectation with respect to the probability p".

Likewise, the forward price of the underlying in Equation 11.13 is exactly the expectation of the underlying's price at the future time T with respect to the probability p.

$$S(t,T) = \mathrm{E}_p\,[S(T)] \quad \text{Forward Price of } S \tag{11.21}$$

In the case of forward prices (which are not tradable financial *instruments* themselves) the expectation is not discounted.

By substituting Equation 7.1 for $S(t,T)$ into the above equation it follows that the dividend-adjusted spot price of the underlying (which is a tradable financial instrument) likewise can be expressed as the *discounted* expectation with respect to this probability

$$\tilde{S}(t,T) = \mathrm{E}_p\,[S(T)]\,B_r(t,T) \quad \text{Spot price } S \tag{11.22}$$

As in Sections 8.1.3 and 8.2.1, it does not matter if the underlying is expected to rise or fall in the real world. This plays no role in the valuation of derivatives on the underlying. The valuation is independent of the expected changes in the underlying. In contrast to the real world where investors are compensated for taking the risk of investing in an underlying by the underlying's mean return, this mean return doesn't play any role at all in the artificial world used for pricing derivatives. This artificial world is thus neutral to the risk inherent in the underlying and is therefore called the *risk-neutral world* and the probability p is called the *risk-neutral probability.* We are again confronted with the *risk neutrality* described in Sections 8.1.3 and 8.2.1.

At first glance, this result may seem rather unintuitive, but upon further consideration we see that it is reasonable both economically and intuitively. The construction of a binomial tree with a *subjective* real world probability would yield different trees and thus different option prices for each investor. Trading with derivatives would thus hardly be possible since the market participant could not agree on a price in most cases. However, from arbitrage considerations making use of the fact that options and futures can be replicated by a portfolio consisting of the underlying and risk-free assets, it follows that a purely objective, risk-neutral probability p for an up-move in a risk-neutral world exists, eliminating this subjectivity (see Equation 11.9).

To see how powerful Equations 11.20, 11.21 and 11.22 are, we have to be more specific. We now choose a stochastic process for the underlying. In what follows, we will assume that the relative changes of $S(t)$ behave as a random walk as in Equation 3.13, i.e. we will rely on Assumption 7 from Section 5. But Equation 3.13 was established to model the behavior of S in the *real* world[6]. We will show later[7], however, that if S performs a random walk in the real world, it *also* performs a random walk in the risk-neutral world. The distribution and first moments of such a random walk are those given in Table 3.5 at the end of Section 3.3. Thus, the underlying is lognormally distributed with expectation

$$\langle S(T) \rangle = S(t) e^{(\mu + \sigma^2/2)(T-t)} \tag{11.23}$$

In the risk-neutral world (i.e. in the world we need for pricing) this expectation has to be equal to the expectation $\mathrm{E}_p[S(T)]$ with respect to the risk-neutral probability p:

$$\langle S(T) \rangle \stackrel{!}{=} \mathrm{E}_p[S(T)]$$

Substituting Equation 11.22 and 11.23 into this requirement completely determines the drift μ of the underlying in the risk-neutral world, i.e. the drift to be used for pricing:

$$\begin{aligned} & S(t) e^{(\mu + \sigma^2/2)(T-t)} \stackrel{!}{=} \frac{\tilde{S}(t,T)}{B_r(t,T)} \\ & \Rightarrow \\ & \mu \stackrel{!}{=} \frac{1}{T-t} \ln \left(\frac{\tilde{S}(t,T)}{S(t) B_r(t,T)} \right) - \frac{\sigma^2}{2} \end{aligned}$$

Using the forward price equation 11.21 instead, we obtain the drift from the ratio of forward price to spot price:

$$\begin{aligned} & S(t) e^{(\mu + \sigma^2/2)(T-t)} \stackrel{!}{=} S(t,T) \\ & \Rightarrow \\ & \mu \stackrel{!}{=} \frac{1}{T-t} \ln \left(\frac{S(t,T)}{S(t)} \right) - \frac{\sigma^2}{2} \end{aligned} \tag{11.24}$$

[6] A "risk-neutral" world was never mentioned in the vicinity of Equation 3.13, nor in the whole of Chapter 3.

[7] We will explicitly show this in great detail and on a much more fundamental basis in Chapter 14 when we discuss the famous Girsanov theorem.

With a dividend yield q and continuous compounding, we obtain, for example,

$$\mu = \frac{1}{T-t} \ln \left(\frac{B_q(t,T)}{B_r(t,T)} \right) - \frac{\sigma^2}{2} = r - q - \frac{\sigma^2}{2} \tag{11.25}$$

where the first equality holds as a result of the assumed dividend yield and the second is valid for continuous compounding. But this is exactly Equation 8.19.

As was pointed out in Equation 3.14, the drift μ is exactly equal to the expected return of the underlying. This means that the expected return in the risk-neutral world (i.e. with respect to the probability p) is objectively given through the risk-free interest rate and the dividends (or through the ratio of forward price to spot price) and the volatility, independent of an investor's opinion as to whether the price will rise or fall. The parameter μ thus determined is called the *risk-neutral yield* or the *risk-neutral drift*.

We note for later reference that all this holds for any arbitrary time span $T-t$, for instances also for one time step dt in a binomial tree:

$$\mu \, dt = \ln \left(\frac{B_q(t)}{B_r(t)} \right) - \frac{\sigma^2}{2} dt \tag{11.26}$$

11.3 The Relationship between Random Walk and Binomial Parameters

Risk neutrality is the essential link connecting the stochastic model describing the underlying with the pricing method used for a derivative. This will be demonstrated for the binomial model. The parameters u and d must first be determined before the binomial model can be applied in pricing derivatives. The choice of these parameters has a significant influence on derivative and forward prices thus calculated. To make a reasonable choice of u and d further assumptions concerning the behavior of the underlying must be made, i.e. a stochastic process for the underlying must be specified. We will again assume that the relative changes of $S(t)$ behave as a random walk as in Equation 3.13 and are therefore normally distributed with moments of the form given in Table 3.5 at the end of Section 3.3. On the other hand we know from the previous sections that in a binomial tree the underlying S is distributed according to the binomial distribution, see Equation 11.14, where for pricing purposes we have to replace the real world probability p' by the risk-neutral "probability" p. We will now relate the random walk parameters μ and σ to the parameters u and d of the binomial tree by *matching the moments* of the random walk distribution to the moments of the binomial distribution. We will ensure that we work in the risk-neutral world (i.e. that we determine the parameters needed for pricing) by using p as defined in Equation 11.9 for the binomial tree and by using the risk-neutral drift defined in Equation 11.25 for the random walk.

In the following derivation, we will assume that the parameters u and d are constant over time until maturity. After j up-moves and $n-j$ down-moves, the final value $S(T)$

and thus the logarithm of the relative price change is

$$S(T) = u^j d^{n-j} S(t) \quad \Rightarrow \quad \ln\left(\frac{S(T)}{S(t)}\right) = j \ln\left(\frac{u}{d}\right) + n \ln(d)$$

$S(T)$ (and thus j) is binomially distributed in our binomial model and from Equation A.43, it follow that $\langle j \rangle = np$ and $\text{var}(j) = np(1-p)$.

For the random walk model, on the other hand, the expectation and variance of the logarithmic changes of S are equal to the drift and the square of the volatility, each multiplied by the time difference $T - t$ (see the first column of Table 3.5). Thus matching the first two moments of the random walk distribution to the distribution induced by the binomial tree yields

$$\begin{aligned} \mu(T-t) &= \left\langle \ln\left(\frac{S(T)}{S(t)}\right) \right\rangle = \underbrace{\langle j \rangle}_{np} \ln\left(\frac{u}{d}\right) + n \ln(d) \\ \sigma^2(T-t) &= \text{var}\left(\ln\left(\frac{S(T)}{S(t)}\right)\right) = \underbrace{\text{var}\,(j)}_{np(1-p)} \left(\ln\left(\frac{u}{d}\right)\right)^2 \end{aligned} \tag{11.27}$$

Because $dt = (T-t)/n$ this can be written as

$$\begin{aligned} \mu dt &= p \ln\left(\frac{u}{d}\right) + \ln(d) \\ \sigma^2 dt &= p(1-p) \left(\ln\left(\frac{u}{d}\right)\right)^2 \end{aligned}$$

Now we use Equation 11.9 for the risk-neutral probability p and Equation 11.26 for the risk-neutral drift μ to establish a system of two (non-linear!) equations for the two unknown binomial parameters u and d:

$$\begin{aligned} \ln\left(\frac{B_q}{B_r}\right) - \frac{\sigma^2}{2} dt &= \frac{B_q/B_r - d}{u-d} \ln\left(\frac{u}{d}\right) + \ln(d) \\ \sigma^2 dt &= \frac{(B_q/B_r - d)\,(u - B_q/B_r)}{(u-d)^2} \left(\ln\left(\frac{u}{d}\right)\right)^2 \end{aligned} \tag{11.28}$$

There exist several closed form solutions to this system which are exact up to linear order in dt. One such solution is given by

$$u = \frac{B_q}{B_r} e^{-(\sigma^2/2)dt + \sigma\sqrt{dt}} \quad , \quad d = \frac{B_q}{B_r} e^{-(\sigma^2/2)dt - \sigma\sqrt{dt}} \tag{11.29}$$

Inserting this into Equation 11.9 for the risk-neutral probability yields

$$p = \frac{e^{(\sigma^2/2)dt} - e^{-\sigma\sqrt{dt}}}{e^{+\sigma\sqrt{dt}} - e^{-\sigma\sqrt{dt}}} = \frac{e^{(\sigma^2/2)dt} - e^{-\sigma\sqrt{dt}}}{2\,\sinh(\sigma\sqrt{dt})}$$

where we have used the definition of the hyperbolic sine function in the last step. Using Equation 11.26 (which is equivalent to $\exp(\mu\, dt) = \exp\left(-\sigma^2 dt/2\right) B_q/B_r$) we

can bring the parameters u and d into a more intuitive form using the risk-neutral drift:

$$u = e^{\mu\, dt + \sigma\sqrt{dt}} \quad , \quad d = e^{\mu\, dt - \sigma\sqrt{dt}}$$

Thus, in this solution the drift only appears in the parameters u and d, while the probability p is determined solely from knowledge of the volatility. For small values of dt and assuming continuous compounding, the Taylor series representation of the exponential function expanded up to linear terms in dt gives

$$u \approx 1 + \sigma\sqrt{dt} + (q - r)\, dt \;, \;\; d \approx 1 - \sigma\sqrt{dt} + (q - r)\, dt \;, \;\; p \approx 1/2 \tag{11.30}$$

Since in this solution both the volatility and the risk neutral drift appear in u and d, we must assume constant volatilities and because of Equation 11.25 also constant yields and dividends, i.e. Assumptions 9, 11 and 12 from Section 5, to ensure that the parameters u and d are constant over time and, in consequence, that the tree recombines.

Another frequently used solution of Equation 11.28 for which it suffices to assume constant volatilities (Assumption 11) is

$$u = e^{+\sigma\sqrt{dt}} \quad , \quad d = e^{-\sigma\sqrt{dt}} \Rightarrow p = \frac{B_q/B_r - e^{-\sigma\sqrt{dt}}}{2\, \sinh(\sigma\sqrt{dt})} = \frac{e^{(q-r)dt} - e^{-\sigma\sqrt{dt}}}{2\, \sinh(\sigma\sqrt{dt})} \tag{11.31}$$

where the last step is of course only valid for continuous compounding. In this solution, the volatility alone completely determines the parameters u and d. Observe that $u(t) = 1/d(t)$ holds. As long as the volatility is constant (allowing the parameters u and d to remain constant over time), the tree is recombining since the starting price is recovered after an up-move followed by a down-move:

$$u(t)\, d(t + dt) = u(t)d(t) = u(t)/u(t) = 1$$

The ease in the construction and analysis of binomial trees resulting from this relation prompts us to utilize the solution given by Equation 11.31 exclusively in the remainder of this book whenever we use binomial trees. For small time intervals dt, the Taylor series representation of the exponential function expanded up to and including terms of linear order in dt yields the following approximations for u, d and p:

$$u \approx 1 + \sigma\sqrt{dt} + \frac{\sigma^2}{2}dt, \;\; d \approx 1 - \sigma\sqrt{dt} + \frac{\sigma^2}{2}dt \;, \;\; p \approx \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{dt}\right)$$

where again the risk-neutral drift μ is used to simplify the last expression.

A detailed demonstration of the application of binomial trees for the valuation of an option portfolio is provided in the Excel workbook BinomialTree.xls. In anticipation of Chapter 13, the evaluation of the *Greeks* (derivatives of the option price with respect to its parameters) using binomial trees also receives attention. This workbook can be used as a small but fully functioning option calculator (as always, the yellow fields are the input fields).

11.4 The Binomial Model with Infinitesimal Steps

In this section, the Black-Scholes option pricing formula is derived directly from the binomial model for European options as given by Equation 11.19. A deeper insight into the relationship between these two important methods in option pricing (finding solutions to a differential equation on the one hand and the calculation of (discounted) expectations on the other) can be gained from an understanding of this derivation. The reader less interested in mathematics may choose to continue on to the next section.

A classical result from statistics, the *Moivre-Laplace theorem,* states that the binomial distribution converges towards a normal distribution as the number n of the observed trials approaches infinity. The statement of the theorem in integral form can be expressed as

$$B_{n,p}(a \leq \frac{j-np}{\sqrt{np(1-p)}} \leq b) \stackrel{n\to\infty}{\longrightarrow} \mathrm{N}(b) - \mathrm{N}(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz$$

The left-hand side of the equation denotes the probability that the standardized form of a *binomially distributed* random variable j (i.e. j less its expectation divided by its standard deviation) will lie in the interval between a and b, while the right-hand side is simply the probability that a *standard normally distributed* random variable will take on values lying within the same interval.

We can exploit this theorem to see what happens to Equation 11.19 as the time interval dt converges towards zero, i.e. as the number n of steps in the binomial tree approaches infinity. This will be demonstrated for a call, the procedure for a put being completely analogous.

We need, for example, to determine the limit of $B_{n,p}(y \leq j)$ in Equation 11.19 For arbitrary constants f and g with $g > 0$, the probability that $y \leq j$ is of course equal to the probability that $(y-f)/g \leq (j-f)/g$. We take advantage of this fact[8] to manipulate $B_{n,p}(y \leq j)$ into a suitable form to apply the Moivre-Laplace theorem:

$$\begin{aligned} B_{n,p}(y \leq j) &= B_{n,p}\left(\frac{y-np}{\sqrt{np(1-p)}} \leq \frac{j-np}{\sqrt{np(1-p)}} \leq \infty \right) \\ &\stackrel{n\to\infty}{\longrightarrow} \underbrace{\mathrm{N}(\infty)}_{1} - \mathrm{N}\left(\frac{y-np}{\sqrt{np(1-p)}} \right) \\ &= \mathrm{N}\left(\frac{np-y}{\sqrt{np(1-p)}} \right) \end{aligned}$$

where in the last step the symmetry property of the normal distribution, Equation A.53, is used.

Equations 11.27 and 11.17 deliver the necessary elements for computing $(np - y)/\sqrt{np(1-p)}$:

$$np = \frac{1}{\ln(u/d)} \left(\mu(T-t) - n\ln(d) \right) \ , \quad \sqrt{np(1-p)} = \frac{1}{\ln(u/d)} \sigma \sqrt{T-t}$$

[8] With the choice $f = np$ and $g = \sqrt{np(1-p)}$

and

$$y = \frac{\ln\left(\frac{K}{S(t)d^n}\right)}{\ln(u/d)} + \varepsilon = \frac{\ln\left(\frac{K}{S(t)}\right) - n\ln(d) + \varepsilon\ln(u/d)}{\ln(u/d)} \quad \text{with} \quad 0 < \varepsilon \leq 1$$

Here, ε represents the difference between $\ln(K/Sd^n)/\ln(u/d)$ and the smallest whole number greater than this value. We will show immediately that the term $\varepsilon\ln(u/d)$ becomes arbitrarily small. Substituting accordingly yields the argument for the standard normal distribution above:

$$\frac{np - y}{\sqrt{np(1-p)}} = \frac{\ln(S(t)/K) + \mu(T-t) - \varepsilon\ln(u/d)}{\sigma\sqrt{T-t}}$$

In both solutions 11.29 and 11.31 the ratio u/d converges towards 1 as $n \to \infty$ (i.e. $dt \to 0$), thus by continuity, $\ln(u/d)$ converges towards zero. Thus, the limit for infinitely many binomial steps becomes

$$\frac{np - y}{\sqrt{np(1-p)}} \overset{n\to\infty}{\longrightarrow} \frac{\ln(S(t)/K) + \mu(T-t)}{\sigma\sqrt{T-t}}$$

Using now Equation 11.25 for the risk-neutral drift we can write

$$\frac{np - y}{\sqrt{np(1-p)}} \overset{n\to\infty}{\longrightarrow} = x - \sigma\sqrt{T-t}$$

where we have defined the abbreviation x as in Equation 9.5:

$$x := \frac{1}{\sigma\sqrt{T-t}} \ln\left(\frac{S(t)\,B_q(t,T)}{K\,B_r(t,T)}\right) + \frac{1}{2}\sigma\sqrt{T-t}$$

The limit of $B_{n,p}(y \leq j)$ is thus established. Proceeding analogously, we can calculate the limit of the other binomial probability in Equation 11.19. In summary, for an infinite number of binomial steps in a finite time interval, the behavior of the binomial distribution is given by

$$B_{n,p}(y \leq j) \overset{n\to\infty}{\longrightarrow} \mathrm{N}\left(x - \sigma\sqrt{T-t}\right) \;, \quad B_{n,\hat{p}}(y \leq j) \overset{n\to\infty}{\longrightarrow} \mathrm{N}\,(x)$$

Using these convergence relations, we obtain the value of a call as the number of binomial steps approaches infinity to be

$$c_s(t) \overset{n\to\infty}{=} S(t)B_q(t,T)\mathrm{N}(x) - KB_r(t,T)\mathrm{N}(x - \sigma\sqrt{T-t}) \tag{11.32}$$

This is in complete agreement with Equation 9.6 and is thus (again!) the famous *Black-Scholes option pricing formula*.

11.4.1 Components of the Black-Scholes Option Pricing Formula

In the above section the Black-Scholes formula was derived from Equations 11.18 and 11.19. We can see from this derivation that the cumulative normal distribution found

$\mathrm{N}(x - \sigma\sqrt{T-t})$	Probability for a call to be exercised
$\mathrm{N}(-x + \sigma\sqrt{T-t})$	Probability for a put to be exercised
$KB_r(t,T)$	Present value (PV) of the money paid upon exercise
$B_q(t,T)\mathrm{N}(x)$	Number of underlyings to be long for replicating the call
$B_q(t,T)\mathrm{N}(-x)$	Number of underlyings to be short for replicating the put
$KB_r(t,T)\mathrm{N}(x - \sigma\sqrt{T-t})$	Money to be borrowed to replicate the call = PV of money paid upon exercise × exercise probability
$KB_r(t,T)\mathrm{N}(-x + \sigma\sqrt{T-t})$	Money to be invested to replicate the put = PV of money received upon exercise × exercise probability

Table 11.1: Interpretation of the various components in the Black-Scholes option pricing formulae for plain vanilla calls and puts. All probabilities mentioned are risk-neutral.

next to the discounted strike price K is the risk-neutral probability for the price of the underlying to be larger than the strike price. This is referred to as the *risk-neutral exercise probability*. A comparison with the replicating portfolio in Equation 11.3 shows that the number $\Delta(t)$ of underlyings needed to replicate the option is given by the factor next to $S(t)$ in Equation 11.32, namely $B_q(t,T)\mathrm{N}(x)$, while the amount $g(t)$ of money in the bank account is given by the second summand in Equation 11.32. The intuitive interpretations of these values in the Black-Scholes formulae for puts and calls are collected in Table 11.1.

11.5 Trinomial Trees

Trinomial trees present us with an alternative method to binomial trees. The form of a trinomial tree is represented graphically in Figure 11.2. The j^{th} step of the tree at time t_j is connected, not with two other nodes in the next step (as was the case for the binomial tree), but with three. The price paid for this additional degree of freedom is additional computational effort. The advantage is that a trinomial tree can always be constructed in such a way that it recombines and in addition, achieves the same degree of accuracy as the binomial tree with fewer time steps. The trinomial tree has $2j+1$ nodes after j steps where the time t is indexed with $j = 0$. The length of the time steps may vary. The value of the underlying at the ith node after j steps is denoted by S_{ji} where

$$i = -j, -j+1, \ldots, j-1, j$$

Each node at time step j branches into three nodes at time step $j+1$, with a probability being associated with each of these branches[9]. Starting from S_{ji}, we denote by p^+ the probability that the underlying will increase from S_{ji} to the value $S_{j+1,i+1}$ at time step $(j+1)$. Correspondingly, denote by p^0 and p^- the probabilities that the underlying

[9] All probabilities appearing in this context are risk-neutral probabilities.

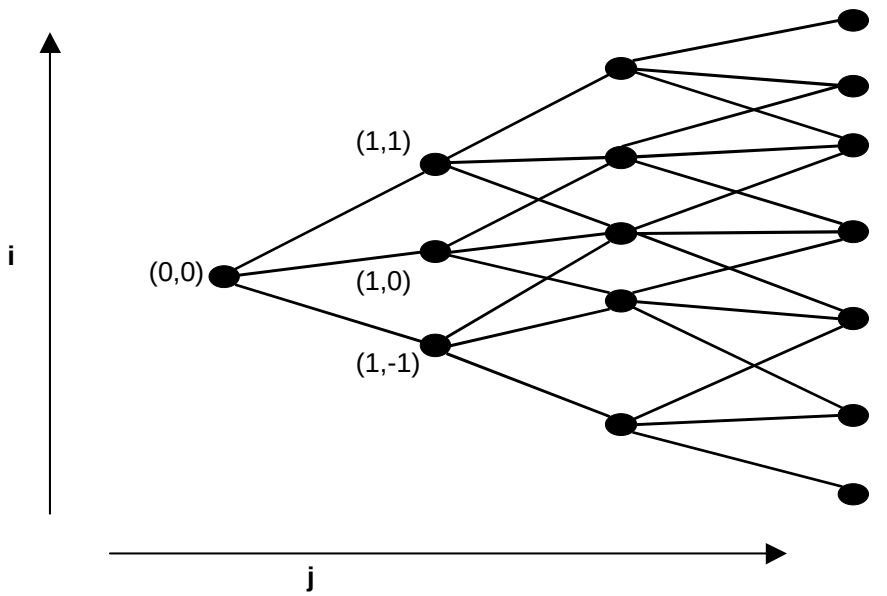


Figure 11.2: A simple trinomial tree.

at time step $(j+1)$ will take on the values $S_{j+1,i}$ and $S_{j+1,i-1}$, respectively. Each of these probabilities must be ≥ 0 and ≤ 1. In addition:

$$p^+ + p^0 + p^- = 1 \tag{11.33}$$

For the *bi*nomial tree[10], a portfolio can be constructed consisting of a position Δ in the underlying and money g in a bank account whose value replicates the option price exactly at each time step. This is not possible for trinomial trees since the addition of the *third* attainable value for the underlying after one time step makes it impossible to find a replicating portfolio on the basis of just *two* parameters Δ and g. However, it is still possible to price options with the help of trinomial trees.

To do so, it is sufficient to choose the probabilities p^+, p^0 and p^- and the nodes S_{ji} so that the tree reflects the probability distribution of the underlying. Again, we assume that the underlying price is lognormally distributed (this corresponds to Assumption 7 from Section 5). The lognormal distribution is completely determined by two parameters, the expectation and the variance of the logarithm. Thus, two conditions are sufficient to adapt the trinomial tree to a lognormal distribution. Like in Equation 11.23, the expectation for the lognormally distributed random variable after a time step of length dt starting from the node S_{ji} (with the risk-neutral drift Equation 11.25) is given by

$$\langle S(t_{j+1}) \rangle = S_{ji} e^{(r-q)dt}$$

[10]The binomial tree usually assumes a constant time step dt. In the most general case, this assumption is not necessary.

On the other hand, from the tree we have

$$\langle S(t_{j+1})\rangle = p^+ S_{j+1,i+1} + p^0 S_{j+1,i} + p^- S_{j+1,i-1}$$

Setting these two expressions equal to one another yields one equation for the determination of the probabilities:

$$S_{ji} e^{(r-q)dt} = p^+ S_{j+1,i+1} + p^0 S_{j+1,i} + p^- S_{j+1,i-1} \tag{11.34}$$

Analogously, taking the expression for the variance of the lognormal distribution shown in Table 3.5 at the end of Section 3.3 we have

$$\text{var}\left[S(t_{j+1})\right] = S_{ij}^2 e^{2(r-q)dt}\left(e^{\sigma^2 dt} - 1\right)$$

It is sometimes easier to work with the expectation of $S^2(t_{j+1})$ rather than the variance and such is the case here. With the help of Equation A.7 we obtain this expectation as

$$\begin{aligned}\left\langle S^2(t_{j+1})\right\rangle &= S_{ji}^2 e^{2(r-q)dt}\left(e^{\sigma^2 dt} - 1\right) + S_{ji}^2 e^{2(r-q)dt} \\ &= S_{ji}^2 e^{2(r-q)dt} e^{\sigma^2 dt}\end{aligned}$$

Expressed in terms of the probabilities for the trinomial tree, the same expectation is given by

$$\left\langle S^2(t_{j+1})\right\rangle = p^+ S_{j+1,i+1}^2 + p^0 S_{j+1,i}^2 + p^- S_{j+1,i-1}^2$$

Combining the two above expressions gives

$$S_{ji}^2 e^{2(r-q)dt} e^{\sigma^2 dt} = p^+ S_{j+1,i+1}^2 + p^0 S_{j+1,i}^2 + p^- S_{j+1,i-1}^2 \tag{11.35}$$

Equations 11.33, 11.34 and 11.35 are sufficient to fit the trinomial tree to the lognormal distribution. Of course, only three of the six parameters p^+, p^0, p^-, $S_{j+1,i+1}$, $S_{j+1,i}$ and $S_{j+1,i-1}$ will be determined. In general, these will be the probabilities. The nodes can then be arbitrarily selected.

The actual option pricing now proceeds as for a binomial tree. V_{ji} denotes the price of the option at the ith node of the jth time step. We assume that the tree consists of N time steps with $j = 0, 1, \ldots, N$. To price a European option, the nodes are initialized with the payoff profile of the option at maturity $t_N = T$. In the case of a call option:

$$V_{Ni} = \max(S_{Ni} - X, 0)$$

The calculation then *rolls backwards* through the tree. The option value is calculated iteratively for a time step using the values just calculated at the next time step starting with $j = N - 1$ and working back to $j = 0$:

$$V_{ji} = B_r(t_j, t_{j+1})\left[p^+ V_{j+1,i+1} + p^0 V_{j+1,i} + p^- V_{j+1,i-1}\right]$$

V_{00} is the present value of the option at time $t_0 = t$ (assuming that S_{00} is the price of the underlying at $t = t_0$). It should be emphasized that the model admits both time-dependent interest rates and volatilities. To take this into consideration either the nodes need to be selected accordingly or the probabilities must be made time-dependent. American options are treated in the same manner as they are treated in binomial trees. Barrier options should be calculated by choosing the nodes such that they lie *directly* on the barrier.

11.5.1 The Trinomial Tree as an Improved Binomial Tree

After *two* time steps a recombining binomial tree has exactly three distinct nodes. This is equal to the number of nodes in the trinomial tree after *one* step. Since the nodes of the trinomial tree can be freely chosen, it is possible to generate a *tri*nomial tree (with an even number of time steps) corresponding to any given recombining *bi*nomial tree. Such a trinomial tree yields the exact same results as the binomial tree, but in only half the time steps. This will be demonstrated for the binomial tree with parameters u, d, and p as given in Equation 11.31 serving as an example. Starting from the node S_{ij}, we can choose the nodes at time $t_{j+1} = t_j + 2\,dt$ as follows:

$$\begin{aligned} S_{j+1,i+1} &= u^2 S_{ij} \\ S_{j+1,i} &= S_{ij} \\ S_{j+1,i-1} &= d^2 S_{ij} \end{aligned}$$

where dt is the length of one time step in the binomial tree and hence $2\,dt$ is the length of one time step in the trinomial tree. The probabilities for the trinomial tree are easily obtained from the probability p in the binomial tree:

$$p^+ \;=\; p^2\;, \quad p^0 \;=\; 2p(1-p)\;, \quad p^- \;=\; (1-p)^2$$

The values for the probabilities are consistent with those in Equations 11.33, 11.34 and 11.35.

The trinomial tree converges faster than the corresponding binomial tree because it requires only half as many steps. Approximately half of the nodes in the binomial tree need not be computed. This advantage is, however not quite as great as it may seem at first glance. It is known that the results of the binomial tree oscillate strongly when the number of time steps increases by one. The best results are obtained by averaging two calculations with N and $N+1$ time steps. This trick can not be exploited when using trinomial trees. Moreover, the parameters specified above are not an optimal choice for the trinomial tree.

11.5.2 Relationship to the Explicit Finite Difference Method

Numerically, the trinomial tree and the explicit method of finite differences are equivalent, at least in certain special cases. Consider the ith node of the time step j. This corresponds to an underlying price S_{ji}. This node is connected to the nodes $i+1$, i and $i-1$ of the time step $j+1$. The value of the option V_{ji} at node i of the time step j is given by

$$V_{ji} = B_r(t_j, t_{j+1}) \left[p^+ V_{j+1,i+1} + p^0 V_{j+1,i} + p^- V_{j+1,i-1} \right] \;.$$

The analogous expression for the explicit finite difference method is given by Equation 10.25 with the coefficients given by Equation 10.13 (or Equation 10.11 for uniform grids):

$$\begin{aligned} W_{i,j} = {} & (1 - (t_{j+1} - t_j) B_{i,j+1})\, W_{i,\;j+1} - (t_{j+1} - t_j) A_{i,j+1}\, W_{i-1,j+1} \\ & - (t_{j+1} - t_j) C_{i,j+1}\; W_{i+1,j+1} \end{aligned}$$

We immediately see the similarity in structure. The probabilities can be easily associated with the coefficients of $W_{i+1,j+1}$, $W_{i,j+1}$ and $W_{i-1,j+1}$. This requires the grid points of the finite difference grid to correspond exactly to the nodes of the trinomial tree and thus the restriction $S_{ji} = S_{j+1,i}$ to hold for the nodes of the trinomial tree. The finite difference grid was constructed specifically to satisfy just this very condition (even for non-uniform grids) to prevent the expressions for the finite differences from becoming unnecessarily complicated. The trinomial tree appears more flexible from this point of view. However, it has all the disadvantages associated with the explicit method. Moreover, a procedure corresponding to the implicit or Crank-Nicolson method is not available for trinomial trees.

12

Monte Carlo Simulations

Having recognized the fact that prices of financial instruments can be calculated as discounted future expectations (with respect to a risk-neutral probability measure), the idea of calculating such expectations by simulating the (stochastic) evolution of the underlyings several times and subsequently averaging the results somehow is not far removed. In fact, this relatively simple idea is widely used and is successful even in the valuation of very exotic options for which other methods are either too complicated or completely unsuitable, the only requirement being the availability of sufficient computation time. Before proceeding with financial *applications* of Monte Carlo techniques, we begin with a presentation of the technique itself.

If random events occur often enough, they can be used to answer diverse questions statistically. This has long been common knowledge in science and we have seen a vast increase in applications with the advance of modern computers over the last twenty years or so, since computers suddenly made it possible to generate "random" events cheaply and in large numbers, or in the language of the specialist, to *simulate* them. Ever since, *computer simulations* have been indispensable in science, and since lately also in the modern financial world. Since generating "random" events lies at the core of such simulations, the name *Monte Carlo simulation* has become accepted despite the fact that the method's namesake city in Monaco could never generate as many random events as are sometimes necessary in practice, even if all the casinos in Monte Carlo were open non-stop for business every day for a million years. From this point of view, the computer can far outperform the roulette table.

Without simulations, there is only one way of calculating a problem on a computer, and that is to find a solution (or at least an approximate solution) either numerically or analytically from the a basic theory (where the theory is already a simplified, in other words, only an approximate representation of reality). Such solutions exist usually only for highly simplified models of the real situation and are often, even after these simplifications, quite complicated. For many real world problems arising in practice, no analytical solutions are available *at all*. We are dependent on measurements (in finance that means observations of the market) which have the distinct disadvantage that the only observable events are those that have already occurred.

The advent of computer simulations brought about a fundamental change to this

situation: it is now possible to deal with even the most complicated problems using a computer since artificial "observations" can be generated by computer simulations. In consequence, the most important practical advantages of theory over experimentation can be exploited even for questions which will never be answered by a "formula". These "theory-advantages" are:

- Fast solutions of the problem through computer computations.
- The ability to predict and perform what-if analyses, since the parameters governing a certain problem can be arbitrarily selected, even when they take on values which have never occurred in reality (yet).
- The answers produced by a simulation are usually much easier to understand, compared to theoretical solutions which are often highly complicated, if they exist at all.

For the end user, for example, a trader who simply enters the parameters of an option into the computer to obtain a price, it makes absolutely no difference whether the computation is accomplished using the Black-Scholes formula or a Monte Carlo simulation. In both cases, a result is obtained from the computer. Herein lies the incredible potential utility of computer simulations.

Assumptions

A Monte Carlo simulation in its most general form consistent with arbitrage-free pricing requires only the assumptions eliminating arbitrage opportunities, i.e. Assumptions 1, 2, 3, 4 and 5 from Section 5. There is one additional assumption, however, namely that a right to exercise early is of no value. This assumption is necessary because, as we will show below, the standard Monte Carlo simulation cannot be used to evaluate the price of instruments with *American* exercise features[1].

The method presented here is based on the random walk equation 3.13 with constant drift and volatility. Hence, because of Equation 11.25 for pricing in a risk-neutral world, constant yields must also be assumed. This means that for the method presented, the additional assumptions 7, 11 and 12 from Section 5 are also made. The assumptions of constant interest rates and volatilities naturally imply that interest rates and volatilities are non-stochastic, i.e. Assumptions 8 and 10 hold as well.

These extensive assumptions are made to allow a clear presentation of the material but are, in principle, not necessary for performing Monte Carlo simulations. For example, not only the underlying price, but also the volatility can be simultaneously simulated as stochastic processes (for instance also as a random walk). Since then two of the parameters involved are stochastic processes, the random walk occurs now in two dimensions. Of course, the volatility of the volatility and also the drift of the volatility would be required as parameters in such a situation. The simulation then proceeds as follows: first, a value for the volatility is simulated. Using this simulated volatility as a

[1]There are only very few Monte Carlo methods which can be used to price *American* options and these are quite complicated. This topic is to be viewed as a matter for research. The practical applications are as yet quite limited.

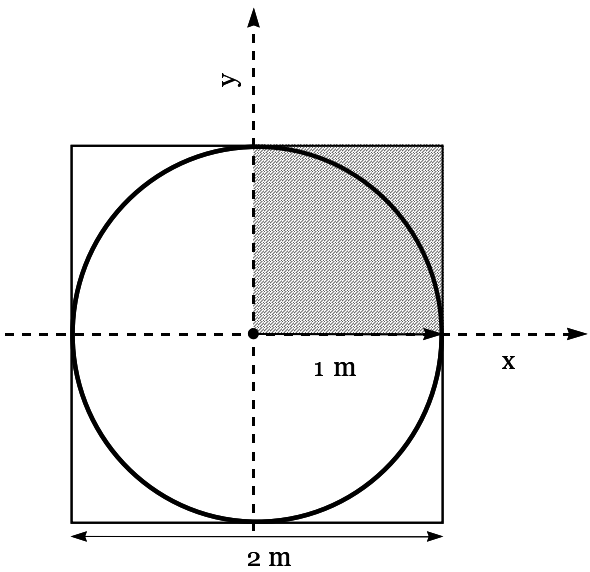


Figure 12.1: A disk with diameter $2m$ in a square with sides of length $2m$.

parameter, the next step for the underlying price is simulated. The two random walks are thus not independent of one another since the random walk describing the volatility affects that of the price. In reality, it can often be observed that the converse holds as well, i.e. the price of the underlying has an influence on the volatility (low volatility for rising prices and high volatility for falling prices). This effect can be incorporated into the simulation using a special form for the drift of the volatility random walk, which depends on the underlying price, for example a volatility-drift $rdt/S(t)$.

When simulating interest rates for pricing caps or floors, for example, Equation 3.13 can naturally be replaced by a more complex process corresponding to a term structure model such as Heath-Jarrow-Morton, Ho-Lee, Hull-White, etc. Mean reversion, for example, intuitively corresponds to a random walk under the influence of an external force which has the effect of driving the random walk back to an asymptotic mean value.

12.1 A Simple Example: The Area of a Disk

It is necessary to first understand the essence of the Monte Carlo simulation before attempting to apply it. Therefore, at this point we present a very simple example of the idea behind the Monte Carlo method before entering into a discussion of applications in the financial world. As known from elementary mathematics, the area of a disk of radius R is πR^2, the product of the square of the radius with the constant π, already known to the ancient Egyptians: $\pi = 3.14159...$ A disk with a radius of one meter, $R = 1m$, thus has an area of $\pi \times (1m)^2 = 3.14159m^2$. Supposing we had never heard of

the number π, this fact can be ascertained with the help of random events. To do so, we simply place square box with sides of length $2m$ containing a "pie dish" as shown in Figure 12.1 out in the rain. The pie dish represents the disk of radius 1m whose area is to be determined. We know that the square has an area of $2m \times 2m = 4m^2$. The random events are the falling raindrops. Assuming that the raindrops fall evenly on the square, then the area of the disk can be given by

$$\text{Area of the disk} = \frac{\text{Number of raindrops falling on the disk}}{\text{Number of raindrops falling on the entire square}} \times 4m^2$$

In this way, the area of a disk can be determined with the help of random events.

The same "Experiment" admits another interpretation. If the equation "disk area $= \pi R^2$" is known, then the value of π can be determined:

$$\pi = \frac{\text{disk area}}{R^2} = \frac{\text{disk area}}{1m^2} = 4 \times \frac{\text{Number of raindrops falling on the disk}}{\text{Number of raindrops falling on the entire square}}$$

Thus, with the help of randomly falling raindrops, the value of the natural constant π has been determined.

The primary application of Monte Carlo simulations – in particular in the financial world – is the calculation of integrals. For example, the price of an option can be expressed as the integral of its payoff profile with respect to an appropriate probability measure associated to the price of its underlying at maturity. Making use of our disc-example, we give a third interpretation of the experiment described above to illustrate how random events can be used to calculate an integral: as is known from elementary mathematics, a semi-circle can be represented as the graph of a function. Choosing the x- and y-axis as shown in Figure 12.1, the ancient *Pythagorean Theorem* says that the points (x, y) on the circle all satisfy

$$x^2 + y^2 = R^2 = 1$$

Thus the upper semi-circle is the graph of the function $y = \sqrt{1 - x^2}$. The lower semi-circle is the graph of the function $y = -\sqrt{1 - x^2}$. As is well know from any introduction to mathematical analysis, integrating a function gives the area under the curve given by the graph of that function. Since the upper semi-circle is represented by the graph of the function $y = \sqrt{1 - x^2}$, the area of the half-disk can be determined by integrating:

$$\begin{aligned}\int_{-1}^{1} \sqrt{1 - x^2}dx &= \frac{1}{2}\text{Area of the disk} \\ &= 2\frac{\text{Number of raindrops falling on the disc}}{\text{Number of raindrops falling on the entire square}}\end{aligned}$$

Thus, we have calculated this integral with the help of randomly falling raindrops.

Note that the quantity we want to calculated using random events is by no means random itself: the area of a disk of radius $1m^2$ never changes nor does the value of π or the value of the above integral. We could ask the question: does rain know anything about geometry and circles? Or about the natural constant π? Or about integration?

Probably not. However, this information can be obtained if we are clever enough to ask the right questions! The only condition, the randomly falling raindrops had to meet, was that they fall evenly, or more precisely, that the probability for them to fall on any particular point in the square is exactly the same as to fall on any other point in the square. For example, on a square field with an area of 1.000 square miles, it might be raining in some places and sunny in others. Such a field would be completely unsuitable for our experiment.

The "machine" used to generate the random events, whether the rain or a computer, is completely irrelevant. The events merely have to be "sufficiently random" and their probability distribution must be known. If this is the case, calculations can be performed with an accuracy which is limited solely by the machine's capacity to produce the random events: according to the laws of statistics, the accuracy of results improves as the number of random events involved in the simulation increases (see Section 29.3).

How would the above experiment have been conducted with a computer? Nowadays, most programming languages such as Pascal, Fortran, C, C++, etc. and the common spread sheet programs, such as Microsoft Excel (or Visual Basic) or Lotus 1-2-3, are equipped with *random number generators*. These are (small) programs which usually generate uniformly distributed *random numbers* between 0 and 1. Simulating a random event such as a falling raindrop with a computer is accomplished by generating *two* such random numbers Z_1 and Z_2 to produce the coordinates (one x- and one y-coordinate) of the point on which the raindrop falls. Since the simulated random numbers are all between 0 and 1 the generator producing such coordinates simulates raindrops falling only in the shaded area in Figure 12.1. To simulate raindrops falling on the entire square, the following transformation must be made:

$$x = 2Z_1 - 1 \quad , \quad y = 2Z_2 - 1$$

Because this transformation is linear, x and y are uniformly distributed random variables, as are Z_1 and Z_2. Thus, the coordinates of a raindrop in the square have been determined. The simulation of the random event is complete and we can continue with the evaluation. As in the case of raindrops, two events must be counted:

- The total number of "raindrops" falling in the square. This, however, is exactly the number of simulated random events, since no coordinates were generated which lie outside of the square (we make no unnecessary simulations). This means that this counter will be increased by one after each random event has been simulated.

- The number of "raindrops" falling within the circle. According to *Pythagoras*, these are the random events whose coordinates x and y satisfy the inequality $x^2 + y^2 \leq 1$. This means that after each random event, the counter is only increased by one if the generated event satisfies this inequality.

If , for example, 10,000 events were simulated (this requires the generation of 20,000 random numbers, one for each coordinate), of which, for example, 7851 satisfy the

above inequality, we obtain

$$\text{Area of the disk} \approx 0,7851 * 4m^2 = 3,1404m^2$$
$$\pi \approx 0,7851 * 4 = 3,1404$$
$$\int_{-1}^{1} \sqrt{1-x^2}dx \approx 0,7851 * 2 = 1,5702$$

Of course, these results are not exact, but the error involved can be determined as described in Section 29.3 and in principle, can be made arbitrarily small by increasing the number of simulated random events.

12.2 The General Approach to Monte Carlo Simulations

We can extract the general procedure for conducting a Monte Carlo simulation from the above example which will be summarized here to provide the reader with a "recipe" for performing such simulations. Each step will be will be explained by reference to its corresponding step in the above disk experiment as well as to the simulation of a random walk.

1. **Generate the random numbers required for a Monte Carlo step, usually from a uniform distribution between 0 and 1.**
   In the disk experiment, a Monte Carlo step corresponds to the falling of a simulated "raindrop". Two random numbers were required. To simulate a random walk, a Monte Carlo step represents a step in the random walk. One random number is required for each dimension of the space in which the random walk occurs.

2. **Transform the random numbers to generate numbers according to a desired distribution.**
   In the disk experiment, the uniformly distributed random numbers generated between 0 and 1 were transformed to yield uniformly distributed random numbers between -1 and 1. For a random walk, they are transformed to normally distributed random numbers. Section A.5 presents methods for doing this.

3. **Perform a Monte Carlo step using the random numbers generated.**
   In the case of the disk , this corresponds to the falling of a "raindrop". For the random walk, it consists of adding one step to the random walk.

4. **Repeat steps 1, 2 and 3 until the system reaches the state required for the proposed investigation.**
   For a random walk consisting of n steps, the required state is attained when n Monte Carlo simulations have been carried out. For the disk, the evaluation can take place after each single Monte Carlo step.

5. **Measurements: measure the system variables of interest.**
   In the case of the disk , we measure the number of "raindrops" falling within the disk . For a random walk, for example, the "end-to-end" distance could be measured.

6. **Repeat steps 1, 2, 3, 4 and 5 until enough systems for a statistical analysis have been generated**.
   For the disk , many "raindrops" must be simulated; for the price evolution of an underlying, many random walks must be generated.

7. **Final Analysis: compute the means of the measured variables and determine the statistical error.**
   For the disk , this was the ratio of the number of raindrops falling within the disk to the total number simulated. For a random walk it could be, for example, the mean length of all measured end-to-end vectors (or the square of their Euclidean norm).

12.3 Monte Carlo Simulation of Risk Factors

12.3.1 Simulation of the Evolution of a *Single* Risk Factor

On the basis of Equation 3.13, the time interval from t to T is divided into n intervals of length δt where n equals the number of steps. The price of a risk factor is to be simulated as a random walk over this time interval. The random walk equation 3.13 (or its equivalent form 3.20) holds for infinitesimal changes in $\ln(S)$ occurring in an infinitesimally small time span dt. The time step δt used for the simulation is not infinitesimally small, however. Therefore we do not use the stochastic PDE 3.13 (or 3.20) itself but its solution, Equation 3.25. Taking the logarithm on both sides of this solution yields[2]:

$$\ln\left(S(t+\,\delta t)\right) = \ln\left(S(t)\right) + \mu\,\delta t + \sigma X\sqrt{\delta t} \quad \text{with} \quad X \sim \mathrm{N}(0,1)$$

This is a *recursion*: Knowledge of $\ln(S)$ at time t enables the calculation of $\ln(S)$ at the next time point $t+\delta t$. To emphasize this point, we enumerate the time points, i.e. we introduce the following notation:

$$t_i = t + i\;\delta t \;\text{ where }\; i = 0,\,1,\,\ldots,\,n \quad \text{i.e.} \quad t_0 = t\;,\quad t_n = T$$

Denoting the i^{th} (standard normally distributed) random number by $X_i = X(t_i)$, we obtain the basis of the simulation of a risk factor whose behavior is governed by Equation 3.13 in terms of the notation just introduced:

$$\ln\left(S(t_i)\right) = \ln\left(S(t_{i-1})\right) + \mu\,\delta t + \sigma\,X_i\sqrt{\delta t}\quad,\quad i = 1,\,\ldots\,,\,n \tag{12.1}$$

[2]To be able to work with standard normally distributed random numbers we also used Equations 3.12 and 3.24 here. Those Equations say that the Wiener-Process δW has the same distribution as $\sqrt{\delta t}$ times a standard normally distributed random number:

$$\delta W \sim X\sqrt{\delta t} \quad \text{mit} \quad X \sim \mathrm{N}(0,1)$$

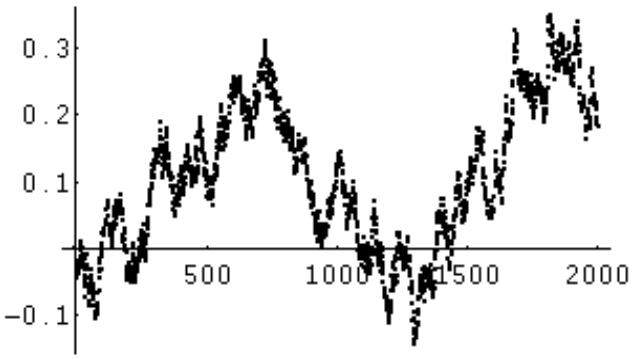


Figure 12.2: Simulation of $\ln(S(t)/S(0))$ over 2000 steps. 1 step corresponds to 1 day. The annual volatility is 20%. The drift is 0.

This equation is suitable for a direct simulation: $\ln(S(t_{i-1}))$ is the end-point of a random walk after $i-1$ steps. In the next step, the value "$\mu\,\delta t + \sigma\,X_i\sqrt{\delta t}$" will be added on to this end point, yielding the end point of the random walk after i steps, namely $\ln(S(t_i))$.

A concrete example at this point may clarify this procedure. With a volatility of $\sigma = 20\%$ per year, a drift of $\mu = 6\%$ per year and a time step $\delta t = 1$ day $= 1/365$ years, Equation 12.1 gives the following simple recursion relation:

$$\begin{aligned}\ln\left(S(t_i)\right) &= \ln\left(S(t_{i-1})\right) + \frac{0,06}{365} + 0,2\,\sqrt{1/356}\,X_i \\ &= \ln\left(S(t_{i-1})\right) + 0,0001644 + 0,01046\,X_i\end{aligned}$$

The contribution of the drift to each step is approximately one hundred times smaller than that of the volatility. This explains the negligible effect of the drift for small time spans $T-t$. However, over longer time periods, the drift cannot be ignored. Figure 12.3 shows the result of a simulation over a longer time span of $2,000$ days with respect to this recursion. To emphasize the effect of the drift, the $2,000$ randomly generated values of X_i were saved and used again for the recursion, this time setting $\mu = 0$. The result is presented in Figure 12.2. The values at the end of the simulation performed with the drift are approximately twice as large as those for the simulation with zero drift.

Such a curve represents one possible price progression over time, called a *path* of the risk factor. Repeating the simulation, we obtain an additional path. The simulation of many such paths yields the probability distributions for the price at each time point t_i in the simulated time span, in other words the probability distribution of the paths. In particular, we obtain a distribution of the values of the risk factor at the end point $t_n = T$. The simulated price of the risk factor at time $t_n = T$ is obtained by adding up the n steps generated in accordance with Equation 12.1:

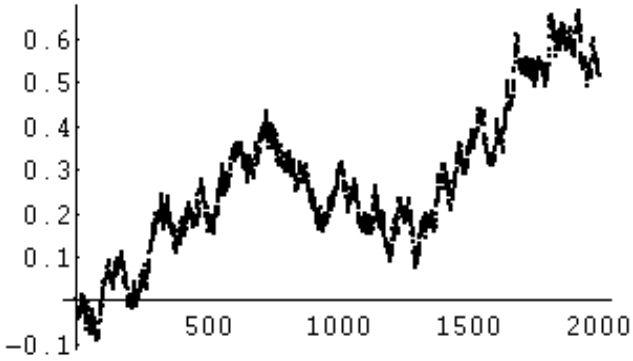


Figure 12.3: The same random walk as in Figure 12.2 but with a drift (mean return) of 6% per year.

$$\ln\left(S(t_n)\right) = \ln\left(S(t_0)\right) + \mu \sum_{i=1}^{n} \delta t + \sum_{i=1}^{n} X_i \, \sigma \sqrt{\delta t}$$

If the volatility σ is assumed to be constant, it can be factored out of the above sum. The length of the time steps δt are the same for each Monte Carlo step and thus

$$\ln\left(S(t_n)\right) = \ln\left(S(t_0)\right) + \mu \, n \, \delta t + \sigma \sqrt{\delta t} \sum_{i=1}^{n} X_i$$

The sum of n independent, standard normally distributed random variables is again a normally distributed random variable with expectation 0 and variance n, i.e. the standard deviation of the sum is $\sqrt{n}$. If only the end distribution (the distribution of S_n) is of interest and not the *path* of the underlying (each of the S_i) taken *during* the time span under consideration, the sum of the n standard normally distributed random variables X_i can be replaced by a standard normally distributed random variable X multiplied by $\sqrt{n}$:

$$\sum_{i=1}^{n} X_i \rightarrow \sqrt{n} \; X$$

In doing so, the simulated underlying price at the end of the time period under consideration can be generated with a *single* random number *directly*. Using $n\delta t = T - t$ from the definition of δt we can thus write

$$\ln\left(S(T)\right) = \ln\left(S(t)\right) + (T-t)\mu + \sigma \sqrt{T-t} \; X \tag{12.2}$$

The path taken by the underlying *during* the simulation is of interest only for exotic, *path-dependent* derivatives (see Chapter 19). Otherwise, in particular for pricing and risk management of European-style instruments where the underlying price is relevant

only at maturity (or at the end of the *liquidation period* in the case of risk management), the efficiency of a Monte Carlo simulation can be significantly improved by this simplification.

12.3.2 Simulation of *Several* Correlated Risk Factors

We are often interested in the progression of several risk factors rather than just one. The risk management of a portfolio, for example, requires the simulation of all risk factors affecting the portfolio. Since those are usually not statistically independent of one another, we are confronted with the question of how the *correlation* between risk factor processes can be incorporated into the simulation. The general approach for an arbitrarily large number of different securities is presented in Section 23.1. Here, we restrict the discussion to the important special case of *two* correlated prices. This can be used, for example, in determining the price of *exchange options*[3].

The two price processes S_1 and S_2 have drifts μ_1 and μ_2, volatilities σ_1 and σ_2, respectively, and a correlation ρ_{12}. The logarithm of the random walks will again be used to model the time evolution of the risk factors:

$$\begin{aligned}\delta \ln S_1(t) &= \mu_1\,\delta t \;+\; Y_1 \\ \delta \ln S_2(t) &= \mu_2\,\delta t \;+ Y_2 \quad \text{with correlated random variables} \quad Y_1,\, Y_2\end{aligned}$$

How should correlated pairs of random variables be constructed? First, the two equations for $\delta \ln(S_i)$ are combined by interpreting the indexed equations as components of a random vector.

$$\begin{pmatrix} \delta \ln S_1(t) \\ \delta \ln S_2(t) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \delta t + \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

If the prices were uncorrelated, the random variables Y_i would be independent, normally distributed random variables with variance $\sigma_i^2 \delta t$. However, since the prices are correlated, it is not sufficient to simply specify the variance of both variables in order to fully determine the distribution. Instead the *co*variance is needed to describe both the variances *and* the correlations. Since several Y_i (in this case, two) may come into play when constructing the random vector, the covariance is not a single number but a matrix, called the *covariance matrix*. As will be presented in detail in Section 21.4, the covariance matrix $\delta\mathbf{\Sigma}$ of two random variables is composed of the correlations and the standard deviations of the associated random variables as follows:

$$\delta\mathbf{\Sigma} = \begin{pmatrix} \delta\Sigma_{11} & \delta\Sigma_{12} \\ \delta\Sigma_{21} & \delta\Sigma_{22} \end{pmatrix} \quad \text{where} \quad \delta\Sigma_{ij} = \underbrace{\rho_{ij}}_{\substack{\text{Correlation} \\ \text{of } i \text{ with } j}} \underbrace{\sigma_i \sqrt{\delta t}}_{\substack{\text{Standard} \\ \text{dev. of } i}} \underbrace{\sigma_j \sqrt{\delta t}}_{\substack{\text{Standard} \\ \text{dev. of } j}} \text{ for } \; i, j = 1, 2 \tag{12.3}$$

Correlations are symmetric, i.e. $\rho_{ij} = \rho_{ji}$. Thus, $\rho_{ii} = 1$, in other words a risk factor is always fully correlated with itself. With $\rho = \rho_{12} = \rho_{21}$ the covariance matrix of two risk factors becomes

$$\delta\mathbf{\Sigma} = \delta t \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

[3] See Section 19.2.5.

In order to generate normally distributed random numbers Y_i with correlation matrix $\delta\mathbf{\Sigma}$ from independent, standard normally distributed random variables X_i, the "square root" $\mathbf{A}$ of the matrix $\delta\mathbf{\Sigma}$ is needed. This matrix satisfies the condition

$$\mathbf{A}\mathbf{A}^T = \delta\mathbf{\Sigma}$$

where $\mathbf{A}^T$ denotes the *transpose* of the matrix obtained by writing the column vectors of the matrix $\mathbf{A}$ as row vectors. As shown in detail in Section 21.4.2, this matrix yields the desired transformation

$\mathbf{Y} = \mathbf{A}\mathbf{X}$ where

$\mathbf{X}$ = Vector of standard normally distributed, uncorrelated random variables

$\mathbf{Y}$ = Vector of normally distributed, correlated random variables with covariance matrix $\delta\Sigma$

The "square root of a matrix" can be obtained using a procedure from linear algebra called the *Cholesky decomposition*. The general form of this decomposition is given in Section 21.4.2. Here, we restrict our consideration to 2×2 matrices, carrying out the procedure explicitly for this case. We begin by assuming that the matrix $\mathbf{A}$ has the following form:

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \Longrightarrow \mathbf{A}^T = \begin{pmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{pmatrix}$$

The components a_{ij} can now be determined from the requirement that the equation $\mathbf{A}\mathbf{A}^T = \delta\mathbf{\Sigma}$ be satisfied.

$$\mathbf{A}\mathbf{A}^T = \delta\mathbf{\Sigma}$$

$$\begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{pmatrix} = \begin{pmatrix} \delta\Sigma_{11} & \delta\Sigma_{12} \\ \delta\Sigma_{21} & \delta\Sigma_{22} \end{pmatrix}$$

$$\begin{pmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{11}a_{21} & a_{21}^2 + a_{22}^2 \end{pmatrix} = \begin{pmatrix} \delta\Sigma_{11} & \delta\Sigma_{12} \\ \delta\Sigma_{21} & \delta\Sigma_{22} \end{pmatrix}$$

Comparing the components on both sides yields a linear system of equations for the a_{ij}:

$$a_{11}^2 = \delta\Sigma_{11} \;\Rightarrow\; a_{11} = \sqrt{\delta\Sigma_{11}} = \sigma_1\sqrt{\delta t}$$

$$a_{11}a_{21} = \delta\Sigma_{12} \;\Rightarrow\; a_{21} = \frac{\delta\Sigma_{12}}{\sqrt{\delta\Sigma_{11}}} = \rho\ \sigma_2\sqrt{\delta t}$$

$$a_{21}^2 + a_{22}^2 = \delta\Sigma_{22} \;\Rightarrow\; a_{22} = \sqrt{\delta\Sigma_{22} - \frac{\delta\Sigma_{12}^2}{\delta\Sigma_{11}}} = \sqrt{1-\rho^2}\,\sigma_2\sqrt{\delta t}$$

Where Equation 12.3 was also used. Now that the matrix elements have been determined, this matrix can be used to generate the correlated random numbers:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \sqrt{\delta t} \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{(1-\rho^{2)}}\sigma_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

and thus

$$\begin{aligned} Y_1 &= \sqrt{\delta t}\ \sigma_1\, X_1 \\ Y_2 &= \sqrt{\delta t}\ \rho\,\sigma_2\, X_1 + \sqrt{\delta t}\ \sqrt{1-\rho^2}\,\sigma_2\, X_2 \end{aligned} \tag{12.4}$$

We can now formulate a random walk equation for two *correlated* price processes in terms of two *un*correlated *standard* normally distributed random variables X_1, X_2:

$$\begin{aligned} \delta \ln S_1(t) &= \mu_1\,\delta t \ +\ \sigma_1\, X_1 \sqrt{\delta t} \\ \delta \ln S_2(t) &= \mu_2\,\delta t \ +\ \sigma_2\,(\rho X_1 + \sqrt{1-\rho^2} X_2)\sqrt{\delta t} \end{aligned} \tag{12.5}$$

The first equation has the form of a single random walk. The correlation affects only the second equation. The interpretation of this representation is that the *second* price is correlated with the *first*. Since the correlations are symmetric, this interpretation is irrelevant for the final result. Assuming another form for the matrix **A** (or simply renaming the Y_i) would yield the result that the *first* price is correlated with the *second*. This holds in general: we can select any risk factor (for example, the one we feel most comfortable working with) as the leading factor and simulate it independently. Then all correlations appear in the evolutions of the other risk factors.

Exactly as in the previous section, the random walk equations 12.5 now provide the basis for the recursion which can be programmed in a Monte Carlo simulation:

$$\begin{aligned} \ln S_1(t_i) &= \ln S_1(t_{i-1}) + \mu_1\,\delta t \ +\ \sigma_1\, X_1(t_i)\sqrt{\delta t} \quad , \quad i = 1, \ldots, n \\ \ln S_2(t_i) &= \ln S_2(t_{i-1}) + \mu_2\,\delta t \ +\ \sigma_2 \left[\rho X_1(t_i) + \sqrt{1-\rho^2} X_2(t_i)\right] \sqrt{\delta t} \end{aligned} \tag{12.6}$$

Or, if only the values at the end of the time period under consideration are of interest but not the paths of the risk factors:

$$\begin{aligned} \ln S_1(T) &= \ln S_1(t) + \mu_1\,(T-t) \ +\ \sigma_1 \sqrt{T-t} X_1 \\ \ln S_2(T) &= \ln S_2(t) + \mu_2\,(T-t) \ +\ \sigma_2 \sqrt{T-t}\ \left[\rho X_1 + \sqrt{1-\rho^2} X_2\right] \end{aligned} \tag{12.7}$$

The Excel workbook MonteCarloDemo.xls can be found on the CD-ROM accompanying this book and includes a demonstration of the Monte Carlo simulation of a risk factor. For demonstration purposes, the simulation is accomplished using Excel cell functions without programming in Visual-Basic. Of course, this would be too slow for use in real applications. We therefore also provide a workbook entitled MonteCarloSimulation.xls on the CD-ROM which contains an executable Visual Basic module for a Monte Carlo simulation.

12.4 Pricing

According to Equation 11.20 the value at time t of a financial instrument can be determined from the expectation of its price at a future time T, if this expectation

is taken with respect to the risk-neutral probability distribution of the underlying. A clever choice of T can make pricing the instrument using Monte Carlo simulations quite easy. If T is chosen to be the maturity of the derivative, then the derivative's price at time T is simply given by the payoff profile P. Thus, if the underlying S is simulated according to Equation 12.1 up to time T we can easily obtain a *simulated* probability distribution for the payoff values. This works even for path dependent instruments if we measure the relevant path-dependent quantities along the way (like for instance averages of the simulated underlying values in the case of Asian options). The *mean* of all these simulated payoff values is then used as an estimator for the *expectation* of the payoff. According to Equation 11.20, discounting[4] this estimator back to time t yields an estimator for the value of the instrument at time t, *if* the calculation of the mean has been carried out with respect to the risk-neutral probability. To ensure that this is the case, the risk factor must be *simulated* with respect to this probability. This is accomplished by simply choosing the *drift* of the random walk to be risk-neutral in accordance with Equation 11.25.

Thus, the approach for pricing derivatives using Monte Carlo simulations is basically clear: We simulate the underlying (or more precisely the logarithm of the underlying value) in a risk-neutral world up to the time of maturity in accordance with Equation 12.1 (for path-*in*dependent derivatives we can even use Equation 12.2 and save a lot of computing time), and then measure the (discounted) mean of the payoff profile. In order to determine the error involved as discussed in Section 29.3, the mean of the square of the payoff profile must be measured, too. That's all. The calculation of option prices using the Monte Carlo method merely involves determining the mean of certain functions during the simulation.

An essential point, however, should not be overlooked. The time of exercise must be known, otherwise it is not clear "up to when" the price is to be simulated. This is why, as mentioned several times before, the Monte Carlo simulation presented here can be used to price the most exotic *European* options but is unsuitable for pricing American derivatives.

A detailed demonstration of the application of the Monte Carlo method to the valuation of an option portfolio is provided in the Excel workbook MonteCarloSimulation.xls. By making appropriate adjustments in the valuation part of the Visual Basic module, this program can be used to price all sorts of European derivatives[5] in the context of the Black-Scholes world (constant yields and volatility). In anticipation of Chapter 13, the calculation of *Greeks*, i.e. the sensitivities of the derivative's price with respect to its parameters, is also demonstrated in the workbook. The workbook can be used as a small but complete option calculator (as always, the yellow fields are the input fields).

[4] For *future styled* instruments, whose value changes are settled by daily adjustments in a margin account (as is the case for futures, for example), today's price is directly related to this expectation without any discounting, see Equations 7.5 and 11.21.

[5] See also Section 19.4.1.

13

Hedging

The replication of derivatives with a portfolio consisting of underlyings and a bank account as, for example, in Equation 11.3, can of course also be used to *hedge* the derivative's risk resulting from the stochastic movement of its underlying (or conversely a derivative could be used to hedge such a portfolio). This is accomplished by going short in the portfolio and long in the derivative or vice versa. This idea can be extended to hedging against influences other than the underlying price, for example, changes in the volatility, interest rate, etc. Such concepts of safeguarding against a risk factor have already made their appearance in arbitrage arguments in previous chapters and will be presented in their general form in this chapter. In addition to the fundamental Assumptions 1, 2, 3, 4 and 5 from Section 5, continuous trading will also be assumed below, i.e. Assumption 6. We will allow the underlying to perform a general Ito process[1] of the Form 3.16 and assume that it pays a dividend yield q.

13.1 Replicating Portfolios as Synthetic Derivatives

Consider first the change in the value V_S of a derivative resulting from the change in the price S of its underlying. V_S, as a function of the stochastic process S, is also a stochastic process. From Ito's lemma in the form of Equation 3.19 this process is explicitly

$$dV_S(S,t) = \left[\frac{\partial V_S}{\partial t} + \frac{1}{2}b(S,t)^2\frac{\partial^2 V_S}{\partial S^2}\right]dt + \frac{\partial V_S}{\partial S}dS(t) \tag{13.1}$$

13.2 Hedging Derivatives with Spot Transactions

In a similar manner as in Section 11, we will now construct a portfolio which replicates (hedges) the derivative in Equation 13.1. This portfolio will consist of a certain number Δ_s of underlyings and an amount g of money borrowed from or invested in the capital

[1] Thus, we are a more general here than Assumption 7 from Section 5.

market. This portfolio is constructed to have the same value as the derivative,

$$V_S = \Delta_S\, S + g \tag{13.2}$$

To ensure that the derivative is perfectly hedged for at least a short time span dt (for example, one step in a binomial model), we require the change in its value over dt to be the same as that of the derivative. The change in value of the derivative is given by Equation 13.1. The value of the portfolio consisting of money and the underlying changes as already described in Equation 11.3, for example. For small time intervals dt, B_r and $1/B_r$ approach $(1 - r\ \ dt)$ and $(1 + r\ dt)$, respectively, for all compounding conventions. The same approximation holds for B_q. Denoting the change in S by dS, and the number of underlyings in the portfolio by Δ_S, to emphasize the fact the we are hedging with respect to the spot price, the change in the value of the portfolio can be expressed by

$$\begin{aligned} d(\Delta_S\, S + g) &= \underbrace{\Delta_S(S + dS)(1 + q\, dt) + g(1 + r\, dt)}_{\text{New Value}} - \underbrace{(\Delta_S\, S + g)}_{\text{Old Value}} \\ &= \Delta_S\, dS + (\Delta_S\, q\, S + r\, g)dt + \Delta_S\, q\, dS\, dt \\ &= \Delta_S\, dS + (\Delta_S\, q\, S + r\, g)dt + \ \cdots \end{aligned} \tag{13.3}$$

where in the last line only the first order terms in dS and dt are considered. To ensure that this change corresponds exactly to a change in the value of the derivative, the coefficients of dt and dS must be the same as those appearing in Equation 13.1. Equating the coefficients of dS yields the well known result that the number of underlyings required for the hedge, called the *hedge ratio*, must be equal to the sensitivity of the derivative with respect to its underlying. Combined with Equation 13.2 we can determine the amount of money in the bank account required for the hedge as well:

$$\Delta_S(t) = \frac{\partial\, V_s}{\partial\, S(t)} \ \Rightarrow\ g(t) = V_s - S(t)\frac{\partial\, V_s}{\partial\, S(t)} \tag{13.4}$$

Thus, with information on the value V_s and its first partial derivative with respect to the underlying, it is possible to construct a portfolio hedging the derivative *perfectly* over the next time step. The partial derivatives of the price function are often referred to as the *sensitivities* (see Section 13.5).

We have yet to exploit the second condition, which is the equality of the coefficients of dt in Equations 13.1 and 13.3. This yields

$$\Delta_S S(t)q + gr = \frac{\partial V_S}{\partial t} + \frac{1}{2}b(S,t)^2\frac{\partial^2 V_S}{\partial S^2}$$

Inserting Equation 13.4 for Δ_S and g, we recognize the famous differential equation of Black and Scholes for derivatives on the spot price S on an underlying:

$$\frac{\partial\, V_S}{\partial\, t} + (r - q)S\frac{\partial\, V_S}{\partial\, S} + \frac{1}{2}b(S,t)^2\frac{\partial^2 V_S}{\partial\, S^2} = r\, V_S \tag{13.5}$$

If we write (without loss of generality) the parameters a and b as in Equation 8.1 we immediately see that Equation 13.5 is exactly equal to Equation 8.8.

As all other derivation of the Black-Scholes PDE, this most simple derivation of the *Black-Scholes differential equation*, does not specify the derivative under consideration in any way. No information about the payoff profile or any other property characterizing the derivative is necessary. Only the properties of the *underlying* are used, these being that the underlying earns a dividend yield q and that its price is governed by a stochastic process of the form 3.16. Thus, there is only *one* equation for the construction of the hedging portfolio and *one* differential equation for the price of *all* derivatives on this underlying. The different derivative instruments can be distinguished from one another solely on the basis of their respective initial and/or boundary conditions (for example, the payoff profile of an option at maturity or the fact that a forward contract is worth nothing at the time when the contract is entered into).

13.2.1 Forwards and Futures as Derivatives

Equations 13.4 and 13.5 will now be applied to forward and futures contracts. These contracts will be interpreted as derivatives which will be hedged with spot transactions, i.e. through the spot price of the underlying. To allow the calculations to be performed explicitly, we adopt the convention of continuous compounding.

Consider first a forward contract. The value of a forward with delivery price K, assuming continuous compounding, is given by Equation 7.4 as

$$f_S(t,T,K) = [S(t,T) - K]\, B_r(t,T) = e^{-q(T-t)}S(t) - e^{-r(T-t)}K$$

The partial derivatives of this function appearing in Equations 13.4 and 13.5 are

$$\frac{\partial f}{\partial t} = q\, e^{-q(T-t)}S - r\, e^{-r(T-t)}K \quad , \quad \frac{\partial f}{\partial S} = e^{-q(T-t)} \quad , \quad \frac{\partial^2 f}{\partial S^2} = 0$$

Thus, Equation 13.4 yields

$$\Delta_S(t) = e^{-q(T-t)} \quad , \quad g(t) = -e^{-r(T-t)}K$$

In order to hedge a (short) forward, $e^{-q(T-t)}$ underlyings must be purchased and an amount $e^{-r(T-t)}K$ of cash must be raised. Substituting the value of the forward and its derivative into Equation 13.5 shows that the Black-Scholes differential equation is satisfied:

$$q\, e^{-q(T-t)}S - r\, e^{-r(T-t)}K + (r-q)Se^{-q(T-t)} = re^{-q(T-t)}S - re^{-r(T-t)}K$$

Now consider a future. The value of a future with delivery price K, assuming continuous compounding, is given by Equation 7.5 as

$$F_S(t,T,K) = S(t,T) - K = e^{(r-q)(T-t)}S(t) - K$$

The partial derivatives of the price function of the future are

$$\frac{\partial F}{\partial t} = -(r-q)\, e^{(r-q)(T-t)}S \quad , \quad \frac{\partial f}{\partial S} = e^{(r-q)(T-t)} \quad , \quad \frac{\partial^2 f}{\partial S^2} = 0$$

Substituting into Equation 13.4 yields

$$\Delta_S(t) = e^{(r-q)(T-t)} \quad , \quad g(t) = -K$$

Thus, in order to hedge a (short) future, $e^{(r-q)(T-t)}$ underlyings must be purchased and a cash amount K must be raised. The value of the futures position and its derivatives substituted into Equation 13.5 show that the Black-Scholes differential equation is generally *not* satisfied:

$$\underbrace{-(r-q)\,e^{(r-q)(T-t)}S + (r-q)Se^{(r-q)(T-t)}}_{=0} = r(e^{(r-q)(T-t)}S - K)$$

While the left-hand side is identically equal to zero, the right-hand side of the equation (the value of the future multplied by r) is only equal to zero when the value of the future is zero, i.e. when the delivery price is equal to the forward price.

The price functions of all futures contracts whose delivery price is not equal to the current forward price *fail* to satisfy the Black-Scholes differential equation!

This is because the concept of variation margins, enforced by the futures exchanges, introduces an inconsistency: the variation margin has the effect that the daily changes in the value of a futures position *must immediately and without discounting* be paid into respectively credited to the variation account, *although these cash flows are actually due at the maturity of the future!* This inconsistency is responsible for the future's failure to satisfy the equation which, theoretically, must be satisfied by the price function of all derivatives on a particular underlying. The mistake of not discounting the variation margin can only be neglected if the balance of the variation margin is zero. This is precisely the case when the current forward price is equal to the delivery price. As we have already seen, the differential equation is satisfied in this case. This is a typical example of how a market convention can cause an inconsistency. Many such examples can be found in the financial world.

A trader intending to use forwards or futures to hedge a portfolio consisting of underlyings and a bank account, however, is interested in precisely *that* point in time when the contract is concluded. But at the time when the contract is concluded the differential equation is satisfied for both forwards *and* futures.

13.3 Hedging Derivatives with Forward Contracts

The portfolio for the synthetic derivative is now to be constructed using *forward* contracts instead of spot transactions. We cannot merely replace the spot price with the forward price in Equation 13.3 since the forward price itself is not tradable. Forward and futures contracts on the other hand are tradable.

13.3.1 Hedging with Forwards

Since futures satisfy the Black Scholes differential equation only upon conclusion of the contract, we will avoid this difficulty at this point and begin by using forwards to construct the synthetic derivative. The hedging portfolio is then

$$\Delta_f(t) f_S(t, T', K) + g(t) = \Delta_f(t) \left[S(t, T') - K\right] B_r(t, T') + g(t) \quad \text{with} \quad T' \geq T > t$$

where the subscript f indicates that the hedge is accomplished with forwards. T denotes the maturity of the derivative and T' the maturity of the forward contract. Note that the condition $T' \geq T$ is not strictly necessary for our considerations. But it is convenient since it ensures that the hedge does not "vanish" before the maturity T of the derivative. But $T' < T$ is not a problem as long as we can roll over the hedge into another forward contract (with a later maturity) at time T' or earlier.

In the time span dt this portfolio changes its value as a result of changes in the forward price and the discount factor as follows:

$$\begin{aligned} d(\Delta_f f + g) &= \underbrace{\Delta_f \left[S(t, T') + dS(t, T') - K\right] B_r(t, T')(1 + r dt) + g(t)(1 + r dt)}_{\text{New Value}} \\ &\quad - \underbrace{\Delta_f \left[S(t, T') - K\right] B_r(t, T') + g(t)}_{\text{Old Value}} \\ &= \Delta_f B_r(t, T') dS(t, T') + \Delta_f \left[S(t, T') - K\right] B_r(t, T') r\, dt + g r\, dt + \cdots \\ &= \Delta_f B_r(t, T') \frac{\partial S(t, T')}{\partial S(t)} dS(t) + \left[\Delta_f(t) f_S(t, T', K) + g\right] r\, dt + \cdots \end{aligned} \tag{13.6}$$

where again we consider only changes which are of linear order in dt or dS. In the last equation, the change in the forward price $dS(t, T')$ was rewritten in terms of a change in the spot price to facilitate the comparison of the coefficients with the associated coefficients in Equation 13.1 for the change in the price of the derivative. Equating the coefficients of dS in the above mentioned expressions now yields:

$$\Delta_f(t) = \frac{1}{B_r(t, T')} \frac{\partial V_S / \partial S(t)}{\partial S(t, T') / \partial S(t)} = \frac{1}{B_r(t, T')} \frac{\partial V_S}{\partial S(t, T')}$$

Using Equation 7.1 for the dividend yield, we can calculate the derivative of the forward price with respect to the spot price and thus obtain the following expression for the hedge ratio:

$$\Delta_f(t) = \frac{1}{B_q(t, T)} \frac{\partial V_S}{\partial S(t)} = \frac{1}{B_r(t, T')} \frac{\partial V_S}{\partial S(t, T')}$$

In other words, the number of forwards required to hedge the derivative is equal to the sensitivity of the derivative with respect to the spot price compounded at the dividend yield up to maturity of the derivative, or equivalently, is equal to the sensitivity with respect to the forward price compounded at the risk-free rate up to maturity of the hedging instrument (the forward).

Since by construction, the value of the replicating portfolio is equal to the value of the derivative we can determine the amount of cash needed for the hedge:

$$g(t) = V_s - f_S(t, T', K)\, \Delta_f(t) \tag{13.7}$$

We have thus completely determined the hedging portfolio using solely information derived from the price of the derivative and its sensitivities.

Before we establish a mathematical expression obtained from comparison of the coefficients of dt in Equations 13.1 and 13.6, we change the coordinates in Equation 13.1 to transform the second derivative of the price with respect to *spot* price $S(t)$ into a derivative with respect to the *forward* price $S(t,T')$. This is accomplished using 7.1.

$$\begin{aligned}\frac{\partial}{\partial S(t)} &= \frac{\partial S(t,T')}{\partial S(t)}\frac{\partial}{\partial S(t,T')} = \frac{B_q(t,T')}{B_r(t,T')}\frac{\partial}{\partial S(t,T')} \\ \frac{\partial^2}{\partial S(t)^2} &= \frac{\partial}{\partial S(t)}\frac{B_q(t,T')}{B_r(t,T')}\frac{\partial}{\partial S(t,T')} = \frac{B_q(t,T')^2}{B_r(t,T')^2}\frac{\partial^2}{\partial S(t,T')^2}\end{aligned} \tag{13.8}$$

Equating now the coefficient of dt and using Equation 13.7 for g yields

$$\frac{\partial V_S}{\partial t} + \frac{1}{2}b(S,t)^2\frac{B_q(t,T')^2}{B_r(t,T')^2}\frac{\partial^2 V_S}{\partial S(t,T')^2} = rV_s$$

This is the Black-Scholes differential equation for derivatives on the forward price of an underlying.

To put this in a more familiar form we rewrite (without loss of generality) the parameter b as in Equation 8.1 and use Equation 7.1 to introduce the forward price into the parameter b, i.e.

$$b(S,t) =: \sigma(S,t)S(t) = \sigma(S,t)\frac{B_r(t,T')}{B_q(t,T')}S(t,T') \tag{13.9}$$

With this substitution we arrive at the well-known form of the Black-Scholes differential equation for derivative on the forward price of an underlying:

$$\frac{\partial V_S}{\partial t} + \frac{1}{2}\sigma(S,t)^2 S(t,T')^2\frac{\partial^2 V_S}{\partial S(t,T')^2} = rV_s \tag{13.10}$$

As in Section 13.2, only *one* equation is needed for the construction of the hedging portfolio with forward contracts, and there is only *one* differential equation for *all* derivatives on the forward price. Different derivatives can be distinguished from one another only through their respective initial and/or boundary conditions.

Note the absence of a first derivative term with respect to the forward price. Equation 13.5 (which is the corresponding Black-Scholes PDE for derivatives on the *spot* price of an underlying) would look quite similar to Equation 13.10 for an underlying which earns a dividend yield q being exactly equal to the risk free rate r. In fact, both equations would look identical if this underlying would have $S(t,T')$ as its *spot* price. Herein lies the deeper explanation for the "recipe" for pricing options on futures as given in Equation 18.4.2: a derivative on the forward price can be priced as a derivative on the spot price of an (artificial) underlying whose spot price is equal to $S(t,T')$ and whose dividend yield equals the risk-free rate.

13.3.2 Hedging with Futures

Although futures fail to satisfy the Black-Scholes differential equation (except at the time when the contract is concluded) it is still legitimate to ask how many futures are required to hedge a derivative. If futures are used to construct a synthetic derivative, the value of the hedging portfolio is given by[2]

$$\Delta_F(t)F_S(t,T',K) + g(t) = \Delta_F(t)\left[S(t,T') - K\right] + g(t) \quad \text{for} \quad T' \geq T > t$$

Changes in the forward price and the bank account in the time interval dt cause the value of the portfolio to change as follows

$$\begin{aligned} d(\Delta_F F_S(t,T',K) + g) &= \underbrace{\Delta_F\left[S(t,T') + dS(t,T') - K\right] + g\,(1+rdt)}_{\text{New Value}} \\ &\quad - \underbrace{\Delta_F\left[S(t,T') - K\right] + g}_{\text{Old Value}} \\ &= \Delta_F\, dS(t,T') + gr\,dt + \cdots \\ &= \Delta_F\,\frac{\partial S(t,T')}{\partial S(t)} dS(t) + gr\,dt + \cdots \end{aligned} \tag{13.11}$$

where again we consider only the parameter changes of linear order in dt or dS, neglecting all higher order terms. In the last equation, the change in the forward price has been replaced by the corresponding expression with respect to a change in the spot price, facilitating a direct comparison with the coefficients in Equation 13.1 for the change in the price of the derivative. Setting the coefficients of dS in both expressions equal yields

$$\Delta_F(t) = \frac{\partial V_S/\partial S(t)}{\partial S(t,T')/\partial S(t)} = \frac{\partial V_S}{\partial S(t,T')}$$

Equation 7.1 for the forward price of an underlying earning a dividend yield q gives the partial derivative of V_S with respect to the forward price in terms of the partial derivative with respect to the spot price.

$$\Delta_F(t) = \frac{B_r(t,T')}{B_q(t,T')}\frac{\partial V_S}{\partial S(t)} = \frac{\partial V_S}{\partial S(t,T')} \tag{13.12}$$

In other words, the number of futures required to hedge a derivative is equal to the sensitivity of the derivative with respect to the *spot* price compounded at the rate obtained by taking the difference in the risk-free rate and the dividend yield. Or equivalently, the number of futures needed equals the sensitivity with respect to the *forward* price.

By definition, the value of the hedging portfolio must be equal to the value of the derivative. It follows that the amount of money needed for the hedge is given by

$$g(t) = V_s - F_S(t,T',K)\Delta_F(t) \tag{13.13}$$

[2]Here again the condition $T' \geq T$ is not strictly necessary but convenient since it ensures that the hedge does not "vanish" before maturity T of the derivative. But again, $T' < T$ would not be a problem as long as we can roll over the hedge into another futures contract (with a later maturity) at time T' or earlier.

We have thus completely determined the hedging portfolio using solely the information derived from the value of the derivative instrument and its sensitivities.

13.3.3 The Differential Equation for Derivatives on Futures

As was seen in Section 13.2.1, futures do *not* satisfy the Black-Scholes differential equation. Nevertheless, a differential equation can be derived from the condition requiring that the market be arbitrage free, in particular, that a portfolio can be found which replicates derivatives on futures. Such derivatives are financial instruments whose *underlying* is a specific *future contract*. The underlying is not a forward or spot *price* but a future *contract* with a specific delivery price K and maturity T'.

Equating the coefficients of dt in Equations 13.1 and 13.11 yields

$$\begin{aligned}\frac{\partial V_S}{\partial t} + \frac{1}{2}b(S,t)^2\frac{\partial^2 V_S}{\partial S^2} &= gr \\ &= rV_s - rF_S(t,T',K)\frac{B_r(t,T')}{B_q(t,T')}\frac{\partial V_S}{\partial S} \\ &= rV_s - r\left[S(t,T') - K\right]\frac{B_r(t,T')}{B_q(t,T')}\frac{\partial V_S}{\partial S} \\ &= rV_s - r\left[\frac{B_q(t,T')}{B_r(t,T')}S(t) - K\right]\frac{B_r(t,T')}{B_q(t,T')}\frac{\partial V_S}{\partial S}\end{aligned}$$

where in the second line we used Equations 13.13 and 13.12 for g and $\Delta_F(t)$, respectively. In the third line we inserted the value 7.5 of a futures position and in the last line we used Equation 7.1 for the forward price. Thus the Black-Scholes differential equation for derivatives on futures is

$$\frac{\partial V_S}{\partial t} + r\left[1 - \frac{B_r(t,T')}{B_q(t,T')}\frac{K}{S(t)}\right]S(t)\frac{\partial V_S}{\partial S(t)} + \frac{1}{2}b(S,t)^2\frac{\partial^2 V_S}{\partial S(t)^2} = rV_s \tag{13.14}$$

The more natural variable for futures derivatives is certainly the *forward* price and not the spot price. Replacing the partial derivatives with respect to the spot price by those taken with respect to the forward price as in Equation 13.8 and replacing $b(S,t)$ as in Equation 13.9, all on the basis of 7.1, we obtain the following form for the Black-Scholes differential equation for the value of derivatives on futures

$$\frac{\partial V_S}{\partial t} + r\left[1 - \frac{K}{S(t,T')}\right]S(t,T')\frac{\partial V_S}{\partial S(t,T')} + \frac{\sigma^2}{2}S(t,T')^2\frac{\partial^2 V_S}{\partial S(t,T')^2} = rV_s \tag{13.15}$$

This reduces to the Black-Scholes differential equation for derivatives on *forwards*, Equation 13.10, *only if* the delivery price K is equal for the current forward price $S(t)B_q/B_r$. Again, this is because, for $S(t,T') = K$, the variation margin disappears and with it the reason for the future's failure to satisfy the Black-Scholes differential equation.

The difference between the Black-Scholes differential equation and the differential equation governing the price of futures derivatives receives little mention in the literature[3]. The reason for this is probably that – as mentioned at the beginning of this section – the difference between Equation 13.15 and the Black-Scholes PDE 13.10 only arises for derivatives whose underlying is a *specific* futures *contract* with a *specific* delivery price K and maturity T'. The derivatives on futures actually traded, however, (such as plain vanilla calls and puts on futures whose prices have already been given in Equation 18.4.2 for the Black-Scholes world), do *not* have a specific futures contract with a specific fixed delivery price K as their underlying. They always have the *current* future as their underlying, i.e. the future with delivery price K equal to the current forward price $S(t, T')$. This has already been discussed for options on futures at the beginning of section 9.3.2: when exercising the option the exerciser goes long (in the case of a call) or short (in the case of a put) in the *current* futures contract with the then valid forward price as the delivery price so that the futures contract – as always – has zero value when entered into[4]. But if the current future is the underlying, i.e. for $K = S(t, T')$, then Equation 13.15 is identical to the Black-Scholes PDE 13.10.

13.4 Hedge-Ratios for Arbitrary Combinations of Financial Instruments

An analogous approach to those described above can be taken to derive a formula for the number of an *arbitrary* hedge instrument required to replicate an *arbitrary* derivative. The number of hedging instruments is always equal to the sensitivity of the derivative with respect to this hedging instrument. Intuitively, if the value h of a hedge instrument changes by the amount dh, then the value V of the derivative changes by the amount dV. The quotient dV/dh is precisely the number of hedging instruments needed to compensate for the change dV in the derivative. For very small changes, this quotient approaches the differential quotient. Any instrument can assume the role of either the derivative or the hedging instrument. Since the value obtained by differentiating a derivative with respect to an arbitrary hedging instrument cannot, in general, be calculated directly, a common reference variable, usually the spot price $S(t)$, is introduced with respect to which both the partial derivative of the value as well as the partial derivative of the hedging instrument is known. The equation for this general *hedge ratio* is therefore

$$\Delta_{\substack{\text{Derivative } V \text{ with} \\ \text{Hedging Instrument } h}} = \frac{\partial V}{\partial h} = \frac{\partial V}{\partial S}\left[\frac{\partial h}{\partial S}\right]^{-1} \tag{13.16}$$

Consequently, the sensitivities of all financial instruments with respect to the spot price must be available if the hedge ratio is to be calculated for an *arbitrary* combination of these instruments. The sensitivities already dealt with in the last section, i.e. the

[3] In fact, to the best knowledge of the author, the above two equations for futures-derivatives have been published for the first time in this text.

[4] In addition the exerciser receives payment of the value of the futures contract which has the option's strike K as its delivery price.

Underlying (Spot)	1	
Future	$B_q(t,T)B_r(t,T)^{-1}$	
Forward	$B_q(t,T)$	
	Call	**Put**
Option on Spot Price	$B_q(t,T)\,\mathrm{N}(x)$	$-B_q(t,T)\,\mathrm{N}(-x)$
Option on a Future	$B_q(t,T')\frac{B_r(t,T)}{B_r(t,T')}\mathrm{N}(x')$	$-B_q(t,T')\frac{B_r(t,T)}{B_r(t,T')}\mathrm{N}(-x')$
Option on a Forward	$B_q(t,T')\mathrm{N}(x')$	$-B_q(t,T')\mathrm{N}(-x')$

Table 13.1: Sensitivities for options and forward contracts with respect to the spot price $S(t)$. x and x' are as defined in Equations 9.7 and 18.4.2.

sensitivities of the option, forward and spot transactions with respect to the spot price are listed in Table 13.1 (for the Black-Scholes world). For forward contracts, they follow from Equations 7.1, 7.4 and 7.5, which were derived using arbitrage arguments. For options, the sensitivities follow directly from the Black-Scholes equations 9.7, 18.4.2 and 9.9.

Readers wishing to reproduce these results themselves are advised to use the following property[5] of $\mathrm{N}'(x-\sigma\sqrt{T-t})$:

$$\mathrm{N}'(x-\sigma\sqrt{T-t}) = \frac{B_q(t,T)\,S(t)}{B_r(t,T)\,K}\mathrm{N}'(x) \tag{13.17}$$

Here, N' denotes the derivative of the cumulative standard normal distribution with respect to its argument:

$$\mathrm{N}'(y) \equiv \frac{d}{dy}\mathrm{N}(y) = \frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}$$

The entries in Table 13.1 together with Equation 13.1 provide all the information needed to determine the hedge ratio for every conceivable combination of these financial instruments in the Black-Scholes world. We now present three examples to illustrate this.

a) Hedging a put on the spot price with futures:

$$\begin{aligned}\Delta_{\substack{\text{Put on spot}\\ \text{with future}}} &= \frac{\partial V}{\partial S}\left[\frac{\partial h}{\partial S}\right]^{-1} = -B_q(t,T)\,\mathrm{N}(-x)\left[\frac{B_q(t,T')}{B_r(t,T')}\right]^{-1}\\ &= -B_r(t,T')\frac{B_q(t,T)}{B_q(t,T')}\mathrm{N}(-x)\end{aligned}$$

[5] To prove Equation 13.17 we first write

$$\mathrm{N}'(x-\sigma\sqrt{T-t}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\sigma\sqrt{T-t})^2} = \underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}_{\mathrm{N}'(x)}e^{x\sigma\sqrt{T-t}-\frac{1}{2}\sigma^2(T-t)}$$

The relation is now obtained by substituting the definition of x, Equation 9.5, into the second exponential term.

A long put on the spot can thus be *replicated* by a *short* position consisting of

$$B_r(t,T')\mathrm{N}(-x)/B_q(T,T'|\,t)$$

futures. Or interpreted as a hedge, a long put can be *hedged* by a *long* position in this number of futures.

b) Hedging a put on a future with forward contracts:

$$\begin{aligned}\Delta_{\substack{\text{Put on future}\\ \text{with forward}}} &= \frac{\partial V}{\partial S}\left[\frac{\partial h}{\partial S}\right]^{-1} = -B_q(t,T')\frac{B_r(t,T)}{B_r(t,T')}\mathrm{N}(-x')\left[B_q(t,T')\right]^{-1}\\ &= -\frac{B_r(t,T)}{B_r(t,T')}\mathrm{N}(-x')\end{aligned}$$

A long put on the future can be *replicated* with a *short* position of

$$\mathrm{N}(-x')/B_r(T,T'|\,t)$$

forwards. Or interpreted as a hedge, a long put can be *hedged* by a *long* position in this number of forwards.

c) Hedging a call on the spot with puts on a future:

$$\begin{aligned}\Delta_{\substack{\text{Call on spot with}\\ \text{put on future}}} &= \frac{\partial V}{\partial S}\left[\frac{\partial h}{\partial S}\right]^{-1} = B_q(t,T)\,\mathrm{N}(x)\left[-B_q(t,T')\frac{B_r(t,T)}{B_r(t,T')}\mathrm{N}(-x')\right]^{-1}\\ &= -\frac{B_r(t,T')}{B_r(t,T)}\frac{B_q(t,T)}{B_q(T,T')}\frac{\mathrm{N}(x)}{\mathrm{N}(-x')}\end{aligned}$$

A long call on the spot can thus be *replicated* by a *short* position in

$$\frac{B_r(T,T'|\,t)\mathrm{N}(x)}{B_q(T,T'|\,t)\mathrm{N}(-x')}$$

puts on the future. Or interpreted as a hedge, the long call can be *hedged* by a *long* position in this number of puts on the future.

13.5 “Greek” Risk Management with Sensitivities

13.5.1 Sensitivities and a Portfolio’s Change in Value

The *sensitivity* of a financial instrument with respect to a parameter is defined as a variable that, when multiplied by a (small) parameter change, yields the change in the value of the instrument resulting from this parameter change.

Price change equals sensitivity multiplied by parameter change

The sensitivity is thus the derivative of the price with respect to the parameter. Since options depend on several such parameters such as the price of the underlying, the

Symbol	Name	Definition	Interpretation
Δ	Delta	$\frac{\partial V}{\partial S}$	Price change resulting from a change in the underlying spot price
Γ	Gamma	$\frac{\partial^2 V}{\partial S^2}$	Change in the delta resulting from a change in the underlying spot price
Ω	Omega	$\frac{S}{V}\frac{\partial V}{\partial S}$	Relative price change (in %) resulting from a relative change (in %) in the underlying spot price
Ψ	Vega	$\frac{\partial V}{\partial \sigma}$	Price change resulting from a change in the volatility of the underlying
Θ	Theta	$\frac{\partial V}{\partial t}$	Price change resulting from a change in the time to maturity
ρ	Rho	$\frac{\partial V}{\partial r}$	Price change resulting from a change in the risk-free rate
ρ_q	Rho_q	$\frac{\partial V}{\partial q}$	Price change resulting from a change in the dividend yield of the underlying

Table 13.2: Definitions of the "Greeks".

time to maturity, the volatility, the dividend yield of the underlying and the risk-free market rate, the sensitivities listed in Table 13.2 are of particular interest.

The first of these sensitivities, namely Δ, has already been used in Section 13.2 to construct a replicating portfolio. In Equation 13.16 the hedge ratio Δ was defined in greater generality as the sensitivity of the derivative with respect to the price change of the hedging instrument. In the special case where the hedging instrument is the underlying itself, this general hedging ratio and the delta listed in Table 13.2 agree.

Any method used to calculate the price of an option can be used to calculate the sensitivities, even if the explicit expressions for the sensitivities cannot be obtained directly. The price model must simply be applied twice: the first time to calculate the price with the current parameters and a second time for a valuation with one of the parameters slightly changed. The difference between these two option values divided by the parameter change yields an approximation for the sensitivity of the option with respect to that parameter. Approximating the second derivative (as for gamma, for example) requires the price to be calculated with respect to three different parameter values. In summary: as soon as a pricing method is available we can calculate sensitivities by approximating the differential quotients in table 13.2 with difference quotients as follows:

$$\begin{aligned}
\text{Delta} &\approx \frac{V(S+\delta S)-V(S)}{\delta S}, \quad \text{Gamma} \approx \frac{V(S+\delta S)-2V(S)+V(S-\delta S)}{\delta S^2} \\
\text{Vega} &\approx \frac{V(\sigma+\delta\sigma)-V(\sigma)}{\delta\sigma}, \quad \text{Theta} \approx \frac{V(t+\delta t)-V(t)}{\delta t} \\
\text{Rho}_r &\approx \frac{V(r+\delta r)-V(r)}{\delta r}, \qquad \text{Rho}_q \approx \frac{V(q+\delta q)-V(q)}{\delta q}
\end{aligned} \tag{13.18}$$

The risk of an instrument (or a portfolio) can be measured by its reaction to changes

in the parameters influencing its price. In other words, the more sensitive it is to parameter changes, the "riskier" the instrument is. For this reason, the sensitivities are also known as *risk ratios.* A commonly used method of risk management is to control these sensitivities, i.e. to set limits on or targets for their values, for instance. The possibilities and limitations of this kind of risk management become clear, when we consider the Taylor series representation for small parameter changes: The value V depends on all of the above listed parameters: $V = V(t, S, \sigma, q, r)$. A change in value is thus

$$\begin{aligned} dV &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{\partial V}{\partial \sigma}d\sigma + \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial q}dq + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}dS^2 + \cdots \\ &= \Theta\, dt + \Delta\, dS + \Psi\, d\sigma + \rho\, dr + \rho_q\, dq + \frac{1}{2}\Gamma\, dS^2 + \cdots \end{aligned}$$

The coefficients of all linear parameter changes are given by sensitivities. A linear approximation holds only for small parameter changes. This is especially true for highly non-linear instruments. Because of the particularly strong influence of the underlying price, the effect of its quadratic change has been included in the above expression resulting in the appearance of the sensitivity Γ.

All risk variables in Table 13.2 (with the exception of the variable Ω) have the important property that the sensitivity of a portfolio with respect to a risk factor is equal to the sum of the sensitivities of its component instruments to this risk factor, i.e. the sensitivities are *linear.* For a portfolio consisting of n instruments of type A and m of type B we can therefore write

$$\Delta(n\,A + m\,B) = n\,\Delta(A) + m\,\Delta(B) \quad \text{likewise for } \Gamma, \Theta, \Psi, \rho, \rho_q \tag{13.19}$$

This property makes it possible to make a portfolio *delta-neutral*, for example: given n, $\Delta(A)$ and $\Delta(B)$, it is possible to make the portfolio's delta, $\Delta(nA+mB)$, equal to zero, by choosing $m = -n\Delta(A)/\Delta(B)$. This is called a *delta hedge.* The portfolio is then (in linear approximation) insensitive to small changes in the price of the underlying S. In the special case that the instrument B is the underlying itself we have $\Delta(B) = 1$. This is reflected in the result given in Section 13.2 that the number of underlyings needed to hedge a particular instrument is obtained by taking the derivative of the value of the instrument with respect to the underlying. We can now observe that the hedge methods introduced in Section 13.1 were delta hedges and represent only in linear approximation a safeguard against the risk resulting from a change in the underlying.

Likewise, a portfolio can be constructed which is gamma, theta, vega or rho neutral. Even *several* sensitivities of the portfolio can be neutralized *simultaneously* but more than one hedging instrument is then required to do so. In general, we need at least as many *different* hedging instruments as the number of sensitivities to be neutralized. For example, if a portfolio consisting of n instruments of type A is to be delta and gamma hedged, we would require two hedging instruments B_1 and B_2. The condition that the delta and the gamma of the portfolio be equal to zero means that

$$\begin{aligned} \Delta(n\,A + m_1\,B_1 + m_2\,B_2) &= n\,\Delta(A) + m_1\,\Delta(B_1) + m_2\,\Delta(B_2) = 0 \\ \Gamma(n\,A + m_1\,B_1 + m_2\,B_2) &= n\,\Gamma(A) + m_1\,\Gamma(B_1) + m_2\,\Gamma(B_2) = 0 \end{aligned}$$

	Strike = 70 EUR	Strike = 90 EUR
Preis	10.18 EUR	0.04 EUR
Delta	0.992	0.024
Omega	7.79	47.41

Solving this system of equations yields the number of hedging instruments m_1 and m_2 for the delta and gamma hedged portfolio:

$$m_1 = -n\frac{\Gamma(B_2)\Delta(A) - \Delta(B_2)\Gamma(A)}{\Gamma(B_2)\Delta(B_1) - \Delta(B_2)\Gamma(B_1)}\ ,\ m_2 = -n\frac{\Gamma(B_1)\Delta(A) - \Delta(B_1)\Gamma(A)}{\Gamma(B_1)\Delta(B_2) - \Delta(B_1)\Gamma(B_2)}$$

From the Black-Scholes differential equation 13.5 we can obtain a relation, which the different sensitivities of *all* financial instruments *must* satisfy (if they are not future styled). Simply replace the partial derivatives in Equation 13.5 by the corresponding sensitivities to obtain

$$\Theta + (r-q)S\Delta + \frac{1}{2}b^2\Gamma = r\,V_S$$

One consequence of this equation is, for example, the following statement: The change value of a delta and gamma neutral portfolio per time (i.e. Θ) is equal to its current value multiplied by the risk-free interest rate.

13.5.2 Omega and Beta

In contrast to the delta which gives the *absolute* price change of the derivative (in monetary units) resulting from an *absolute* change of one monetary unit in the underlying, *omega* is the *relative* price change of the derivative (in %) resulting from a *relative* change in the underlying[6]:

$$\Omega \equiv \frac{[V(S+dS) - V(S)]\,/\,|V(S)|}{dS/S} = \frac{S}{|V(S)|}\Delta$$

In situations where a trader is instructed to invest a specific *amount* of money rather than in a certain *number* of instruments, this allows a significantly better assessment of the risk than the (much more prominent) delta. A simple example will serve to clarify this point.

Assume that 10,000 euros are to be invested either in a call with a strike price of 90 euros on an underlying paying no dividend with a spot price of 80 euros or in a call on the same underlying with strike price of 70 euros. Suppose the annualized volatility of this underlying is 20%. Both options mature in 30 days. At a risk-free rate of 3% per annum, Equation 9.7 (implemented in the Excel workbook BlackScholesModel.xls on the accompanying CD-ROM) gives the following prices and risk ratios:

The sensitivity delta is thus almost 50 times greater for the call with a strike price of 70 euros than for the call with the strike price of 90 euros. Is the call with a strike of 70 euros riskier? If we have the choice of buying a certain *number* of the strike-70

[6]The absolute value of $V(S)$ in this definition has the effect that omega always has the same sign as delta, even for a position with a negative present value.

calls or the same number of strike-90 calls then yes, since we could loose a lot more with a certain number of the strike-70 calls than with the same number of strike-90 calls. But this is solely because the *price* of the strike-70 call is so much higher than the price of the strike-90 call! The higher risk is entirely due to the significantly larger investment: if we invest more euros we can loose more euros.

But if the *same amount* of money is to be invested, a portfolio with strike-90 calls (250,250 of these calls can be purchased for 10,000 euros) is significantly riskier than a portfolio with strike-70 calls (only 982 of these calls can be purchased for 10,000 euros). This risk is impressively quantified by omega, also called the *elasticity*. For the strike-70 call portfolio, a 1.60 euros (=2%) decline in the underlying's price results in a loss of value of 7.79 * 2 %, or 1,558 euros. However, the strike-90 call portfolio loses 47.41*2% of its value, a sum of 9,482 euros. Investing in the strike-90 calls, a relatively small change in the price of the underlying (2%) almost wipes out the entire investment. On the other hand, a slight increase in the price of the underlying can just as easily double the value of the strike-90 call portfolio. Thus, the volatility of the option equals omega times the volatility of the underlying

$$\sigma_{\text{Option}} = \Omega \, \sigma_{\text{Underlying}} \tag{13.20}$$

A similar relation can also be shown to hold for the expected return of an option (in the real world). The market requires, as a compensation for the risk of an investment, a higher expected return of the security under consideration compared to the risk-free rate. Since the risk (as represented by the volatility) of an option on that security is omega times as high as the risk of the security itself, the compensation for taking the risk of investing in the option must be omega times as high as well:

$$\mu_{\text{Option}} - r = \Omega \, (\mu_{\text{Underlying}} - r) \tag{13.21}$$

We can now establish a connection with the *CAPM* (*capital asset pricing model*). The variable β (*beta*) in the CAPM relates the risk premium of a security to the risk premium of a portfolio representing the entire market (with the corresponding diversification). *Indices* such as the *DAX* or the *Dow Jones* are constructed to represent such entire markets. To be more specific: the beta of a security with a price S with respect to an index with a price I is defined as the covariance of the price with the index divided by the variance of the index (see Equation A.20 in Section A):

$$\beta_{SI} = \frac{\text{cov}\,[S, I]}{\text{var}\,[I]} = \frac{\sigma_S}{\sigma_I} \rho_{SI} \tag{13.22}$$

This variable can be interpreted as follows: should the value I of the index increase by 1%, then the price S of the security will increase by beta %. The relative change of S is thus beta times the relative change in I.

$$\begin{aligned} \beta_{SI} &= \frac{\text{relative change in } S \text{ resulting from a change in } I}{\text{relative change in } I} \\ &= \frac{(dS)/S}{(dI)/I} = \frac{I}{S} \frac{dS}{dI} \end{aligned} \tag{13.23}$$

This corresponds to the definition of omega if S is interpreted as a financial instrument having the index I as an underlying. Within the capital asset pricing model the risk premium for investing in S is shown to be

$$\mu_S - r = \beta_{S,I}\,(\mu_I - r) \tag{13.24}$$

Substituting this for the risk premium of the underlying S in Equation 13.21 gives

$$\mu_{\text{Option}} - r = \Omega\beta_{S,I}\,(\mu_I - r)$$

Thus, the risk premium for investing in the option is omega times beta multiplied by the risk premium for investing in the entire market. We could therefore also use Equation 13.24 to define a beta for the option as well.

$$\beta_{\text{Option}} = \Omega\,\beta_{\text{Underlying}} \tag{13.25}$$

Volatility, risk premium and the beta of an option are thus obtained by simply multiplying the corresponding variables for the underlying by the elasticity omega.

13.5.3 Summation of Sensitivities of Different Underlyings

The summation of sensitivities presented in Section 13.5.1 is only possible if they are sensitivities of instruments on the same underlying. It is often the case, however, that a portfolio consists of instruments on several different underlyings. The net delta of such a portfolio is *not* simply the sum of the deltas of the individual instruments. For example, in a chemical portfolio with options on Bayer and Hoechst stocks, the delta of the Bayer option with respect to the Bayer stock price cannot be added to the delta of the Hoechst option with respect to the Hoechst stock price to obtain the portfolio's delta. We must first answer the question: Which stock should we chose as the reference underlying in calculating the sensitivity of our portfolio in the above example? Bayer or Hoechst?

Having chosen a reference underlying S_i from among all the instruments S_k in the portfolio with respect to which the sensitivities of the entire portfolio are to be calculated, we must then proceed by expressing the sensitivities of the remaining underlyings in terms of the sensitivity of this chosen reference. The Betas of the S_k with respect to the S_i can be used in the calculation of such an expression. From the volatilities of the prices and the pairwise correlation between the underlyings, we can use Equation 13.22 or A.20 to calculate the necessary Betas, even when neither of the prices is an index. From the sensitivity Δ_k of an instrument with value V_k on the underlying S_k we can then deduce the sensitivity Δ_i of this instrument with respect to *another* underlying S_i by using Equation 13.23:

$$\Delta_i = \frac{dV_k}{dS_i} = \frac{dS_k}{dS_i}\frac{\partial V_k}{\partial S_k} = \frac{dS_k}{dS_i}\Delta_k = \frac{S_k}{S_i}\beta_{ki}\Delta_k = \frac{S_k\sigma_k}{S_i\sigma_i}\rho_{ki}\Delta_k \tag{13.26}$$

In the second to last step, Equation 13.23 in the form "$dS/dI = \beta_{SI}S/I$ " is applied while in the last step we make use of Equation 13.22. Having in this manner transformed the sensitivities with respect to the underlyings S_k into sensitivities with respect

to the selected reference underlying S_i, we can now simply add them up to calculate the total sensitivity of the portfolio since they now refer to the same underlying. The sensitivity of the entire portfolio with respect to the reference underlying is thus given by

$$\Delta_i^{\text{Portfolio}} = \frac{1}{S_i} \sum_k S_k \, \beta_{ik} \, \Delta_k = \frac{1}{S_i \sigma_i} \sum_k S_k \, \sigma_k \, \rho_{ik} \, \Delta_k \tag{13.27}$$

This reference underlying can (but need not) be an index on a relevant business sector, for example. Such a combination of securities on different underlyings makes sense when the underlyings are highly correlated. The correlation of the underlyings, or at least the correlation with the selected reference underlying, must be known for the following two reasons:

- They are necessary for the transformation of the deltas in accordance with Equation 13.26.
- Sound decisions as to a reasonable choice of the above mention combination of securities can only be made on the basis of the correlations.

13.6 Computation of the Greek Risk Variables

Having introduced the *Greeks* and their applications in the last section, we will now give a short discussion on their determination using the most important pricing methods.

13.6.1 Sensitivities in the Binomial Model

In order to obtain the price changes in the binomial model resulting from a change in the underlying's spot price, under the assumption that all other parameters (including the time) remain the same, we require several different spot prices for the same time. This is the case for all times except at the beginning of the tree. To obtain more than one spot price at time t, we start the tree at least one time step earlier for each order of differentiation desired. The highest order derivative with respect to S which will be needed is that for gamma, i.e. the second derivative. Thus we let the tree start two time steps in the past, i.e. at time $t - 2dt$. In doing so, we generate (in a recombinant tree) three different price evolutions uuS, udS and ddS. The tree should be constructed in such a way that the middle price evolution udS represents the actual existing spot price $S(t)$. This implies that the tree must start at $S(t - 2dt) = e^{-2\mu dt} S(t)$ if we wish to apply the procedure given by Equation 11.30. If Equation 11.31 is to be utilized instead, the tree should begin with $S(t - 2dt) = S(t)$. Let $V_S(t)$ denote the value of any given financial derivative (or portfolio) on the spot price S. For a binomial tree beginning two time steps before the present time t, the following sensitivities can be

calculated *directly* within the same tree:

$$\Delta \approx \frac{V_S^{uu}(t) - V_S^{dd}(t)}{(u^2 - d^2)S(t-2dt)} \approx \frac{V_S^{uu}(t) - V_S^{ud}(t)}{u(u-d)S(t-2dt)} \approx \frac{V_S^{ud}(t) - V_S^{dd}(t)}{d(u-d)S(t-2dt)}$$

$$\Omega = \frac{u\,d\,S(t-2dt)}{V_S^{ud}(t)}\Delta \tag{13.28}$$

$$\Gamma = 2\frac{d\,V_S^{uu}(t) - (u+d)V_S^{ud}(t) + uV_S^{dd}(t)}{ud(u-d)(u^2-d^2)S(t-2\delta t)^2}$$

$$\Theta = \frac{V_S^{uudd}(t+2dt) - V_S(t-2dt)}{4dt} \approx \frac{V_S^{uudd}(t+2dt) - V_S^{ud}(t)}{2dt} \approx \frac{V_S^{ud}(t) - V_S(t-2dt)}{2dt}$$

Other sensitivities such as vega and rho can only be determined via Equation 13.18: we have to calculate the price for a slightly modified volatility or risk-free rate, respectively, with a completely new tree, subsequently dividing the difference between the old and the new option price by the respective change in the parameter to obtain the desired sensitivity.

For the first-order partial derivatives of the price, namely delta, omega and theta, the binomial model provides three different choices: a symmetric derivative, an asymmetric derivative for which the price change with respect to an increasing parameter value is used, and an asymmetric derivative for which the price change with respect to a declining parameter value is used. All these alternatives are displayed in Equation 13.28.

We have previously introduced a somewhat different equation for delta (see Equation 11.8). The difference is that in the earlier expressions for delta, the time was not assumed to be strictly constant but changed by dt. The price change induced by this time change is compensated for by the other form of the delta so that both methods are correct up to the order of precision obtained by using the binomial model.

The equation for Theta is, strictly speaking, only correct for the procedure given by Equation 11.31 since in this procedure, the derivative prices used all contain the *same* underlying price. In the procedure given by Equation 11.30, however, not only does the time change but, because of the drift appearing in u and d, the underlying S changes as well. Consequently, a change in the option price resulting *solely* because of a time change is not obtained. This is one (more) reason why Equation 11.31 is used almost exclusively in this text.

The expression for gamma can be obtained from the following argument: gamma is by definition the change of delta per change of the underlying. At time t we have two deltas at our disposal, these being the two unsymmetric derivatives in Equation 13.28

$$\Delta_u \approx \frac{V_S^{uu}(t) - V_S^{ud}(t)}{u(u-d)S(t-2\delta t)} \quad , \quad \Delta_d \approx \frac{V_S^{ud}(t) - V_S^{dd}(t)}{d(u-d)S(t-2\delta t)}$$

Δ_u is the delta between the upper and middle node at time t. Similarly Δ_d is the delta between the middle and lower node. The difference between these two deltas will serve as the delta change needed to calculate gamma. In addition we need the values of the underlying belonging to those two deltas since gamma is the delta change per corresponding underlying change. As the underlying value belonging to Δ_u we use

the average (denoted by $\overline{S}_u$) of the underlying values at the upper and middle nodes. Similarly we use the average (denoted by $\overline{S}_d$) of the underlying values at the lower and middle nodes as the underlying value belonging to Δ_d.

$$\overline{S}_u = \frac{1}{2}\left[uu + ud\right] S(t-2\delta t) = \frac{1}{2}u(u+d)S(t-2\delta t)$$
$$\overline{S}_d = \frac{1}{2}\left[dd + ud\right] S(t-2\delta t) = \frac{1}{2}d(u+d)S(t-2\delta t)$$

With respect to *these* underlying values Δ_u and Δ_d are even *symmetric* derivatives. Gamma is now simply the delta difference divided by the difference in underlying values:

$$\begin{aligned}
\Gamma &= \frac{\Delta_u - \Delta_d}{\overline{S}_u - \overline{S}_d} \\
&= \frac{1}{\overline{S}_u - \overline{S}_d} \frac{\left[V_S^{uu}(t) - V_S^{ud}(t)\right]/u - \left[V_S^{ud}(t) - V_S^{dd}(t)\right]/d}{(u-d)S(t-2\delta t)} \\
&= \frac{1}{\overline{S}_u - \overline{S}_d} \frac{d\left[V_S^{uu}(t) - V_S^{ud}(t)\right] - u\left[V_S^{ud}(t) - V_S^{dd}(t)\right]}{ud(u-d)S(t-2\delta t)} \\
&= \frac{2}{(u-d)(u+d)S(t-2\delta t)} \frac{dV_S^{uu}(t) - (u+d)V_S^{ud}(t) + uV_S^{dd}(t)}{ud(u-d)S(t-2\delta t)}
\end{aligned}$$

This corresponds exactly to the gamma in Equation 13.28.

A detailed demonstration on how to determine the sensitivities of an option portfolio using the binomial model is provided in the Excel workbook BinomialTree.xls.

13.6.2 Sensitivities in the Black-Scholes Model

Since the Black-Scholes model provides analytic solutions for option prices, the sensitivities can be determined immediately by calculating the required partial derivatives directly. The sensitivities of a plain vanilla call and put on an underlying with spot price S earning a yield q will serve as an example and are given below. The expressions appearing here arise from differentiating[7] the function in Equation 9.7 and exploiting the relation in Equation 13.17. The sensitivities of options on futures and forwards can be calculated similarly by differentiating the functions given in Equations 18.4.2

[7] Only the *explicit* dependence and not the *implicit* dependence of the option on the parameters under consideration play a role when calculating these partial derivatives. In determining theta, for example, the time dependence of S should not be taken into consideration, since this dependence is *implicit* for the option price. This means that in taking the partial derivatives, S, t, r, q and σ are to be viewed as independent parameters.

and 9.9.

$$\Delta_{\text{Call}} = B_q(t,T)\,\mathrm{N}(x) \quad , \quad \Delta_{\text{Put}} = B_q(t,T)\,[\mathrm{N}(x)-1]$$

$$\Gamma_{\text{Call}} = \frac{B_q(t,T)\,\mathrm{N}'(x)}{S(t)\sigma\sqrt{T-t}} = \Gamma_{\text{Put}}$$

$$\Theta_{\text{Call}} = -\frac{B_q(t,T)S(t)\,\mathrm{N}'(x)\sigma}{2\sqrt{T-t}} + S(t)\mathrm{N}(x)\frac{\partial B_q(t,T)}{\partial t} - K\mathrm{N}(x-\sigma\sqrt{T-t})\frac{\partial B_r(t,T)}{\partial t}$$

$$\Theta_{\text{Put}} = -\frac{B_q(t,T)S(t)\,\mathrm{N}'(x)\sigma}{2\sqrt{T-t}} - S(t)\mathrm{N}(-x)\frac{\partial B_q(t,T)}{\partial t} + K\mathrm{N}(-x+\sigma\sqrt{T-t})\frac{\partial B_r(t,T)}{\partial t}$$

$$\Psi_{\text{Call}} = B_q(t,T)S(t)\,\mathrm{N}'(x)\sqrt{T-t} = \Psi_{\text{Put}}$$

$$\rho_{\text{Call}} = -K\mathrm{N}(x-\sigma\sqrt{T-t})\frac{\partial B_r(t,T)}{\partial r} \quad , \quad \rho_{\text{Put}} = K\mathrm{N}(-x+\sigma\sqrt{T-t})\frac{\partial B_r(t,T)}{\partial t}$$

$$\rho_{q_{\text{Call}}} = S(t)\mathrm{N}(x)\frac{\partial B_q(t,T)}{\partial q} \quad , \quad \rho_{q_{\text{Put}}} = -S(t)\mathrm{N}(-x)\frac{\partial B_q(t,T)}{\partial q} \tag{13.29}$$

Here, x is as defined in Equation 9.5 and N' again denotes the derivative of the cumulative standard normal distribution with respect to its argument:

$$\mathrm{N}'(x) = \frac{d\mathrm{N}(x)}{dx} = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

The Excel workbooks BLACKSCHOLES.XLS and STRADDLE.XLS demonstrate these sensitivities. In BLACKSCHOLES.XLS, the hedging of a portfolio consisting of puts, calls and underlyings is demonstrated and in STRADDLE.XLS, the behavior of the sensitivities of a straddle as a function of the underlying is described.

13.6.3 Sensitivities by Means of Finite Difference Methods

In principle, every numerical algorithm used to compute the price of an option can also be applied for the determination of the sensitivities or Greeks. All numerical procedures for computing the sensitivities are in essence based on the same fundamental idea displayed in Equation 13.18. Two option prices are computed. These prices differ since one of the risk factors used in their determination was slightly different. The sensitivity of the option price with respect to that particular risk factor is then the difference between the two option prices divided by the difference in the risk factor.

With finite difference methods – as with binomial trees – some sensitivities can be calculated (as finite differences) *directly on the grid* used for pricing without having to go through the lengthy procedure based on Equation 13.18. In fact, a finite difference scheme does nothing other than calculate (approximations for) the derivatives with respect to S and t at each time step. The relevant formulas have already been presented in Section 10. Explicitly, Equation 10.9 directly yields *delta*, Equation 10.10 *gamma*, and Equation 10.12 *Theta*. Calculating Greeks directly on the grid has proved to yield results which are significantly more exact[8] than procedures based on Equation 13.18. Hence, the Greeks should be calculated on the grid whenever possible.

[8] For instance, we might otherwise observe that the sensitivities oscillate as a function of the spot price of the underlying.

Vega and *rho,* however, cannot be calculated directly from an (S, t)-grid and so for these sensitivities we have to rely on Equation 13.18, i.e. the calculation must be run through twice, in each case with a slight change in the respective parameter.

The determination of the sensitivities of an options portfolio using finite difference methods is shown in the Excel workbook FiniteDifferenceMethod.xls on the accompanying CD-ROM.

13.6.4 Sensitivities by Means of Monte Carlo Simulations

When using Monte Carlo simulations in determining the sensitivities, the computations are always performed in accordance with Equation 13.18. This means the two simulations are performed with slightly different values for the parameter under consideration but with *exactly the same random numbers.*

The difference between the two option prices divided by the difference in the parameter value gives an approximation of the desired sensitivity. Differentiating twice, as for gamma, for example, is accomplished in accordance with Equation 13.18 with three different simulations for three different parameter values. Each simulation must be performed *with the same random numbers.* The computation of the sensitivities by means of Monte Carlo simulations is demonstrated in full detail for an option portfolio in the Excel workbook MonteCarloSimulation.xls on the accompanying CD-ROM.

14

Martingale and Numeraire

14.1 The Martingale Property

The most important and profound concept that the reader may have gained from the material presented in this book so far is that of risk neutrality, which can be summarized as follows:

Today's price of a (tradable) financial instrument is equal to the discounted expectation of its future price if this expectation is calculated with respect to the risk-neutral probability measure.

At this point, we recommend that the reader review Sections 8.1.3, 8.2.1 and 11.2.3. We will now elaborate on the concept of risk neutrality.

The *risk-neutral probability* is an example of a *martingale measure*. Martingale measures are a specific class of probability measures satisfying the property – as we are about to see – described in Equation 14.1. For an intuitive explanation of the term *measure*: probability distributions can be interpreted as *measures* since the expectation of a function $f(X)$ of a random variable X, having a distribution with density function p, can be interpreted as an integral with respect to a certain *integral measure*:

$$\mathrm{E}\left[f(X)\right] = \int f(x) \underbrace{p(x)dx}_{\text{Integral Measure}}$$

By replacing the *Riemann integral measure* dx with the *probability measure* $p(x)dx$, we obtain the expectation of the random variable $f(X)$ by integrating the function f with respect to this probability measure.

Martingale theory is an intensively researched filed of mathematics, and recognizing that the price process of a derivative can be interpreted as a martingale allows the application of martingale theory in the valuation of financial instruments. Because of its generality, the utilization of martingale theory requires very few of the assumptions listed in Section 6, namely Assumptions 1, 2, 3, 4, 5. The methods presented here

will not be discussed in complete mathematical detail, but will be motivated by our "experience" of the subject gained in previous sections.

The results in Section 11.2.3 are based on the somewhat surprising observation that the probability p' of an upward move u in the underlying price in the real world does not appear in Equations 11.8 or 11.13. Only the risk-neutral probability plays a role. This observation led to Equation 11.20 and the interpretation of derivative prices as expectations taken with respect to a certain probability measure.

At this stage, we wish to generalize this concept. Consider again the derivation of Equation 11.8. The specific type of the financial instrument V does not come into play. The discussion took place in the context of options merely for simplicity's sake, making use of none of the properties particular to options. Hence, Equation 11.20 holds for all financial instruments whose value is governed by the price of the underlying S. Furthermore, at no point in the derivation did the process S have to have a particular form. This means that Equation 11.20 holds for *arbitrary* instruments on *arbitrary* underlyings, in other words, for all general processes of the form 3.16. Such processes are much more general than the simple random walk given by Equation 3.20, for example. This is of decisive importance in the analysis of term structure models. In fact, the generality of Equation 11.20 goes still further. Using $B(T,T) = 1$, Equation 11.20 can be written as

$$\frac{V_S(t)}{B(t,T)} = \mathrm{E}\left[\frac{V_S(T)}{B(T,T)}\right]$$

where we recall that the expectation at time t has been taken with respect to the risk-neutral probability p. Defining the *normalized* price by

$$Z_S(t) := \frac{V_s(t)}{B(t,T)}$$

then

$$Z_S(t) = \mathrm{E}\left[Z_S(T)\right] \tag{14.1}$$

In other words:

The normalized price at time t is given by the expectation (with respect to p) of the future normalized price.

This expresses the martingale property in its "purest" form. We can also say:

The normalized price is a martingale.(with respect to p)

Normalizing the price means nothing other than expressing the price of the instrument in units of zero bonds maturing at T rather than in monetary units such as euros. Thus, the numerical price does not tell us how much the instrument costs in euros but how much the instrument costs in terms of zero bonds.

Having shown that the martingale property holds (within limits[1]) for *arbitrary financial instruments* on *arbitrary underlyings*, we now show that it also holds for *arbitrary normalizing factors*. This indicates the truly general character of the martingale property.

[1]The limits are that the underlying has to follow a general Ito process of the form 3.16 and that the financial instrument as well as the normalizing factor have to be traded instruments, see Section 15.4.

14.2 The Numeraire

In financial literature, the normalizing instrument is commonly referred to as *numeraire*. In our discussions here, we will frequently use the more intuitive expression *normalization*. We will now show that not only zero bonds but *arbitrary* (tradable) financial instruments may serve as normalizing factors, i.e. as numeraires. The numeraire used does not even have to refer to the underlying S (the zero bond does not do so either). Let S be the price of an underlying, V_S the price of an arbitrary financial instrument on this underlying, and Y another arbitrary financial instrument (in our previous discussions, the zero bond was chosen to play the role of Y). As in the derivation of Equation 11.8, let the time evolution of the price of the underlying be described by a *binomial tree* such as given in Equation 11.2, for example.

We intend to construct a portfolio composed of α underlyings whose prices are given by the process S and β of the financial instruments with a price process given by Y:

$$\Pi(t) = \alpha(t)S(t) + \beta(t)Y(t) \tag{14.2}$$

We require this portfolio to have a value equal to that of the derivative V in all states of the world one time step later. Note that both the normalizing factor Y and the underlying S must be *tradable* since otherwise the "α underlyings at price S" or the "β instruments at price Y" could be neither purchased nor sold on the market and the construction of such a portfolio would be impossible. Since the normalizing instrument Y can, as will be shown, be chosen arbitrarily, there is no shortage of candidates for Y; we simply select any *tradable* instrument as the numeraire.

However, it is often the case that while a financial instrument on an underlying is tradable, the underlying *itself* is *not* tradable. This situation arises quite frequently. For example, the forward price $S(t,T)$ (see Equation 7.1) is generally not tradable even if the associated spot price $S(t)$ is the price of a tradable instrument (such as a stock). Despite this fact, the forward price is often used as an underlying; the reader is referred to Equations 18.4.2 or 9.9 for examples. Also, as we will show in detail in Section 15.4, interest rates are not tradable either, in contrast to bonds which are financial instruments having the interest rate as an "underlying". In such cases, a second tradable instrument U_S having S as its underlying is chosen in addition to the numeraire instrument Y. The only restriction in the choice of this second instrument is that it must not be possible to construct U_S by a portfolio consisting solely of the numeraire instrument (we need two truly "linearly independent" instruments).

A portfolio can now be constructed similar to Equation 14.2 by replacing the non-tradable underlying S with the tradable instrument U_S. A non-tradable underlying does not, in principle, complicate the situation as long as a tradable instrument on the underlying U_S (which cannot be represented by the numeraire instrument) can be found[2].

$$\Pi(t) = \alpha(t)U_S(t) + \beta(t)Y(t) \tag{14.3}$$

The portfolio thus constructed is required to have the same value as the derivative V

[2]If S itself is tradable, it can, of course, be chosen as the (tradable) instrument U_S. If this is the case, we merely need to replace U_S with S in all pertinent equations derived in the following material.

after proceeding by one time step. Hence

$$\begin{aligned}\alpha(t)U_{S_u}(t+\delta t)+\beta(t)Y_u(t+\delta t)&=\Pi_u(t+\delta t)\stackrel{!}{=}V_{S_u}(t+\delta t)\\ \alpha(t)U_{S_d}(t+\delta t)+\beta(t)Y_d(t+\delta t)&=\Pi_d(t+\delta t)\stackrel{!}{=}V_{S_d}(t+\delta t)\end{aligned}\tag{14.4}$$

The weights α and β satisfying both of these equations can now be uniquely determined[3].

$$\alpha(t)=\frac{V_{S_u}Y_d-V_{S_d}Y_u}{U_{S_u}Y_d-U_{S_d}Y_u}\quad,\quad \beta(t)=\frac{V_{S_d}U_{S_u}-V_{S_u}U_{S_d}}{U_{S_u}Y_d-U_{S_d}Y_u}\tag{14.5}$$

If in all events the derivative and the portfolio have the same value at time $t+\delta t$, their values must also be equal at time t. Otherwise an arbitrage opportunity would exist.

$$\begin{aligned}V(t)=\Pi(t)&=\alpha(t)U_S(t)+\beta(t)Y(t)\\ &=\frac{V_{S_u}Y_d-V_{S_d}Y_u}{U_{S_u}Y_d-U_{S_d}Y_u}U_S(t)+\frac{V_{S_d}U_{S_u}-V_{S_u}U_{S_d}}{U_{S_u}Y_d-U_{S_d}Y_u}Y(t)\end{aligned}\tag{14.6}$$

It is exactly at this point that the assumption of an *arbitrage free market* enters into our discussion. As we continue with the derivation, we will clearly recognize how this assumption, together with the normalizing factor Y, uniquely determines the martingale probability p. Collecting terms with respect to the coefficients of V_{S_u} and V_{S_d}, we obtain

$$V(t)=V_{S_u}\frac{Y_dU_S(t)-U_{S_d}Y(t)}{U_{S_u}Y_d-U_{S_d}Y_u}+V_{S_d}\frac{U_{S_u}Y(t)-Y_uU_S(t)}{U_{S_u}Y_d-U_{S_d}Y_u}$$

In view of our goal of finding a representation of the value of the derivative normalized with respect to the numeraire instrument Y, we rewrite the above equation in the form

$$\frac{V(t)}{Y(t)}=\frac{V_{S_u}}{Y_u}\frac{Y_dU_S(t)/Y(t)-U_{S_d}}{U_{S_u}Y_d/Y_u-U_{S_d}}+\frac{V_{S_d}}{Y_d}\frac{U_{S_u}-Y_uU_S(t)/Y(t)}{U_{S_u}-U_{S_d}Y_u/Y_d}$$

This equation can now be written as

$$Z_S(t)=Z_S^u(t+\delta t)p_u+Z_S^d(t+\delta t)p_d\tag{14.7}$$

by defining the normalized prices as

$$Z_S(t):=\frac{V(t)}{Y(t)}\quad,\quad Z_S^{u,d}(t+\delta t):=\frac{V_{S_{u,d}}(t+\delta t)}{Y_{u,d}(t+\delta t)}$$

and the "probabilities" as

$$\begin{aligned}p_u&:=\frac{Y_d\frac{U_S}{Y}-U_{S_d}}{U_{S_u}\frac{Y_d}{Y_u}-U_{S_d}}=\frac{Y_uY_dU_S-Y_uYU_{S_d}}{Y_dYU_{S_u}-Y_uYU_{S_d}}=\frac{\frac{U_S}{Y}-\frac{U_{S_d}}{Y_d}}{\frac{U_{S_u}}{Y_u}-\frac{U_{S_d}}{Y_d}}\\ p_d&:=\frac{U_{S_u}-Y_u\frac{U_S}{Y}}{U_{S_u}-U_{S_d}\frac{Y_u}{Y_d}}=\frac{Y_dYU_{S_u}-Y_dY_uU_S}{Y_dYU_{S_u}-Y_uYU_{S_d}}=\frac{\frac{U_{S_u}}{Y_u}-\frac{U_S}{Y}}{\frac{U_{S_u}}{Y_u}-\frac{U_{S_d}}{Y_d}}\end{aligned}\tag{14.8}$$

[3]In order to simplify the notation, the time dependence is suppressed in the arguments. All variables having an index u or d are evaluated at time $t+\delta t$.

where the last expressions are obtained by dividing both the numerator and denominator by $Y_u Y_d Y$. Note that these p are independent of the derivative V_S. They depend explicitly only on the normalizing factor Y and the instrument U_s. The form of these functions arose from the assumption of an arbitrage-free market as expressed in Equation 14.6. *If* p_u and p_d could actually be interpreted as probabilities, Equation 14.7 would have the form indicated in Equation 14.1. Before showing that this is the case, we make the following remark on the computation of p in practice: Instead of expressing p in terms of U_S and Y, as just presented, Equation 14.7 can be used to write p as a function of Z_S (and thus of V_S and Y). Making use of the equality $p_d = 1 - p_u$ yields

$$p_u = \frac{Z_S(t) - Z_S^d(t+\delta t)}{Z_S^u(t+\delta t) - Z_S^d(t+\delta t)} \tag{14.9}$$

This is a method frequently employed in explicitly computing martingale probabilities in practice. This expression has the disadvantage that the independence of p and the derivative V_S is not immediately recognizable.

Before we can interpret p as a probability, it remains to show that p does in fact satisfy all requisite properties. The following conditions, holding for all probabilities, must be checked [4]:

$$p_u + p_d = 1 \quad , \quad p_u > 0 \quad , \quad p_d > 0 \tag{14.10}$$

Using simple algebra, it follows immediately from the explicit representation in Equation 14.8 that $p_u + p_d = 1$ holds. In order to recognize the implications of the other two conditions, note that p_u and p_d have a *common* denominator $\frac{U_{S_u}}{Y_u} - \frac{U_{S_d}}{Y_d}$. This factor is greater than zero if and only if the normalized price of U_S in the "down state" is smaller than in the "up state". This is not necessarily always the case[5], since "up" and "down" are defined by the *un*normalized *underlying* price S ($S_u > S_d$ by definition) and not by the normalized price of the instrument U_S. If such should be the case, i.e. if the *denominator* should be less than zero, *both numerators* must be less than zero as well. Just as both numerators must be greater then zero if the denominator is greater than zero. In summary,

$$\begin{aligned} &\frac{U_{S_u}}{Y_u} > \frac{U_S}{Y} > \frac{U_{S_d}}{Y_d} \quad \text{for} \quad S_u > S > S_d \\ &\qquad \text{or} \\ &\frac{U_{S_u}}{Y_u} < \frac{U_S}{Y} < \frac{U_{S_d}}{Y_d} \quad \text{for} \quad S_u > S > S_d \end{aligned} \tag{14.11}$$

must hold. The normalized price of the instrument U_S must therefore be a strictly monotone function of the underlying price. If this is not the case, we are immediately presented with an arbitrage opportunity. Let us assume for instance, that $\frac{U_{S_u}}{Y_u} < \frac{U_S}{Y} > \frac{U_{S_d}}{Y_d}$. This market inefficiency could be exploited by selling (short selling)

[4] If all three of these conditions hold, it follows immediately that $p_u < 1$ and $p_d < 1$ as well.

[5] Even when S is tradable, allowing U_S to be replaced by the underlying, there are several common instruments that violate this condition when used as a normalizing instrument. For example, the value C of a plain vanilla call on S increases *faster* (in percentage terms) than S itself, implying for the quotient $\frac{S}{C(S)}$ that $\frac{S_2}{C(S_2)} < \frac{S_1}{C(S_1)}$ for $S_2 > S_1$.

the instrument U_S at time t and using the proceeds to purchase $a = U_S/Y$ of the instrument Y. This is always possible as both Y and U_S are tradable instruments. This portfolio has a value at time t of

$$-U_S + aY = -U_S + (\frac{U_S}{Y})Y = 0$$

One time step later, the portfolio's value is, in all events u and d, positive since

$$-U_{S_u} + aY_u = -U_{S_u} + (\frac{U_S}{Y})Y_u = Y_u \underbrace{\left[-\frac{U_{S_u}}{Y_u} + \frac{U_S}{Y}\right]}_{>0} > 0$$

$$-U_{S_d} + aY_d = -U_{S_d} + (\frac{U_S}{Y})Y_d = Y_d \underbrace{\left[-\frac{U_{S_d}}{Y_d} + \frac{U_S}{Y}\right]}_{>0} > 0$$

This strategy leads to a certain profit without placing investment capital at risk. The fact that the value of the portfolio is positive in both possible states u and d can be directly attributed to the assumption that $\frac{U_S}{Y}$ is greater than both $\frac{U_{S_d}}{Y_d}$ and $\frac{U_{S_u}}{Y_u}$, thereby violating the condition in Equation 14.11. Conversely, if $\frac{U_{S_u}}{Y_u} > \frac{U_S}{Y} < \frac{U_{S_d}}{Y_d}$, an analogous arbitrage opportunity arises by following the strategy of going long in U_S and short in $a = U_S/Y$ of the instrument Y.

It follows immediately from these arbitrage considerations that, if the market is arbitrage free, the condition in Equation 14.11 is *automatically* satisfied by every tradable financial instrument playing the role of the numeraire and every tradable instrument U_S on the underlying and, in consequence, need not be verified in practice. p_u and p_d in Equation 14.8 are therefore actually probabilities (they satisfy all the properties in 14.10), implying that the normalized price $Z = V_S/Y$ (and thus the normalized prices of all tradable instruments on S) is a martingale. Or conversely, if it were possible to find a tradable instrument U_S on S whose price, normalized with respect to numeraire instrument Y, is not a martingale, a portfolio consisting of instruments U_S and Y could be constructed, making arbitrage possible.

14.3 Self-Financing Portfolio Strategies

The discussion given above was restricted to one single discrete time step. The extension to arbitrarily many discrete time steps is completely analogous. As described in Section 11, we obtain at every node of the tree a replicating portfolio consisting of the numeraire instrument and a financial instrument on the underlying (or the underlying itself if it is tradable)[6] in order to replicate both possible derivative prices in the next step. Such a replication of the derivative V through a portfolio composed of U_S and Y (or S and Y, if S is tradable) corresponds to a *hedge* of V (for more on this subject,

[6] In Section 11, the normalizing instrument Y was taken to be the zero bond B with a face value of 1 and weight $\beta(t) = g(t)$. The change in value of this zero bond over one time step is $Y_u = B_r(t)^{-1} = Y_d$. The weight $\alpha(t)$ of the underlying was denoted by Δ.

see Section 13). As we move from one time step to the next, this hedge is constantly adjusted by adjusting the value of the weights $\alpha(t)$ and $\beta(t)$ of the (financial instrument on the) underlying and the numeraire. The weights at a specific time t are determined from the values V, Y and U_S (or S) of the derivative, numeraire instrument, and the instrument on the underlying (or the underlying itself) evaluated one time step *later*; see Equation 14.5. We thus calculate backwards through the tree, starting from some future time point T at which the price of the derivative $V(T)$ is known (for example, upon maturity of the derivative). This procedure has already been demonstrated in Section 11. Having calculated backwards, the price of the replicating portfolio at the first node is precisely the desired price of the derivative *provided that* the position adjustments made in the replicating portfolio with the passage of time were accomplished without the injection or withdrawal of capital. To ensure that this is in fact the case, we consider the replicating portfolio at time t

$$\Pi(t) = \alpha(t)U_S(t) + \beta(t)Y(t)$$

This is always set up so that, in accordance with Equation 14.6, it equals the value of the derivative at this time point, i.e.

$$\Pi(t) = V(t)$$

Likewise, the following portfolio is set up at time $t + \delta t$:

$$\Pi(t + \delta t) = \alpha(t + \delta t)U_S(t + \delta t) + \beta(t + \delta t)Y(t + \delta t)$$

This portfolio is also constructed so that its value equals the value of the derivative at the time point in question:

$$\Pi(t + \delta t) = V(t + \delta t)$$

Now consider the value at time $t + \delta t$ of the portfolio which was set up at time t. We denote the value of this portfolio by $\Pi_t(t + \delta t)$, where the index t indicates that this value refers to the portfolio set up at time t. According to the construction in Equation 14.4, this has, in all events, the same value as the derivative at time $t + \delta t$. The *new* portfolio $\Pi_{t+\delta t}$ set up at time $t + \delta t$ also has the same value as the derivative at time $t + \delta t$ as required by Equation 14.6. Thus, the value Π_t of the portfolio constructed at time t is by construction exactly equal to the value of the new portfolio set up at time $t + \delta t$. This implies that we could dissolve the "old" portfolio Π_t at time $t + \delta t$ and use the proceeds to construct the "new" portfolio $\Pi_{t+\delta t}$. In the realization of this strategy, no new capital flows either into or out of the portfolio. Such a strategy is called *self financing*. Thus, the price at the first node required to construct the replicating portfolio is in fact precisely the price of the derivative since no additional capital is required at any time to replicate the derivative nor are any additional proceeds earned from transactions involved in the replication up to the maturity of the derivative.

It is worthwhile to consider this fact from another point of view which allows us to quickly verify whether or not if a portfolio is self financing. The total difference $\delta\Pi(t)$ in the value of a portfolio for a certain trading strategy over a time span δt equals the difference in the value at time $t + \delta t$ of the portfolio set up at time $t + \delta t$ and the value

at time t of the portfolio set up at time t, thus

$$\begin{aligned}\delta\Pi(t) &= \Pi_{t+\delta t}(t+\delta t) - \Pi_t(t) \\ &= \underbrace{\Pi_t(t+\delta t) - \Pi_t(t)}_{\delta\Pi^{\text{Market}}} + \underbrace{\Pi_{t+\delta t}(t+\delta t) - \Pi_t(t+\delta t)}_{\delta\Pi^{\text{Trading}}}\end{aligned}$$

We arrive at the second equation by simply inserting a zero in the form $0 = \Pi_t(t+\delta t) - \Pi_t(t+\delta t)$. It is now easy to recognize both components contributing to the total change in Π: $\delta\Pi^{\text{Market}}$ is the change in value of the portfolio resulting from changes in the market without having adjusted the position in the portfolio. $\delta\Pi^{\text{Trading}}$ is the difference at time $t+\delta t$ between the value of the new portfolio and that of the old portfolio, i.e. the value change resulting solely from trading. For the strategy to be self financing, $\delta\Pi^{\text{Trading}}$ must equal zero since the value of the old portfolio must provide exactly the funds necessary to finance the new portfolio. Consider for example a portfolio composed of two instruments having the form $\Pi(t) = \alpha(t)U_S(t) + \beta(t)Y(t)$ as above:

$$\begin{aligned}\delta\Pi^{\text{Market}} &= \Pi_t(t+\delta t) - \Pi_t(t+\delta t) \\ &= \alpha(t)\underbrace{[U_S(t+\delta t) - U_S(t)]}_{\delta U_S(t)} + \beta(t)\underbrace{[Y(t+\delta t) - Y(t)]}_{\delta Y(t)} \\ \delta\Pi^{\text{Trading}} &= \Pi_{t+\delta t}(t+\delta t) - \Pi_t(t+\delta t) \\ &= U_S(t+\delta t)\underbrace{[\alpha(t+\delta t) - \alpha(t)]}_{\delta\alpha(t)} + Y(t+\delta t)\underbrace{[\beta(t+\delta t) - \beta(t)]}_{\delta\beta(t)}\end{aligned}$$

Then

$$\begin{aligned}\delta\Pi(t) &= \alpha(t)\delta U_S(t) + \beta(t)\delta Y(t) + U_S(t+\delta t)\delta\alpha(t) + Y(t+\delta t)\delta\beta(t) \\ &= \alpha(t)\delta U_S(t) + \beta(t)\delta Y(t) \quad \text{for a self financing strategy}\end{aligned} \tag{14.12}$$

This implies that the strategy is self financing if and only if the *total* change in the portfolio's value from one adjustment period to the next can be explained exclusively by *market* changes. This holds for infinitesimal time steps as well:

$$\begin{aligned}\Pi(t) &= \alpha(t)U_S(t) + \beta(t)Y(t) \quad \text{self financing strategy} \\ &\Longleftrightarrow \\ d\Pi(t) &= \alpha(t)dU_S(t) + \beta(t)dY(t)\end{aligned} \tag{14.13}$$

The weights $\alpha(t)$ and $\beta(t)$, necessary to replicate the derivative one time step *later*, namely at time $t+\delta t$, are already known at time t. A process whose value at a specific time is already known at the previous time step is called a *previsible process*.

From this point of view, we again consider the structure of Equation 14.12. The change $\delta\Pi$ over the next time step is composed of the change δU and the change δY. At time t, neither the value of δU nor of δY are known (we cannot even say if these are "upward" or "downward"), but the *coefficients* $\alpha(t)$ and $\beta(t)$ controlling the influence of these changes on $\delta\Pi$ are known at time t.

Compare this with the general Ito process given in Equation 3.16. The coefficients $a(S,t)$ and $b(S,t)$ which control the next step in the process are also already known at time t. These coefficients a and b are also *previsible processes*.

14.4 Generalization to Continuous Time

The profound and important concepts presented above can be summarized as follows:

- For any arbitrary underlying S which follows a stochastic process of the general form indicated in Equation 3.16,
- and any tradable instrument U with S as its underlying (or S itself if it is tradable)
- and any other arbitrary, tradable, financial instrument Y (*numeraire*),
- the assumption of an arbitrage-free market implies the existence of a (numeraire-dependent) unique probability *measure* p,
- such that the current, normalized price $Z_S = V_S/Y$ (also called the *relative price*) of an arbitrary, tradable, financial instrument V_S on S
- equals the expectation of the future normalized price and thus, Z_S is a martingale with respect to p.
- This statement holds because a *self financing* portfolio strategy with *previsible* weights can be followed which replicates (*hedges*) the price of the financial instrument V_S at all times.

The expectation is taken at time t with respect to the Y-dependent probability measure p. This dependence is given explicitly by Equation 14.8 (in the context of a binomial tree over one time step). To emphasize this dependence, the expectation is often equipped with the subscript t and the superscript Y:

$$Z_S(t) = \mathrm{E}_t^Y\left[Z_S(u)\right] \qquad \forall u \geq t \tag{14.14}$$

The intuitive interpretation of Equation 14.14 is that the *expected* change in the normalized price of the tradable instrument is zero with respect to this probability measure, in other words, the normalized price has *no drift*. This implies that in this measure such a normalized price process has the form given in Equation 3.16 with no drift term:

$$dZ_S(t) = \widetilde{g}_Z(S,t)\, d\widetilde{W} \qquad \text{where} \quad d\widetilde{W} \sim X\sqrt{dt},\ X \sim \mathrm{N}(0,1) \tag{14.15}$$

Here, $d\widetilde{W}$ is a Brownian motion and $\widetilde{g}_Z$ a (*previsible*) process which is different for each different instrument.

The material presented thus far has been restricted to *discrete* time steps. But the above statements hold in continuous time as well. To see this, several fundamental theorems from stochastic analysis are required. In the following, the insights gained

in the study of discrete processes will be carried over to the continuous case and the necessary theorems from stochastic analysis will be used (without proof) when needed. In addition to extending the results already obtained to the time continuous case, the object of the following discussion is to provide a deeper understanding of the general approach to pricing derivatives, in particular, an understanding of role of the *drift* of an underlying.

We therefore consider a very general (not necessarily tradable) underlying S, which, *in the real world,* is governed by an Ito process satisfying Equation 3.16, i.e.

$$dS(t) = a(S,t)\,dt + b(S,t)\,dW \qquad \text{where} \quad dW \sim X\sqrt{dt},\ X \sim \mathrm{N}(0,1) \tag{14.16}$$

Let $U(S,t)$ denote the price of a tradable, financial instrument with an underlying S. The process for U *in the real world,* according to Ito's lemma, is given by Equation 3.18 as

$$\begin{aligned} dU(S,t) &= a_U(S,t)dt + \frac{\partial U}{\partial S}\,b(S,t)\,dW \qquad \text{where} \\ a_U(S,t) &:= \frac{\partial U}{\partial S}a\,(S,t) + \frac{\partial U}{\partial t} + \frac{1}{2}\frac{\partial^2 U}{\partial S^2}b(S,t)^2 \end{aligned} \tag{14.17}$$

Here $a_U(S,t)$ denotes the drift of U.

Furthermore, we select an arbitrary, tradable instrument Y as the numeraire instrument. Note, however, that the choice of the numeraire Y is not completely arbitrary. It has always been tacitly assumed that our market is driven by just one *single* random factor (*one-factor model*), namely the *Brownian motion* dW in Equation 3.16. The numeraire instrument may indeed be any arbitrary, deterministic or stochastic instrument, but if it has a stochastic component, it must be driven by *the same* random walk as the underlying S. If not, the resulting model would be a *multi-factor model,* and in consequence could not be completely "spanned" by the two instruments U and Y. Analogous to Equation 3.16, the most general process describing the numeraire instrument Y satisfies

$$dY(t) = m(Y,t)\,dt + n(Y,t)\,dW \qquad \text{with} \quad dW \sim X\sqrt{dt},\ X \sim \mathrm{N}(0,1) \tag{14.18}$$

with (*previsible*) processes m and n and the same random walk dW, which drives the random component of the underlying S in Equation 14.16.

Motivated by our experience with Section 14.1, we seek a probability measure with respect to which the prices of all tradable instruments (which depend on no stochastic factors other than the Brownian motion dW in Equation 14.16) normalized with the numeraire instrument Y are martingales. We are as yet quite far from attaining this goal. We start by aiming at the short-term goal of finding a measure for which the normalized price

$$Z(S,t) := \frac{U(S,t)}{Y(t)} \tag{14.19}$$

of a *single* selected instrument U is a martingale. The product rule establishes the following equation for the process Z[7]:

$$dZ = d\left[Y^{-1}U\right] = Ud\left[Y^{-1}\right] + Y^{-1}dU + dUd\left[Y^{-1}\right] \tag{14.20}$$

[7]We suppress the arguments of U, Y, a, b, m and n in order to keep the notation simple. The arguments of these variables are always those as given in Equations 14.16 and 14.18.

The differential dU was already specified above. The differential of $f(Y) := Y^{-1}$ is obtained through an application of Ito's lemma, Equation 3.18, under consideration of $\partial f/\partial Y = -1/Y^2$, $\partial^2 f/\partial Y^2 = 2/Y^3$ and $\partial f/\partial t = 0$. Simple substitution gives

$$d\left[Y^{-1}\right] = \left[-\frac{1}{Y^2}m + \frac{1}{Y^3}n^2\right]dt - \frac{1}{Y^2}\,n\,dW$$

The last term in Equation 14.20 appears since the product of the two differentials contains not only higher order terms but also a term $\sim dW^2$ which is *linear* in dt (see Equation 3.17), explicitly:

$$\begin{aligned} dU\,d\left[Y^{-1}\right] &= \left(a_U dt + \frac{\partial U}{\partial S}\,b\,dW\right)\frac{1}{Y^2}\left(\left[-m + n^2/Y\right]dt - n\,dW\right) \\ &= -\frac{\partial U}{\partial S}\,b\frac{n}{Y^2}\underbrace{(dW)^2}_{dt} + \mathcal{O}\left(dtdW\right) \end{aligned}$$

This effect clearly stems from the fact that both dU and dY are no ordinary but *stochastic* differentials[8]. Altogether dZ becomes

$$\begin{aligned} dZ &= \frac{U}{Y}\left(\left[\frac{n^2}{Y^2} - \frac{m}{Y}\right]dt - \frac{n}{Y}dW\right) + \frac{a_U}{Y}\,dt + \frac{\partial U}{\partial S}\frac{b}{Y}dW - \frac{\partial U}{\partial S}\,\frac{bn}{Y^2}dt \\ &= \left(\frac{b}{Y}\frac{\partial U}{\partial S} - \frac{n}{Y}\frac{U}{Y}\right)dW + \left(\frac{a_U}{Y} + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right]\frac{U}{Y} - \frac{bn}{Y^2}\frac{\partial U}{\partial S}\right)dt \\ &= \left(\frac{b}{Y}\frac{\partial U}{\partial S} - \frac{n}{Y}\frac{U}{Y}\right)\left\{dW + \frac{a_u + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right]U - b\,\frac{n}{Y}\frac{\partial U}{\partial S}}{b\frac{\partial U}{\partial S} - n\frac{U}{Y}}dt\right\} \end{aligned} \tag{14.21}$$

where the coefficient of dW is factored out "by force" in the final step. We seek a probability measure with respect to which Z is a martingale, or in other words, a process of the form specified in Equation 14.15. The process dZ would have the desired form if a measure existed with respect to which

$$d\widetilde{W} := dW + \frac{a_u + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right]U - b\,\frac{n}{Y}\frac{\partial U}{\partial S}}{b\frac{\partial U}{\partial S} - n\frac{U}{Y}}dt \tag{14.22}$$

were a standard Brownian motion, i.e.

$$d\widetilde{W} \sim X\sqrt{dt} \quad \text{with} \quad X \sim N(0,1)$$

Stochastic analysis delivers just such a theorem, namely the famous Girsanov Theorem:

Theorem 1 (Girsanov) *Let $W(t)$ be a Brownian motion with respect to a probability measure $\mathcal{P}$, and $\gamma(t)$ a previsible process which (for some future time T) satisfies the boundedness condition*

$$\mathrm{E}_{\mathcal{P}}\left[\exp\left(\frac{1}{2}\int_0^T \gamma(t)dt\right)\right] < \infty$$

[8]Equation 14.20 can be proven formally by applying Ito's lemma (in the version for two stochastic variables) to the function $f(U,Y) = U\,Y^{-1}$.

Then there exists a measure $\mathcal{Q}$, equivalent[9] *to $\mathcal{P}$, with respect to which*

$$\widetilde{W}(t) = W(t) + \int_0^t \gamma(s)ds$$

is a Brownian motion. This implies that

$$dW(t) + \gamma(t)dt = d\widetilde{W}(t) \sim X\sqrt{dt} \quad \textit{with } X \sim N(0,1)$$

Conversely, in the measure $\mathcal{Q}$ the original process $W(t)$ is a Brownian motion with an additional drift component, $-\gamma(t)$: $dW(t) = d\widetilde{W}(t) - \gamma(t)dt$

In order to apply the theorem, the coefficient of dt in Equation 14.22 must be identified with the process $\gamma(t)$ in the Girsanov Theorem. We begin by observing that this coefficient depends only on variables which can be evaluated at time t and as such can itself be determined at time t. This implies that it is previsible. Proceeding under the assumption that the technical boundedness condition in the theorem is satisfied (this will always be the case in our models), the theorem provides a measure with respect to which $d\widetilde{W}$ is in fact a simple Brownian motion. dZ in Equation 14.21 then has a form as in Equation 14.15 with

$$\widetilde{g}_Z(S,t) = \frac{b}{Y}\frac{\partial U}{\partial S} - \frac{n}{Y}\frac{U}{Y} \tag{14.23}$$

Can we now conclude that Z is a martingale with respect to this measure? Does the absence of a drift in Equation 14.15 directly imply[10] the martingale property in Equation 14.1? Again, stochastic analysis provides theorems which ensures (if certain technical conditions are satisfied, see [7, page 79], for example) that this is the case. The measure for which dZ has the form 14.15 is thus a martingale measure.

We have thus attained our first goal by finding a measure with respect to which the normalized price of one selected instrument U (or for S if S should be tradable) is a martingale. The only requirement made of the instrument U is that it be tradable. Thus, for every arbitrary, tradable financial instrument (with S as an underlying) there exists a martingale measure. This measure could, at this point in the discussion, be different for each instrument U, in other words, it may be dependent on our choice of U (just as it depends on the choice of numeraire instrument Y). It remains to show that the normalized price of *every* tradable instrument (with S as its underlying) is a martingale with respect to *the same* probability measure. In the discrete case, the essential point was that a self financing strategy for a portfolio consisting of U and Y exists which replicates V exactly. We will need another important theorem from stochastic analysis to establish this for the time continuous case:

Theorem 2 (Martingale Representation) *If Z is a martingale with respect to the probability measure $\mathcal{P}$ with a volatility which is non-zero almost everywhere with probability $P = 1$, i.e. if Z follows a stochastic process satisfying*

$$dZ = b_Z(t)dW \quad \textit{with} \quad P\left[b_Z(t) \neq 0\right] = 1 \quad \forall t$$

[9] Two probability measures are called equivalent if they agree exactly on what is possible and what is impossible. I.e. an event is impossible (probability zero) in one probability measure if and only if it is impossible in all equivalent probability measures.

[10] We have already shown the reverse implication above.

with a previsible process $b_Z(t)$, and if there exists in this measure another martingale X, then there exists a previsible process $\alpha(t)$ such that

$$dX = \alpha(t)dZ$$

Or equivalently in integral form

$$X(t) = X(0) + \int_0^t \alpha(s)dZ(s)$$

The process $\alpha(t)$ is unique. Furthermore, α and b_Z together satisfy the boundedness condition

$$\mathrm{E}\left[\exp\left(\frac{1}{2}\int_0^T \alpha(t)^2 b_Z(t)^2 dt\right)\right] < \infty$$

This theorem states intuitively that (if the volatility is non-zero), two martingales differ at most by a previsible process. This implies that any martingale can be represented by any other martingale and a previsible process.

As yet, we only have one martingale in our measure, namely Z, the normalized price of U. In order to apply the theorem, we need a second martingale. And since we wish to gather information about the price of another arbitrary financial instrument V, we must construct a second martingale from this instrument V. We do this with the help of yet another quite simple theorem:

Theorem 3 (Tower Law) *For any arbitrary function V, depending on events occurring up to some specified future time $T > t$, the expectation at time t of $V(T)$ with respect to any arbitrary probability measure $\mathcal{P}$,*

$$E(t) := E_t^{\mathcal{P}}\left[V(T)\right]$$

is a martingale with respect to $\mathcal{P}$, explicitly

$$E(t) = E_t^{\mathcal{P}}\left[E(u)\right] \quad \forall u > t$$

It is easy to see that this theorem is true: substituting the definition of E into the claim that $E(t)$ is a martingale reads

$$\mathrm{E}_t^{\mathcal{P}}\left[V(T)\right] = \mathrm{E}_t^{\mathcal{P}}\left[\mathrm{E}_u^{\mathcal{P}}\left[V(T)\right]\right] \quad \forall u > t \tag{14.24}$$

This implies that in taking the expectation at time u and subsequently taking the expectation *of this expectation* at an earlier time t we arrive at the same result as if we had directly taken the expectation with respect to the earlier time t in the first place. The reader should become familiar with this idea by verifying it using the binomial or trinomial trees presented in Sections 11.1 or 11.2.

The payoff profile $V(T)$ of a financial instrument with maturity T is a function depending only on events (values of the underlying process S) occurring up to time T. The Tower Law states that the *expectation* of this payoff profile is a martingale with respect to *every* probability measure, in particular with respect to the martingale

measure of Z from Equation 14.19. Thus, we have found two processes which are martingales with respect to this measure, namely $Z(t)$ and $\mathrm{E}_t^Y[V(T)]$.

But we want more. We want the (appropriately normalized) price $V(t)$ itself to be a martingale, not merely the expectation of the payoff profile $V(T)$. In the discrete case (and in the continuous case for U as well), this was accomplished by considering the *normalized* prices. We therefore consider instead of V the payoff profile normalized with Y, $V(T)/Y(T)$. The expectation (taken at time t with respect to the martingale measure of Z) of this function

$$E(t) := \mathrm{E}_t^Y\left[\frac{V(T)}{Y(T)}\right]$$

is, because of the Tower Law, also a martingale, which we denote by $E(t)$ in what follows. Furthermore, the payoff of V at maturity $t = T$ is exactly replicated by the product $Y(t)E(t)$ since

$$Y(T)E(T) = Y(T)\mathrm{E}_T^Y\left[\frac{V(T)}{Y(T)}\right] = Y(T)\frac{V(T)}{Y(T)} = V(T) \tag{14.25}$$

It is now clear how the existence of a replicating portfolio can be established through an application of the martingale representation theorem: the martingale in question is $E(t)$, the expectation of the normalized derivative price at maturity, and as our second martingale we take $Z(t)$ from Equation 14.19, the normalized price of the initially selected tradable instrument U. The martingale representation theorem now states that (if the volatility of Z is always non-zero) the process $E(t)$ differs from the process $Z(t)$ only by a previsible process $\alpha(t)$:

$$dE = \alpha(t)dZ \tag{14.26}$$

We use this previsible process now to construct a portfolio consisting of $\alpha(t)$ of the instrument U and $\beta(t)$ of the numeraire instrument Y as was done in Equation 14.3. This is always possible since $\alpha(t)$ is previsible by the Martingale Representation Theorem and both U and Y are tradable.

$$\Pi(t) = \alpha(t)U(t) + \beta(t)Y(t) \tag{14.27}$$

This portfolio should equal $Y(t)E(t)$ for all times $t \leq T$ since, according to Equation 14.25, it then replicates the payoff of V upon its maturity exactly, i.e. when $t = T$. From this condition, we can derive the number $\beta(t)$ of numeraire instruments required for the replicating portfolio:

$$\begin{aligned} Y(t)E(t) = \Pi(t) &= \alpha(t)U(t) + \beta(t)Y(t) \\ &\Longleftrightarrow \\ \beta(t) &= E(t) - \alpha(t)\frac{U(t)}{Y(t)} = E(t) - \alpha(t)Z(t) \end{aligned} \tag{14.28}$$

We have thus established the existence of a replicating portfolio. It remains to show that this portfolio is *self financing*, because only if no injection or withdrawal of capital

is required throughout the lifetime of the derivative can we deduce the equality of the portfolio's value and the value of the instrument V. To this end, we consider the total change in the value of the portfolio in light of Equation 14.13:

$$\begin{aligned} d\Pi &= d\,(YE) \\ &= EdY + YdE + dEdY \\ &= EdY + Y\alpha dZ + \alpha dZdY \\ &= [\beta + \alpha Z]\,dY + \alpha Y dZ + \alpha dZdY \\ &= \alpha\,[ZdY + YdZ + dZdY] + \beta dY \\ &= \alpha\,\underbrace{d(ZY)}_{U} + \beta\,dY \end{aligned}$$

where in the second equality the (stochastic) product rule has been applied, in the third equality the Martingale Representation Theorem in form of Equation 14.26, and in the fourth, Equation 14.28 in the form $E(t) = \beta(t) + \alpha(t)Z(t)$ has been used. In the last equation the (stochastic) product rule has been applied again. The equation now states that the total change in the portfolio defined in Equation 14.27 results solely from the change in price of the instruments U and Y and *not* from any adjustment of the positions α or β :

$$d\Pi(t) = \alpha(t)dU(t) + \beta(t)dY(t)$$

Thus, via Equation 14.13 the portfolio is self financing. The value of the portfolio is by construction $\Pi(t) = Y(t)E(t)$ for all times. According to Equation 14.25, this replicates the payoff profile $V(T)$ exactly at time T. The value of the portfolio must therefore equal that of the derivative for all previous times as well:

$$V(t) = \Pi(t) = Y(t)E(t) = Y(t)\mathrm{E}_t^Y\left[\frac{V(T)}{Y(T)}\right] \tag{14.29}$$

and thus

$$\frac{V(t)}{Y(t)} = \mathrm{E}_t^Y\left[\frac{V(T)}{Y(T)}\right] \tag{14.30}$$

Therefore, the normalized price of the tradable financial instrument V is a martingale in the *same* probability measure with respect to which the normalized price of the instrument U is a martingale. Since V was chosen arbitrarily, this implies that the normalized price of *all* tradable instruments are martingales with respect to the *same* probability measure.

Furthermore, the process $\alpha(t)$ can be calculated explicitly. Equation 14.26 states that $\alpha(t)$ is the change of E per change in Z or in other words the derivative of E with respect to Z. Using Equations 14.19 and 14.29, both Z and E can be expressed in terms of the prices of tradable instruments (known at time t)

$$\alpha(t) = \frac{\partial E(t)}{\partial Z(t)} = \frac{\partial\,[V(t)/Y(t)]}{\partial\,[U(t)/Y(t)]} \tag{14.31}$$

The process $\alpha(t)$ corresponds to the sensitivity Δ introduced in Chapter 13. It follows that Equation 14.28 can be applied to calculate $\beta(t)$ at time t explicitly as well.

$$\beta(t) = E(t) - \alpha(t)Z(t) = \frac{V(S,t)}{Y(t)} - \frac{U(S,t)}{Y(t)} \frac{\partial [V(S,t)/Y(t)]}{\partial [U(S,t)/Y(t)]}$$

We have thus accomplished our goal, having shown that the normalized prices V/Y of all tradable instruments are martingales with respect to the martingale measure of $Z = U/Y$. The only question remaining is whether this measure is unique or whether several such measures may exist. To answer this question we apply yet another theorem from stochastic analysis which states that for *complete markets*[11] the martingale measure obtained above is unique. The theorem [75] is stated explicitly here[12]:

Theorem 4 (Harrison-Pliska) *A market consisting of financial instruments and a numeraire instrument is arbitrage free if and only if there exists a measure with respect to which the prices of all financial instruments normalized with the numeraire instrument are martingales. This measure is unique if and only if the market it complete.*

Summary

At this stage, it is helpful to summarize what has been done in this section. The summary corresponds to the summary at the beginning of the Section 14.4 (which was done for discrete time steps).

- We select a tradable instrument Y as the numeraire and another tradable instrument U which has S as an underlying (if S itself is tradable, S can be chosen directly).
- We then find the probability measure for which $Z = U/Y$ is a martingale. The *Girsanov-Theorem* guarantees that this is always possible via a suitable drift transformation as long as a technical boundedness condition is satisfied.
- The *Martingale Representation Theorem* and the *Tower Law* enable the construction of a *self financing portfolio* composed of the instruments U and Y which replicates the payoff profile at maturity of any arbitrary, tradable instrument V having S as an underlying. The value of this replicating portfolio is given by Equation 14.29 where the expectation is taken with respect to the martingale measure of $Z = U/Y$.
- This portfolio must be equal to the value of the derivative $V(t)$ for all times before maturity if the market is arbitrage free. This means that, according to Equation 14.30, the normalized price V/Y with respect to the martingale measure of Z is likewise a martingale. Thus, having obtained (via Girsanov) a martingale measure for Z, the normalized price V/Y of all other tradable instruments are martingales with respect to this same measure.

[11] A market is called *complete* if there exists a replicating portfolio for each financial instrument in the market.

[12] As always, the term "if and only if" means that one follows from the other *and vice versa*.

- Finally, the *Harrison-Pliska Theorem* states that this measure is unique in complete markets: in complete markets there exists for each numeraire instrument *one single* measure with respect to which all tradable instruments normalized with this numeraire are martingales.

14.5 The Drift

With respect to the martingale measure, the price processes of all instruments normalized by the numeraire instrument Y are drift-free. The *expected* changes in the normalized prices of tradable instruments are thus exactly equal to zero. What can we say about the process of the *underlying* with respect to this measure? The model Equation 3.16 was set up to describe the underlying in the *real* world. As we have seen however, the *valuation* of financial instruments is accomplished in a world governed by the probability with respect to which normalized prices of tradable instruments are martingales, i.e. processes of the form indicated in Equation 14.15. Hence, it is important to know how the *underlying* process is transformed when the real probability measure is transformed into the martingale measure.

Not only do we know that a martingale measure exists (and is unique in a complete market) for $Z = U/Y$ with respect to which Z can be represented by a process of the form in Equation 14.15, we know from Equations 14.22 and 14.23 the explicit relationships between the variables in the real world and the world governed by the martingale measure. We now make use of Equation 14.22 in particular, to express the underlying S as given by the process 14.16 with respect to the Brownian motion dW in the real world in terms of the Brownian motion $d\widetilde{W}$ with respect to the martingale measure (using the explicit form for a_u from Equation 14.17):

$$\begin{aligned}
dS &= a\,dt + b\,dW \\
&= adt + b\left\{ d\widetilde{W} - \frac{a_u + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U - b\,\frac{n}{Y}\frac{\partial U}{\partial S}}{b\frac{\partial U}{\partial S} - n\frac{U}{Y}} dt \right\} \\
&= adt - \frac{\frac{\partial U}{\partial S}a + \frac{\partial U}{\partial t} + \frac{1}{2}\frac{\partial^2 U}{\partial S^2}b^2 + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U - b\,\frac{n}{Y}\frac{\partial U}{\partial S}}{\frac{\partial U}{\partial S} - \frac{n}{b}\frac{U}{Y}} dt + bd\widetilde{W} \\
&= \frac{b\,\frac{n}{Y}\frac{\partial U}{\partial S} - a\frac{n}{b}\frac{U}{Y} - \frac{\partial U}{\partial t} - \frac{1}{2}\frac{\partial^2 U}{\partial S^2}b^2 - \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U}{\frac{\partial U}{\partial S} - \frac{n}{b}\frac{U}{Y}} dt + bd\widetilde{W}
\end{aligned} \tag{14.32}$$

This equation explicitly specifies the underlying process with respect to the martingale measure.

Note that in the transition from the real world measure to the martingale measure, only the drift of the underlying has changed and *not* the volatility; the coefficient of the Brownian motion remains $b(S,t)$. This is not merely coincidence but a natural consequence of the Girsanov Theorem which intuitively states that a transformation between two equivalent probability measures effects nothing more than a change in the drift.

We should further note that the Harrison-Pliska Theorem states that the martingale measure in complete markets (we will from now on always assume that the market is complete) is *unique.* This implies that the drift of S, used in the valuation of financial instruments on S is *unique* as well, up to the choice of the numeraire instrument Y.

In the last equality in 14.32 the drift $a(S,t)$ in the real world has been written to share a common denominator with the second term $\sim dt$. In doing so, we observe that the terms $\sim \frac{\partial U}{\partial S}a$ cancel each other and the drift in the real world enters into the equation corresponding to the martingale measure only in the form of $\frac{n}{b}\frac{U}{Y}a$. For numeraire instruments Y satisfying Equation 14.18 with $n(Y,t) = 0\ \forall Y,t$, i.e. for numeraires with processes of the form

$$dY(t) = m(Y,t)\,dt \tag{14.33}$$

the drift a of the *real* world disappears completely for the martingale measure associated with Y! Regardless of the drift $a(S,t)$ chosen in the model, $a(S,t)$ is completely irrelevant for the valuation of financial instruments if the numeraire chosen is a process of the form specified in Equation 14.33.

Equation 14.33 by no means implies that Y must be deterministic, since $m(Y,t)$ is not assumed to be deterministic but only *previsible*[13]. This means that at time t the evolution of Y is only known for the step *immediately following* t, but not for later steps.

As has already been mentioned on more than one occasion, the choice of instrument to be used as the normalizing factor (the *numeraire*) is arbitrary but this choice significantly affects whether a specific problem can be solved elegantly or awkwardly. This is analogous to the selection of a suitable system of coordinates when solving problems in physics, for example. We see from Equation 14.32 that an appropriate choice of numeraire can simplify calculations substantially. The numeraire should always be chosen to be of the form 14.33 for some previsible process m. This is always possible in practice and all numeraire instruments in this book satisfy this property[14]. For such numeraires, m/Y in Equation 14.32 is precisely the *yield of the numeraire instrument*, since Equation 14.33 obviously implies:

$$m = \frac{dY}{dt} \quad \Longrightarrow \quad \frac{m}{Y} = \frac{1}{Y}\frac{dY}{dt} = \frac{d\ln Y}{dt} \tag{14.34}$$

Using this equation and $n = 0$, the drift transformation equation 14.22 becomes

$$\begin{aligned} d\widetilde{W} &= dW + \frac{a_U - \frac{d\ln Y}{dt}U}{b\frac{\partial U}{\partial S}}dt \\ &= dW + \frac{1}{b}\left[a_U\left(\frac{\partial U}{\partial S}\right)^{-1} - \frac{d\ln Y}{dt}\left(\frac{\partial \ln U}{\partial S}\right)^{-1}\right]dt \end{aligned} \tag{14.35}$$

[13] A previsible process is a *stochastic* process whose current value can be determined from information available at the previous time step. Intuitively, it is a stochastic process "shifted back" one step in time.

[14] This will be shown below explicitly for all numeraire instruments used.

and the underlying process in Equation 14.32 reduces to

$$dS(t) = \widetilde{a}(S,t)\,dt + b(S,t)d\widetilde{W} \qquad \text{where}$$

$$\widetilde{a}(S,t) = \left(\frac{\partial U(S,t)}{\partial S}\right)^{-1}\left[U(S,t)\frac{d\ln Y(t)}{dt} - \frac{\partial U(S,t)}{\partial t} - \frac{b(S,t)^2}{2}\frac{\partial^2 U(S,t)}{\partial S^2}\right] \quad (14.36)$$

Only the tradable instrument U and the numeraire Y (and their respective derivatives) and the "volatility" $b(S,t)$ of the underlying process S appear in this expression. As already noted, the real underlying drift $a(S,t)$ has disappeared completely.

Since the prices of financial instruments must obviously be independent of the method used to compute them, the following theorem holds irrespective of the choice of numeraire:

Theorem 5 *Suppose there exists a numeraire Y in an arbitrage-free market satisfying Equation 14.33 with a previsible process $m(Y,t)$. Then the drift of the underlying in the real world is irrelevant to the prices of financial instruments. Arbitrage freedom alone determines the prices of financial instruments and not the expectation of the market with respect to the evolution of the underlying.*

In this context, it is interesting to consider the behavior of the *non*-normalized process of the tradable instrument U in the martingale measure. The process 14.17 in the real world is transformed via Equation 14.35 to

$$\begin{aligned} dU &= a_U dt + \frac{\partial U}{\partial S}\,b\,dW \\ &= a_U dt + \frac{\partial U}{\partial S}\,b\left\{d\widetilde{W} - \frac{1}{b}\left[a_U\left(\frac{\partial U}{\partial S}\right)^{-1} - \frac{d\ln Y}{dt}\left(\frac{\partial \ln U}{\partial S}\right)^{-1}\right]dt\right\} \\ &= a_U dt + \frac{\partial U}{\partial S}\,b\,d\widetilde{W} - \left[a_U - U\frac{d\ln Y}{dt}\right]dt \end{aligned}$$

and thus

$$dU(S,t) = \frac{d\ln Y(t)}{dt}U(S,t)dt + b(S,t)\frac{\partial U(S,t)}{\partial S}d\widetilde{W} \quad (14.37)$$

The drift of a tradable instrument in the martingale measure for a numeraire of the form specified in Equation 14.33 is thus simply the product of the price of the instrument and the yield of the numeraire!

Theorem 6 *The expectated yield (defined as the expected relative price change per time) of a tradable financial instrument in the martingale measure with a numeraire of the form 14.33 is always equal to the yield of the numeraire instrument.*

$$\mathrm{E}_t^Y\left[\frac{dU(S,t)}{U(S,t)}\right] = \frac{d\ln Y(t)}{dt}\,dt \quad (14.38)$$

In Equations 14.37 and 14.38, a tradable instrument is denoted by the letter U but these properties naturally hold for all tradable instrument since, as discussed in detail in the previous section, all tradable instruments are martingales with respect to the same probability measure. The instrument U is not essentially different from other tradable instruments in the market.

14.6 The Market Price of Risk

As emphasized several times previously, all tradable instruments U in a complete, arbitrage-free market (normalized with respect to a selected numeraire) have the same martingale measure. This implies that $d\widetilde{W}$ in Equation 14.22 is always *the same* Brownian motion for this measure. Since the Brownian motion dW of the underlying in the real world is not dependent on the specific financial instrument either, this must hold for the difference $d\widetilde{W} - dW$ as well. The change in drift from dW to $d\widetilde{W}$ in Equation 14.22 must be *the same* for every tradable instrument U. This implies that for two arbitrary tradable instruments, $U_1(S,t)$ and $U_2(S,t)$, the following must hold:

$$\frac{a_{u_1} + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U_1 - b\,\frac{n}{Y}\frac{\partial U_1}{\partial S}}{b\frac{\partial U_1}{\partial S} - n\frac{U_1}{Y}} = \frac{a_{u_2} + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U_2 - b\,\frac{n}{Y}\frac{\partial U_2}{\partial S}}{b\frac{\partial U_2}{\partial S} - n\frac{U_2}{Y}} \tag{14.39}$$

where a_{u_i} denotes the drift of U_i in the *real* world in accordance with Equation 14.17. The existence of a unique measure with respect to which all normalized, tradable instruments are martingales satisfying Equation 14.15 can be formulated equivalently as follows: regardless of the appearance of the drift of the financial instrument in the real world, the combination of this drift with other characteristics of the financial instrument and the numeraire as specified in Equation 14.39 must be the same for *all* financial instruments. This combination has its own name; it is known as the *market price of risk*[15]. The market price of risk γ_U for an instrument U is defined by

$$\begin{aligned}\gamma_U(t) &:= \frac{a_u + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] U - b\,\frac{n}{Y}\frac{\partial U}{\partial S}}{b\frac{\partial U}{\partial S} - n\frac{U}{Y}} \\ &= \frac{1}{b}\left[a_U \left(\frac{\partial U}{\partial S}\right)^{-1} - \frac{d\ln Y}{dt}\left(\frac{\partial \ln U}{\partial S}\right)^{-1}\right]\end{aligned} \tag{14.40}$$

with a_u as in Equation 14.17. The second step in Equality 14.40 holds for a numeraire of the form 14.33. From this definition and Equation 14.39 it is now immediate that the following theorem holds.

Theorem 7 *The market price of risk is identical for all tradable instruments in a complete, arbitrage-free market.*

Comparing the definition in Equation 14.40 with Equation 14.22 immediately gives

Theorem 8 *The previsible process $\gamma(t)$ in the Girsanov Theorem effecting the drift transformation for the transition from the probability measure in the real world to the martingale measure is the market price of risk.*

Let us consider the process given by Equation 14.17 representing a tradable instrument U in the real world from this point of view. The drift a_U in the real world can

[15]The motivation for this name will become clear further below, when we look at certain special cases.

by Definition 14.40 be expressed in terms of the market price of risk:

$$a_U = \gamma_U \, b \frac{\partial U}{\partial S} + U \frac{d \ln Y}{dt}$$

Substituting this expression into Equation 14.17 yields the process for a tradable financial instrument in the *real* world, expressed in terms of the market price of risk:

$$dU(S,t) = \left[\gamma_U(t) \, b(s,t) \frac{\partial U(S,t)}{\partial S} + \frac{d \ln Y(t)}{dt} U(S,t) \right] dt + \, b(S,t) \frac{\partial U(S,t)}{\partial S} dW \quad (14.42)$$

The *valuation* of the financial instrument is accomplished not in the real world but in that governed by the martingale measure. In the martingale measure, the instrument U is a process satisfying Equation 14.37. Comparing this process with Equation 14.42 directly yields the following "recipe":

Theorem 9 *Setting the market price of risk equal to zero in the expression for the stochastic process (more explicitly in the differential equation which is satisfied by this process) which governs the financial instrument in the real world immediately yields the stochastic process (i.e. the differential equation) which is to be applied in the valuation of this instrument.*

14.7 Tradable Underlyings

The equations in the previous section appear relatively complicated because we have assumed throughout that the underlying S is not necessarily tradable. If the underlying is in fact tradable, it can be used in place of the instrument U directly with the consequence that

$$U = S \Longrightarrow \frac{\partial U}{\partial S} = 1 \quad , \quad \frac{\partial^2 U}{\partial S^2} = 0 = \frac{\partial U}{\partial t} = 0 \quad (14.43)$$

The general equation 14.32 for this special case reduces to

$$dS = \frac{b \, \frac{n}{Y} - a \frac{n}{b} \frac{S}{Y} - \left[\frac{n^2}{Y^2} - \frac{m}{Y} \right] S}{1 - \frac{n}{b} \frac{S}{Y}} dt + b d\widetilde{W}$$

The corresponding Equation 14.36 for a more appropriate choice of numeraire instrument satisfying Equation 14.33 further reduces to

$$dS(t) = S(t) \frac{d \ln Y(t)}{dt} \, dt + b(S,t) d\widetilde{W} \quad (14.44)$$

which, of course, agrees with Equation 14.37, this equation holding for every tradable instrument. The expectation of the yield (defined as the expected relative price change per time) of a tradable underlying in the martingale measure is thus (as for every tradable instrument) always equal to the yield of the numeraire instrument.

The market price of risk $\gamma_S(t)$ for a tradable underlying is obtained from Equation 14.40 with 14.43 as

$$\gamma_S(t) = \frac{a + \left[\frac{n^2}{Y^2} - \frac{m}{Y}\right] S - b\,\frac{n}{Y}}{b - n\frac{S}{Y}} = \frac{a(S,t)}{b(S,t)} - \frac{S(t)}{b(S,t)}\frac{d\ln Y}{dt} \tag{14.45}$$

where $a = a(S,t)$ is the underlying drift in the real world (see Equation 14.16). Here, the first equality holds for any arbitrary numeraire of the form 14.18, the second for a (well-chosen) numeraire of the form 14.33. The process given by 14.16 in the real world becomes

$$dS(t) = \left[b(S,t)\gamma_S(t) + \frac{d\ln Y(t)}{dt}S(t)\right]dt + b(S,t)\,dW \tag{14.46}$$

(in agreement with Equation 14.42). Comparison with Equation 14.44 again yields the "recipe" (which holds for all tradable instruments and hence for S as well) that the market price of risk must simply be set equal to zero for valuation purposes.

Note that these considerations are valid for a still very general case: In our treatment, it has only been assumed that the underlying is tradable and that the numeraire instrument is of the form 14.33.

14.8 Applications in the Black-Scholes World

Special cases of the general material discussed above have been encountered in Sections 8.1.3, 8.2.1 and 11.2.3 in various "disguises". The irrelevance of the real world drift became most apparent in Section 11.3. There, the requirement that the expectation of the underlying, computed using its stochastic process, must equal the expectation of the underlying resulting from the martingale property[16] led to Equation 11.25, i.e. an explicit specification of the drift to be used in the valuation. The volatility σ, in contrast, was subject to no such condition and remained the same.

Equation 11.25 holds for a very special case, namely when *a)* the underlying behaves as in Equation 3.20 and *b)* is tradable, *c)* the numeraire is given by $Y(t) = B(t,T)$ and *d)* interest rates and volatilities are constant; in short, in the Black-Scholes world. Let us therefore apply the general results of the above sections to this special situation as an example.

Firstly, a zero bond $B(t,T)$ is chosen as the numeraire instrument maturing at some arbitrary future time T. The process for B in the real world is of the form 14.33. For continuous compounding, we have explicitly

$$dY \equiv dB(t,T) = \frac{dB(t,T)}{dt}dt = rB(t,T)dt \Longrightarrow \frac{d\ln Y(t)}{dt} = r \tag{14.47}$$

The special process, Equation 3.20 (or Equation 3.21) corresponds to the general pro-

[16] There, instead of the general formulation "expectation with respect to the martingale measure", the expression "risk neutral expectation" was used.

cess, Equation 14.16 with the parameters

$$a(S,t) = \widetilde{\mu} S(t) = \left(\mu + \frac{\sigma^2}{2}\right) S(t)$$
$$b(S,t) = \sigma S(t)$$

Since the underlying is tradable, Equation 14.44 can be applied directly to obtain the underlying process in the martingale measure

$$dS(t) = rS(t)\,dt + \sigma S(t) d\widetilde{W}$$

This process and not 3.20 is to be used in the valuation of the financial instrument. Comparison of this process with Equation 3.20 shows that the choice

$$\mu = r - \sigma^2/2 \quad \text{or equivalently} \quad \widetilde{\mu} = r \tag{14.48}$$

for the drift transforms the real world process directly to the process to be used for pricing, in agreement with Equation 11.25 (dividend yield equal to zero). The market price of risk of the underlying is simply

$$\gamma_S(t) = \frac{a(S,t) - rS(t)}{b(S,t)S(t)} = \frac{\widetilde{\mu} - r}{\sigma} \tag{14.49}$$

in this special case and is exactly equal to zero for $\widetilde{\mu} = r$. In fact "setting the market price of risk equal to zero" is equivalent to "choosing the correct drift for pricing".

Equation 14.49 provides the motivation for the name "market price of risk". In the special case considered here the underlying drift in the real world is simply $a(S,t) = \widetilde{\mu} S(t)$. The expectation of the underlying-*yield* is thus $a(S,t)/S(t) = \widetilde{\mu}$. This implies that $\widetilde{\mu} - r$ represents the excess yield above the risk-free rate, which is expected from the underlying in the real world. If the volatility σ is viewed as a measure of the risk of the underlying, then the market price of risk γ_S (at least in the context of this special case) can be interpreted as the excess yield above the risk-free rate *per risk unit* σ which the market expects from the underlying. This is, so to speak, the price (in the form of an excess yield above the risk-free rate) which the market demands for the risk of investing in the underlying. In the real world, the market is by no means risk neutral, but rather expects higher yields for higher risks; the market price of risk illustrates this clearly.

To conclude this section we will now show that we can actually use all of this information about martingales and drifts to really calculate something. We consider below only payoff profiles of path-independent instruments. These are payoff profiles which depend solely on the value of the underlying at maturity T and not on the path taken by the process S between t and T. For such processes

$$V(S,T) = V(S(T),T)$$

holds. Therefore we only need the distribution of S at time T (and not the distribution of all paths of S between t and T). Choosing the zero bond as the numeraire instrument,

$Y(t) = B(t,T)$, Equation 14.30 for the price V of a financial instrument becomes

$$V(S,t) = Y(t)\mathrm{E}_t^Y\left[\frac{V(S,T)}{Y(T)}\right] = B(t,T)\mathrm{E}_t^Y\left[V(S(T),T)\right]$$

since $B(T,T) = 1$. We again model the underlying with the simple process in Equation 3.20. The associated underlying process over a *finite* time interval of length $\delta t = T - t$, i.e. the solution of the stochastic differential equation 3.20 for S, has already been given in Equation 3.25 for an arbitrary drift μ, and thus for an arbitrary probability measure, namely

$$\begin{aligned} S(T) &= S(t)\exp\left[\mu(T-t) + \sigma W_{T-t}\right] \quad \text{with} \quad W_{T-t} \sim \mathrm{N}(0, T-t) \\ &= S(t)\exp\left[x\right] \quad \text{with} \quad x \sim \mathrm{N}\left(\mu(T-t), \sigma^2(T-t)\right) \end{aligned} \tag{14.50}$$

The distribution of $S(T)$ is therefore $S(t)$ multiplied by the exponential of the normal distribution with expectation $\mu(T-t)$ and variance $\sigma^2(T-t)$. Using this, the expectation of the function

$$V(S(T),T) = V(S(t)e^x, T) =: g(x)$$

can be computed explicitly:

$$\mathrm{E}_t\left[V(S(T),T)\right] =: \mathrm{E}_t\left[g(x)\right] = \int_{-\infty}^{\infty} g(x)p(x)dx$$

with the probability density of the normal distribution (see Equation A.45)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}}\exp\left[-\frac{\left(x-\mu(T-t)\right)^2}{2\sigma^2(T-t)}\right]$$

Thus

$$\mathrm{E}_t\left[V(S(T))\right] = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}}\int_{-\infty}^{\infty} V\left(S(t)e^x, T\right)\exp\left[-\frac{\left(x-\mu(T-t)\right)^2}{2\sigma^2(T-t)}\right]dx$$

This is the expectation of the payoff profile with respect to a probability measure associated with μ, for instance with respect to the probability measure in the real world if μ represents the drift in the real world. To allow the use of this expectation for the *valuation* of the instrument V, it must be computed with respect to the (in a complete market, unique) *martingale* measure. This is accomplished through the choice of drift in accordance with Equation 14.48. The price of V then becomes

$$\begin{aligned} V(S,t) &= B(t,T)\mathrm{E}_t^Y\left[V(S(T),T)\right] \\ &= \frac{B(t,T)}{\sqrt{2\pi\sigma^2(T-t)}}\int_{-\infty}^{\infty} V\left(S(t)e^x, T\right)\exp\left[-\frac{\left[x-\left(r-\frac{1}{2}\sigma^2\right)(T-t)\right]^2}{2\sigma^2(T-t)}\right]dx \end{aligned} \tag{14.51}$$

This equation holds in complete generality for every financial instrument with a non-path dependent payoff profile $V(S(T), T)$ on an underlying S of the form specified in Equation 3.20.

To be more specific, we now use the payoff profile of a plain vanilla call as an example:

$$V(S(T), T) = \max\{S(T) - K, 0\} = \max\{e^x S(t) - K, 0\}$$

with x as defined in Equation 14.50. This payoff profile is only non-zero when $e^x S(t) \geq K$, or equivalently when $x \geq \ln(K/S(t))$. Equation 14.51 for the price of the call is then

$$V(S,t) = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \int_{\ln(K/S(t))}^{\infty} [e^x S(t) - K] \exp\left[-\frac{\left(x - \left(r - \frac{\sigma^2}{2}\right)(T-t)\right)^2}{2\sigma^2(T-t)} \right] dx$$

This (with $\mu = r - \sigma^2/2$) is in complete agreement with Equation 9.4, for example. In connection with Equation 9.4 it has been demonstrated how this integral can be computed explicitly. The result is the famous Black-Scholes Equation 9.6.

15

Interest Rates and Term Structure Models

The assumption made in the Black-Scholes world, in particular in the Black-76 model Equation 9.10, that interest rates are non-stochastic directly contradicts the very existence of interest rate options. If interest rates were deterministic and hence predictable with certainty for all future times, we would know at time t which options will be in or out of the money upon maturity T. The options which are out of the money at maturity would be worthless at all earlier times $t < T$ as well. The options which are in the money at maturity would be nothing other than forward transactions. Thus, the assumptions made in pricing interest rate options using the Black-76 model imply that these very options should not even exist! In spite of this fact, the option prices obtained by applying the Black-76 model are surprisingly good. The results of recent research [94][126] have shown that the effects of several "false" assumptions (in particular, the assumed equality of forward and futures prices) tend to cancel out each other[1].

Option pricing methods allowing for *stochastic* interest rates have been the subject of very active research in the field of financial mathematics in recent years. *Term structure models* are used to model the stochastic changes observed in interest rate curves, similar to the way stock prices and exchange rates are modeled with an underlying stochastic process $S(t)$.

In contrast to a stock price or exchange rate $S(t)$, an interest rate $R(t, T)$ depends on *two* time variables, namely on the time t at which the interest rate is being considered *and* on the maturity T of the term over which the interest is to be paid. Usually, interest is paid at a higher rate for longer terms $\tau := T - t$ (i.e. later maturities) than for shorter ones. For a fixed time t (today), the set of all interest rates for the various maturity dates T form a curve called the *term structure*. Theoretically, each point in this curve is a stochastic variable associated with a single maturity. The term structure is thus theoretically a continuum of infinitely many stochastic variables. From a practical point of view, these processes are naturally very strongly correlated; the

[1]The Black-76 model can be derived as a special case of the Heath-Jarrow-Morton model.

interest rate over a term of 3 years and 2.34 seconds is (almost) the same as that over a term of 3 years and 0 seconds and so on. In practice, market participants therefore consider only finitely many terms whose lengths lie well removed from one another (for example, terms $\tau = T - t$ of 1 day, 1 month, 3 months, 6 months, 9 months, 1 year, 2 years, 3 years, 5 years, 10 years and so on). Motivated by the results of principle component analysis (see for instance Section 32.2) which show that well over 90% of the dynamics of the yield curve can be explained by just one or two stochastic factors, most term structure models even go a step further and reduce the number of factors driving the stochastic evolution of the entire term structure to one or two (in rare cases three) stochastic variables. These models are referred to as 1-factor, 2-factor and 3-factor models.

15.1 Instantaneous Spot Rates and Instantaneous Forward Rates

In a 1-factor model, a stochastic process of the form specified in Equation 3.16 is usually assumed for very short term rates, called *instantaneous* interest rates[2]. These rates take the form of either an *instantaneous spot rate* (also called the *instantaneous short rate*) or an *instantaneous forward rate.* The terms $\tau = T - t$ belonging to these rates are infinitesimally short, i.e. we consider the limit $\tau \to 0$, or equivalently $T \to t$.

In the following discussion, we adopt the convention of continuous compounding usually observed in the literature. This allows the instantaneous rates to be defined quite easily. The usual index R appearing in the notation for the discount factor $B_R(t,T)$ in Chapter 3 is suppressed in order to simplify the notation. The *instantaneous spot rate* $r(t)$ is defined by

$$\begin{aligned} e^{-r(t)dt} &:= \lim_{dt\to 0} B(t, t+dt) \\ r(t) &= -\lim_{dt\to 0} \frac{\ln B(t,t+dt)}{dt} \end{aligned} \tag{15.1}$$

The *instantaneous forward rate* $f(t,T)$ is defined by

$$\begin{aligned} e^{-f(t,T)dT} &:= \lim_{dT\to 0} B(T, T+dT\ |t) = \lim_{dT\to 0} \frac{B(t,T+dT)}{B(t,T)} \\ f(t,T) &= -\lim_{dT\to 0} \frac{1}{dT} \ln \frac{B(t,T+dT)}{B(t,T)} = -\lim_{dT\to 0} \frac{\ln B(t,T+dT) - \ln B(t,T)}{dT} \end{aligned}$$

where Equation 3.4 has been used. Thus

$$f(t,T) = -\frac{\partial \ln B(t,T)}{\partial T} \tag{15.2}$$

[2]This is no longer the case in more recent models, called *market rate models* or *Brace-Gatarek-Musiela models,* (*BGM models* for short). For such models, an interest rate over a longer term (for example, the 3-month LIBOR) is used as the underlying stochastic process [15].

Integrating this equation over dT and making use of the fact that $B(t,t) = 1$ yields

$$\int_t^T f(t,s)ds = -\int_t^T \frac{\partial \ln B(t,s)}{\partial s} ds = -\ln B(t,T) + \ln \underbrace{B(t,t)}_{1} = -\ln B(t,T)$$

and therefore

$$B(t,T) = \exp\left[-\int_t^T f(t,s)ds\right] \tag{15.3}$$

This, together with Equation 3.4, implies that the forward rate over a finite time interval of length $T' - T$ is the *average* of the instantaneous forward rates over this interval:

$$R(T,T'|\,t) = -\frac{1}{T'-T}\ln\frac{B(t,T')}{B(t,T)} = \frac{1}{T'-T}\int_T^{T'} f(t,s)ds \tag{15.4}$$

The relation between the forward rates and zero bond yields (spot rates over finite time intervals of length $T - t$) can be established as well by inserting the explicit form of the discount factor for continuous compounding into Equation 15.2 and taking the derivative with respect to T :

$$f(t,T) = -\frac{\partial\left(\ln\exp\left[-R(t,T)(T-t)\right]\right)}{\partial T} = \frac{\partial\left[R(t,T)(T-t)\right]}{\partial T}$$

And after application of the product rule

$$f(t,T) = R(t,T) + (T-t)\frac{\partial R(t,T)}{\partial T} \tag{15.5}$$

The forward rates are greater than the spot rates for $\partial R(t,T)/\partial T > 0$, i.e. for term structures[3] whose values increase with T.

Finally, we make note of the relationship between the instantaneous forward rate and the *instantaneous* spot rate. This is simply

$$r(t) = \lim_{T\to t} f(t,T) \tag{15.6}$$

In anticipation of the following sections we stress here that all of the above equations hold in any arbitrary probability measure, i.e. irrespective of any choice of numeraire, since they have been derived directly from the definitions of the instantaneous interest rates.

15.2 Important Numeraire Instruments

As was shown in Section 14 in great detail, the value V of an arbitrary, tradable interest rate instrument[4] normalized with an arbitrary, tradable financial instrument

[3] As a reminder: the term structure is by definition the spot rate curve $R(t,T)$ as a function of maturity T.

[4] In spot rate models, every interest rate instrument can be interpreted as a derivative V on the underlying $S(t) = r(t)$ or $S(t) = \ln r(t)$.

Y is a process $Z = V/Y$ which, according to Equation 14.14, is a martingale

$$Z(t) = \mathrm{E}_t^Y [Z(u)] \qquad \forall u \geq t$$

The martingale measure at time t with respect to which the expectation E_t^Y is calculated depends on the choice of the numeraire instrument Y. In principle, any arbitrary tradable instrument can be used as the *numeraire*. An appropriate choice of numeraire, however, is a deciding factor in enabling an elegant solution of specific problems to be found. Two numeraire instruments are particularly popular (not only for term structure models but other models as well). These will be introduced in the following sections.

15.2.1 The Risk-Neutral Measure

We define $A(t_0, t)$ as the value of a *bank account* or *money market account*. This is the value one monetary unit (for example, 1 euro) has at time t, if it was invested at a time $t_0 < t$ and was subsequently always compounded at the current spot rate, with the interest earnings being immediately reinvested in the same account at the current spot rate. Intuitively, we could imagine reinvesting ever decreasing interest payments earned over ever shorter interest periods, the number of interest periods finally approaching infinity. The value of such an account can be written in terms of the instantaneous spot rate as

$$A(t_0, t) = \exp\left[\int_{t_0}^{t} r(s)ds\right] \tag{15.7}$$

In the *risk-neutral measure* this bank account, Equation 15.7, is used as the numeraire.

$$Y(t) = A(t_0, t) = \exp\left[\int_{t_0}^{t} r(s)ds\right] \qquad \text{for any arbitrary } t_0 < t$$

This numeraire has the advantage of satisfying the important property specified in Equation 14.33. Since

$$\begin{aligned} Y(t+dt) &= A(t_0, t+dt) \\ &= \exp\left[\int_{t_0}^{t+dt} r(s)ds\right] = \exp\left[r(t)dt + \int_{t_0}^{t} r(s)ds\right] \\ &= e^{r(t)dt} \exp\left[\int_{t_0}^{t+dt} r(s)ds\right] = e^{r(t)dt} Y(t) \end{aligned}$$

we have

$$dY(t) = Y(t+dt) - Y(t) = \left(e^{r(t)dt} - 1\right) Y(t) \approx r(t)Y(t)dt$$

where, in the last step of the derivation, the approximation $e^x \approx 1 + x$ (obtained from the Taylor series expansion of the exponential function valid for small values of x) was used. For infinitesimal x (and thus for infinitesimal dt in the above equation) equality holds. Both $r(t)$ and $Y(t)$ are known at time t (even if they are stochastic variables

and as such not yet known for the next time step $t+dt$). This implies that the process $m(t) = r(t)Y(t)$ is previsible as required in Equation 14.33.

With this numeraire the martingale property Equation 14.14 becomes

$$\frac{V(t)}{A(t_0,t)} = \mathrm{E}_t^A\left[\frac{V(u)}{A(t_0,u)}\right] \quad \forall u \geq t \geq t_0$$

Note that the initial time point t_0 in the money account can be chosen at will. For every t_0 we obtain another, different risk-neutral measure. Setting $t_0 = t$ and using the fact that $A(t,t) = 1$ reduces the above expression to

$$V(t) = \mathrm{E}_t^A\left[\frac{V(u)}{A(t,u)}\right] = \mathrm{E}_t^A\left[e^{-\int_t^u r(s)ds}V(u)\right] \quad \forall u \geq t \tag{15.8}$$

which directly yields the price of the financial instrument. In words:

With respect to the risk-neutral measure, today's value of a financial instrument is equal to the expectation of the discounted future value.

This is *not* the same as the discounted future expectation. In this measure, the discounting is performed *first* and *then* the expectation is calculated. Discounting means *division* by the numeraire.

We take as an example, the value at time t of a zero-coupon bond with maturity u, i.e. we set $V(t) = B(t,u)$ and consequently $V(u) = B(u,u) = 1$, in Equation 15.8 to obtain

$$B(t,u) = \mathrm{E}_t^A\left[\frac{1}{A(t,u)}\right] = \mathrm{E}_t^A\left[e^{-\int_t^u r(s)ds}\right] \tag{15.9}$$

The bond price is the expectation with respect to the risk neutral measure of the reciprocal of the bank account. Comparing this with Equation 15.3, which always holds, yields the relationship between the instantaneous forward rates and the future instantaneous spot rates with respect to the risk-neutral measure.

$$e^{-\int_t^u f(t,s)ds} = \mathrm{E}_t^A\left[e^{-\int_t^u r(s)ds}\right]$$

The *future price* of an instrument V is $V(u)$ for $u > t$ and is unknown at time t. It is well known that this must be distinguished from the *forward price* $V(t,u)$, which is known at time t since it follows from arbitrage arguments (see Equation 7.1). If the instrument under consideration pays no dividends between t and u, the forward price is

$$V(t,u) = \frac{V(t)}{B(t,u)} \quad \text{mit} \quad u > t \tag{15.10}$$

The forward price, Equation 15.10, of an interest rate instrument with respect to the risk-neutral measure is given by

$$\frac{V(t)}{B(t,u)} = \frac{1}{B(t,u)}\mathrm{E}_t^A\left[\frac{V(u)}{A(t,u)}\right] = \mathrm{E}_t^A\left[e^{-\int_t^u (r(s)-f(t,s))ds}V(u)\right] \quad \forall u \geq t \tag{15.11}$$

where Equation 15.3 was used, exploiting the fact the instantaneous forward rates are by definition known at time t and thus can be included in or taken out of the expectation operator as desired.

15.2.2 The Forward-Neutral Measure

For the *forward-neutral measure*, a zero bond is used as the numeraire (see Equation 15.3)

$$Y(t) = B(t,T) = \exp\left[-\int_t^T f(t,s)ds\right] \qquad \text{for any arbitrary } T > t$$

This numeraire has the property prescribed in Equation 14.33 as well:

$$\begin{aligned} Y(t+dt) = B(t+dt,T) &= \exp\left[-\int_{t+dt}^T f(t,s)ds\right] \\ &= \exp\left[-\int_t^T f(t,s)ds + \underbrace{f(t,t+dt)}_{r(t) \text{ for } dt\to 0}dt\right] \\ &= e^{r(t)dt}\exp\left[-\int_t^T f(t,s)ds\right] = e^{r(t)dt}Y(t) \end{aligned}$$

thus

$$dY(t) = Y(t+dt) - Y(t) = \left(e^{r(t)dt} - 1\right)Y(t) \approx r(t)Y(t)dt$$

The Taylor series expansion of the exponential function $e^x \approx 1+x$ was used here again. Likewise, equality holds for infinitesimal x (and thus for infinitesimal dt). Both $r(t)$ and $Y(t)$ are known values at time t, implying the previsibility of $m(t) = r(t)Y(t)$ as required in Equation 14.33.

With this choice, the martingale property equation 14.14 becomes

$$\frac{V(t)}{B(t,T)} = \mathrm{E}_t^B\left[\frac{V(u)}{B(u,T)}\right] \quad \forall T \geq u \geq t$$

Observe that the maturity T of the zero bond can be selected arbitrarily. For each choice of T, we obtain a different normalizing factor and thus another forward-neutral measure. Setting $T = u$ and using the fact that $B(u,u) = 1$ we obtain the following price for a financial instrument

$$V(t) = B(t,u)\mathrm{E}_t^B\left[V(u)\right] = e^{-\int_t^u f(t,s)ds}\mathrm{E}_t^B\left[V(u)\right] \quad \forall u \geq t \tag{15.12}$$

Today's value of a financial instrument is equal to the discounted expectation of its future value taken with respect to the forward neutral measure.

The expectation is *first* taken with respect to this measure and *then* discounted. Discounting means *multiplication* by the numeraire.

The forward price, Equation 15.10, for an interest rate instrument is, with respect to this measure, exactly equal to the expected future price, hence the name"forward-neutral"

$$\frac{V(t)}{B(t,u)} = \mathrm{E}_t^B\left[V(u)\right] \quad \forall u \geq t \tag{15.13}$$

15.3 The Special Case of Deterministic Interest Rates

The reader may be somewhat confused that, in earlier Chapters (see in particular Equation 11.20), prices of financial instruments were always calculated by discounting the future expectation using $B_R(t,T)$, which, from the above discussion, would indicate that the *forward*-neutral measure was used in the calculations. The measure was, however, always referred to as the *risk*-neutral measure. In those chapters however, interest rates were always assumed to be completely deterministic (or in many cases even constant). We will now show that for deterministic interest rates, the forward-neutral and the risk-neutral measures are identical.

Consider a portfolio consisting of a zero bond $B(t,T)$ and a loan made at time t to finance the purchase of the zero bond. The interest rate on this loan is floating and is always equal to the current spot rate for each interest period δt. After one such period δt, the loan debt will have grown to $B(t,T)\exp(r(t)\delta t)$. The portfolio thus constructed has no value at time t and the evolution of the entire portfolio over one time step δt is known exactly at time t, and therefore involves no risk. Because arbitrage is not possible and no risk is taken, the portfolio must have no value for all later times as well. Thus, at time $t+\delta t$ the value of the zero bond must be equal to the value of the loan

$$B(t+\delta t,T) = B(t,T)\exp(r(t)\delta t)$$

So far, everything is as it was as for stochastic interest rates. On the right-hand side of the above equation there appear only terms which are known at time t. The difference comes in the next time step: for deterministic interest rates, the spot rates at all later times $u > t$ are known at time t as well. After *two* interest periods, the credit debt will have grown to $B(t,T)\exp(r(t)\delta t)\exp(r(t+\delta t)\delta t)$ with a *known* interest rate $r(t+\delta t)$. Again, no risk is taken since the rate $r(t+\delta t)$ is already known at time t. Therefore, because the market is arbitrage free, the portfolio must still be worthless at time $t+2\delta t$ and the loan must therefore still equal the value of the zero bond. Proceeding analogously over n time steps, we obtain the value of the zero bond as

$$B(t+n\delta t,T) = B(t,T)\exp\left[\sum_{i=0}^{n-1}(r(t+i\delta t)\delta t)\right]$$

Taking the limit as $\delta t \to 0$, the value of the bond at time $u := t+n\delta t$ is

$$B(u,T) = B(t,T)\exp\left[\int_t^u r(s)ds\right] \quad \text{for} \quad u \geq t$$

This holds for every $u \geq t$, in particular for $u = T$. Thus, observing that $B(T,T) = 1$, we obtain the price of a zero bond for deterministic interest rates:

$$B(t,T) = \exp\left[-\int_t^T r(s)ds\right] = \frac{1}{A(t,T)} \tag{15.14}$$

where in the last step the definition of a bank account, Equation 15.7, was used. Hence, for deterministic interest rates, the numeraire $B(t,T)$ associated to the forward-neutral measure is equal to the reciprocal $A(t,T)^{-1}$ of the numeraire of the risk-neutral measure. Substituting this into Equation 15.8, we obtain

$$V(t) = \mathrm{E}_t^A\left[\frac{V(u)}{A(t,u)}\right] = \mathrm{E}_t^A\left[B(t,u)V(u)\right] = B(t,u)\mathrm{E}_t^A\left[V(u)\right]$$

where in the last step we have made use of the fact that $B(t,u)$ is known at time t, is therefore not stochastic and can be factored out of the expectation. The price of a financial instrument must, however, be independent of the measure used in its computation. In consequence, comparison of this equation with Equation 15.12 immediately yields the equation

$$B(t,u)\mathrm{E}_t^B\left[V(u)\right] = V(t) = B(t,u)\mathrm{E}_t^A\left[V(u)\right]$$

and thus $\mathrm{E}_t^B\left[V(u)\right] = \mathrm{E}_t^A\left[V(u)\right]$. This implies that both measures are identical if interest rates are deterministic.

The fundamental difference between the general case and deterministic interest rates is that the price of a zero bond in Equation 15.14 is given by the *future spot rates*. Of course, the general equation 15.3 stating that the price of a zero bond is given by the *current forward rates* continues to hold. We might suspect that this is closely related to the fact that the future instantaneous spot rates must equal the current instantaneous forward rates if interest rates are deterministic. This is in fact the case since the derivative of Equation 15.14 with respect to T gives

$$r(T) = -\frac{\partial}{\partial T}\ln B(t,T)$$

Comparing this with Equation 15.2, which holds in general, yields

$$f(t,T) = r(T) \quad \forall t,T \quad \text{for } t \leq T$$

Thus, if interest rates are deterministic, the instantaneous forward rates are indeed equal to the (known) future instantaneous spot rates. Since the right-hand side of this equation is not dependent on t, this must be true for the left-hand side as well. Hence, if interest rates are deterministic, the instantaneous forward rates are independent of the present time t.

15.4 Tradable and Non-Tradable Variables

As was emphasized at the end of the Section 14.2, both the numeraire Y and the financial instrument V *must* be **tradable** instruments in order for the martingale property Equation 14.1 to hold. Otherwise, potential arbitrage opportunities cannot be exploited by trading and in this way fail to violate the assumption of an arbitrage-free market. This realization, which may seem trivial from our modern perspective, was in the past by no means trivial. In fact, there exist pricing methods which assume that non-tradable variables have the martingale property. The mistake made in doing

so is then (approximately) corrected after the fact by making a so-called *convexity adjustment* (see Section 15.5).

One example of a non-tradable variable is the *yield* of a tradable instrument whose price is a non-linear function of its yield. Take, for example, a zero bond B with lifetime τ. The yield r of the zero bond depends non-linearly on its price (except if the linear compounding convention has been adopted; see Table 3.4). For example, for simple compounding the price of the zero bond is $B = (1 + r\tau)^{-1}$. The zero bond is obviously a tradable instrument. Therefore its future expectation taken with respect to the forward-neutral measure as in Equation 15.13 must be equal to its forward price.

$$B(T, T+\tau|\, t) = \mathrm{E}_t^B\left[B(T, T+\tau)\right] = \mathrm{E}_t^B\left[\frac{1}{1 + r(T, T+\tau)\tau}\right] \neq \frac{1}{1 + \mathrm{E}_t^B\left[r(T, T+\tau)\right]\tau}$$

The inequality in the above expression is meant to clarify that the expectation of the price is *not* equal to the price calculated with the expectation of the yield because of the non-linearity in the relation between the price and the yield.

On the other hand, the forward rate for the time period between T and $T + \iota$ is by definition equal to the yield of a forward zero bond over this period; for linear compounding explicitly:

$$\frac{1}{1 + r_f(T, T+\tau|\, t)\tau} \equiv B(T, T+\tau|\, t)$$

Comparing this with the above expression for the forward price of the bond gives

$$\frac{1}{1 + r_f(T, T+\tau|\, t)\tau} \neq \frac{1}{1 + \mathrm{E}_t^B\left[r_\tau(T)\right]\tau}$$
$$\Longrightarrow$$
$$r_f(T, T+\tau|\, t) \neq \mathrm{E}_t^B\left[r_\tau(T)\right]$$

The forward rate is thus *not* equal to the future expectation of the spot rate taken with respect to the forward-neutral measure. Because of Equation 15.12, this means that the zero bond yields are not martingales with respect to the forward-neutral measure. This contradicts the fact that *all tradable* instruments are martingales with respect to the forward-neutral measure. Hence, zero bond yields are *not* tradable.

For example, the 3-month LIBOR rate as the yield of a 3-month zero bond is not directly tradable, but the 3-month zero bond of course is. This is the deeper reason why, for example, in *forward rate agreements* or *caplets* and *floorlets* (which, of course, are tradable), the difference between the future LIBOR and the strike is not paid out at the LIBOR fixing at the beginning of the relevant interest period but rather is paid out at the *end* of that period (see Sections 17.1 and 18.4.3, respectively). This has the effect that we are not really dealing with a forward contract (or an option) on a future interest rate, but rather with a forward contract (or an option) on future *zero bonds,* i.e. on *tradable* instruments. The same holds for *swaps, caps* and *floors,* which are nothing other than a series of forward rate agreements, caplets and floorlets, respectively, strung together.

15.5 Convexity Adjustments

According to Equation 15.12, today's price $V(t)$ of a financial instrument normalized with respect to the forward-neutral numeraire[5] is equal to its discounted future expectation. If the future time for which the expectation is calculated is chosen to be the maturity date T of the instrument, then the price of the instrument is equal to the discounted future expectation of its *payoff profile* $V(T)$. For instruments whose payoff profiles are *linear* functions of the underlying S, i.e. $V(S,T) = a + bS$, the expectation of the *payoff profile* is equal to the payoff profile of the expectation of the *underlying*:

$$\begin{aligned} \mathrm{E}\left[V(S,T)\right] &= \mathrm{E}\left[a + bS\right] \\ &= \int_{-\infty}^{\infty} \left[a + bS\right] p(S) dS \\ &= a \int_{-\infty}^{\infty} p(S) dS + b \int_{-\infty}^{\infty} S\, p(S) dS \\ &= a + b\mathrm{E}_t^B\left[S\right] \end{aligned} \tag{15.15}$$

where p denotes the probability density of the pertinent martingale measure.

Although the prices of most instruments (for example, the zero bond) are non-linear functions of their underlyings, there do exist transactions with linear payoff profiles for which Equation 15.15 holds, for example forward contracts. The payoff profile of a forward contract with maturity T and delivery price K is known to be $S(T) - K$. The expectation of this payoff profile is simply the expectation of the underlying less the delivery price:

$$\mathrm{E}\left[V(S,T)\right] = \mathrm{E}\left[S(T) - K\right] = \mathrm{E}\left[S(T)\right] - K$$

If the underlying S of the forward contract is itself a *tradable* instrument (for example, a stock), we can now go a step further. With respect to the forward-neutral normalization (or in the case of deterministic interest rates), the future expectation of any tradable instrument is equal to its current forward price in accordance with Equation 15.13, i.e. $\mathrm{E}_t^B\left[S(T)\right] = S(t,T)$. Furthermore, Equation 15.12 states that with respect to this measure, today's price is equal to the discounted future expectation. So in summary (and in agreement with Equation 7.4):

$$\begin{aligned} V(S,t) &= B(t,T)\mathrm{E}_t^B\left[V(S,T)\right] \\ &= B(t,T)\left(\mathrm{E}_t^B\left[S(T)\right] - K\right) \\ &= B(t,T)(S(t,T) - K) \end{aligned}$$

In the first of the above equations, we used Equation 15.12. In the second line we made use of the fact that the payoff profile is a linear function of the underlying and hence, that Equation 15.15 holds, while in the third equation, we finally used of the namesake property of the forward-neutral measure, Equation 15.13.

The mistake made in some traditional pricing methods corresponds to precisely this last step, i.e. simply replacing the future expectation of the underlying as in

[5] or when interest rates are deterministic also in the risk-neutral measure

Equation 15.13 with the forward price of the underlying even if the underlying is *not* a tradable instrument (and as such, is not a martingale with respect to the forward-neutral measure and does not satisfy Equation 15.13). This mistake is then corrected (approximately) after the fact by a *convexity adjustment*.

The convexity adjustment is defined as the difference between the future expectation of the underlying (with respect to the forward-neutral measure) and the forward price of the underlying

$$\text{Convexity Adjustment} \equiv \mathrm{E}_t^B\left[S(T)\right] - S(t,T) \tag{15.16}$$

The most important relevant case is when the underlying is an interest rate. As was shown in Section 15.4, zero bond yields are (except when strictly linear compounding is used) not tradable instruments because the expression relating these yields to the price of the (tradable) instrument under consideration, namely the zero bonds, is non-linear, as indicated in Table 3.4:

$$B_r(t,t+\tau) = \begin{cases} \exp\left(-r\tau\right) & \text{Continuous} \\ \left(1+r\right)^{-\tau} & \text{Discrete} \\ \left(1+r\tau\right)^{-1} & \text{Simple} \\ 1-r\tau & \text{Linear} \end{cases} \tag{15.17}$$

We now want to determine an approximation for the convexity adjustment of the yield of such a zero bond, i.e. we want to determine

$$\text{Convexity Adjustment } = \mathrm{E}_t^B\left[r(T,T+\tau)\right] - r_f(T,T+\tau|\,t)$$

Here we have written out explicitly the time dependence of the the zero bond yields: $r(T,T+\tau)$ is the (unknown future) spot rate at time T for a period of length τ starting at time T. The forward rate for that same period, as known at time t, is denoted by $r_f(T,T+\tau|\,t)$. To find the expectation of the unknown future spot rate we expand the (also unknown) future bond price $B_r(T,T+\tau)$ as a Taylor series up to second order about the (known) forward rate $r_f = r(T,T+\tau|\,t)$:

$$\begin{aligned} B_r(T,T+\tau) &= \sum_{n=0}^{\infty} \frac{1}{n!}\left[r(T,T+\tau) - r_f(T,T+\tau|\,t)\right]^n \left.\frac{\partial^n B_r(T,T+\tau)}{\partial r^n}\right|_{r=r_f} \\ &\approx B_{r_f} + \left[r-r_f\right] B_r'|_{r=r_f} + \frac{\left[r-r_f\right]^2}{2} B_r''|_{r=r_f} \end{aligned}$$

where in the last line we have dropped all time arguments for ease of notation. We now calculate the forward-neutral expectation of this bond price:

$$\underbrace{\mathrm{E}_t^B\left[B_r\right]}_{B_{r_f}} \approx B_{r_f} + B_r'|_{r=r_f}\,\mathrm{E}_t^B\left[r-r_f\right] + \frac{1}{2}\,B_r''|_{r=r_f}\,\underbrace{\mathrm{E}_t^B\left[(r-r_f)^2\right]}_{\approx \mathrm{var}[r]} \tag{15.18}$$

Here, use is made of the fact that B and its derivative are evaluated at r_f which is known at time t. Therefore, B and its derivative can be factored out of the expectation.

On the left-hand side appears the expectation of the bond price with respect to the forward-neutral measure. This is, since bonds are tradable, exactly the forward bond price and, by the definition of the forward rate r_f, is identically equal to $B_{r_f}(T, T+\tau)$.

The expectation $\mathrm{E}_t^B\left[(r-r_f)^2\right]$ is approximately equal to the variance of $r(T, T+\tau)$; it would be exactly this variance if r_f were equal to $\mathrm{E}_t^B[r]$. To express this variance of $r(T, T+\tau)$ in terms of values known at time t, the variance of $r(T, T+\tau)$ is approximated by the variance of the *forward rate* $r_f(T, T+\tau|\, t)$:

$$\mathrm{E}_t^B\left[(r-r_f)^2\right] \approx \mathrm{var}\,[r] \approx \mathrm{var}\,[r_f] = r_f^2\sigma_f^2\,(T-t) \tag{15.19}$$

Here the volatility σ_f of the forward rate r_f, called the *forward volatility*, appears. This is (at least in principle) known at time t.

Substituting all of the above into Equation 15.18 yields

$$0 \approx B_r'|_{r=r_f}\left(\mathrm{E}_t^B[r]-r_f\right) + \frac{1}{2}\,B_r''|_{r=r_f}\,r_f^2\sigma_f^2\,(T-t)$$

This can now be solved for the desired expectation of the future spot rate:

$$\mathrm{E}_t^B[r] \approx r_f \underbrace{-\frac{1}{2}r_f^2\sigma_f^2\,(T-t)\frac{B_r''|_{r=r_f}}{B_r'|_{r=r_f}}}_{\text{Convexity Adjustment}} \tag{15.20}$$

Within these approximations, i.e. by expanding the bond price up to second order (see Equation 15.18) and with the approximations in Equation 15.19, the future expectation of the interest rate can thus be approximated by the forward rate adjusted by the amount:

$$\begin{aligned}
&\mathrm{E}_t^B[r(T, T+\tau)] - r_f(T, T+\tau|\, t) \\
&\approx -\frac{1}{2}r_f(T, T+\tau|\, t)^2\sigma_f^2\,(T-t)\frac{B_r''(T, T+\tau)|_{r=r_f}}{B_r'(T, T+\tau)|_{r=r_f}} \\
&= \begin{cases}
\frac{1}{2}r_f(T, T+\tau|\, t)^2\sigma_f^2\,(T-t)\tau & \text{Continuous} \\
\frac{1}{2}r_f(T, T+\tau|\, t)^2\sigma_f^2\,(T-t)\frac{\tau(\tau+1)}{1+r_f} & \text{Discrete} \\
r_f(T, T+\tau|\, t)^2\sigma_f^2\,(T-t)\frac{\tau}{1+\tau r_f} & \text{Simple} \\
0 & \text{Linear}
\end{cases}
\end{aligned} \tag{15.21}$$

Here, the convexity adjustments for all compounding conventions listed in Equation 15.17 have been explicitly calculated. As expected, there is no convexity adjustment in the case of strictly linear compounding. This is consistent since, for this form of compounding, the interest rate is a linear combination of zero bonds and cash. In other words, it is a portfolio consisting of tradable instruments and as such also tradable. Thus, only for the strictly linear compounding convention is the interest rate itself a martingale and does Equation 15.13 hold in the forward-neutral measure.

15.5.1 LIBOR-In-Arrears Swaps

As an example, we consider a forward contract on a zero bond yield. As was mentioned at the end of Section 15.4, problems regarding tradability do not arise for normal forward rate agreements since the interest rate (difference) fixed at the *beginning* of an interest period is paid at the *end* of the period. As a result, such contracts are effectively forward contracts on (tradable) zero bonds. The same holds for plain vanilla swaps since such a swap can be interpreted as a portfolio of forward rate agreements.

But there are swaps for which the difference between the future LIBOR and the fixed side is paid at the same time when the LIBOR rate is fixed. Or, to put it the other way around: the interest payable is determined only at the time when payment is to be made, i.e. at the *end* of the corresponding interest period. Therefore such instruments are called *LIBOR-In-Arrears Swaps*. Here indeed the underlying is directly the LIBOR rate which is *not* a tradable instrument. If we nevertheless wish to price such an instrument as if we could replace the future expectation of the underlying (with respect to the forward-neutral measure) with the forward rate, the resulting error must be corrected, at least approximately, by the convexity adjustment.

For the sake of simplicity, we consider only one period of a LIBOR-In-Arrears Swap, a *LIBOR-In-Arrears Forward Rate Agreement (FRA)*[6] so to speak. This has a principal N, a fixed interest rate K and extends over a period from T to $T+\tau$. A potential compensation payment is calculated by means of simple compounding and flows at the beginning of this period, i.e. directly at time T when the LIBOR rate $r(T,T+\tau)$ is fixed. Such an FRA has a payoff profile given by

$$V(r(T,T+\tau),T) = N\tau\left[r(T,T+\tau)-K\right] \tag{15.22}$$

This is a *linear* function of the underlying r. The expectation of the payoff profile is thus simply

$$\mathrm{E}\left[V(r(T,T+\tau),T)\right] = N\tau\mathrm{E}\left[r(T,T+\tau)\right] - N\tau K \tag{15.23}$$

and its value today is this expectation, taken with respect to the forward-neutral measure, discounted back to today

$$V(r(T,T+\tau),t) = B(t,T)N\tau\left(\mathrm{E}_t^B\left[r(T,T+\tau)\right]-K\right)$$

Up to now, all the equations are exact. The calculation of the future expectation is now performed either using a term structure model or an approximation by means of the convexity adjustments given in Equation 15.20:

$$\begin{aligned}
&V(r(T,T+\tau),t)\\
&= B(t,T)N\tau\left(\mathrm{E}_t^B\left[r(T,T+\tau)\right]-K\right)\\
&\approx B(t,T)N\tau\left(r_f(T,T+\tau|\,t) - \frac{1}{2}r_f(T,T+\tau|\,t)^2\sigma_f^2\,(T-t)\frac{B_r''(T,T+\tau)|_{r=r_f}}{B_r'(T,T+\tau)|_{r=r_f}} - K\right)\\
&= B(t,T)N\tau\left(r_f(T,T+\tau|\,t) + r_f(T,T+\tau|\,t)^2\sigma_f^2\,(T-t)\frac{\tau}{1+\tau r_f} - K\right)
\end{aligned} \tag{15.24}$$

[6] A LIBOR-In-Arrears Swap is simply a portfolio consisting of such FRAs.

The convexity adjustment for simple compounding has been used in the last step, since the specified instrument prescribes this compounding convention. Thus for r_f, we must use the forward rate *with respect to simple compounding* as well.

Again, we emphasize that above considerations hold only for those instruments whose payoff profiles are *linear* functions of the underlying. Only then does Equation 15.15 hold and only then does the expectation of the *underlying* come into play. For instruments with *non*-linear payoff profiles, on the other hand, the expectation of the *payoff profile* must be calculated directly. For example, for a plain vanilla call the expectation of the payoff profile is, in contrast to Equation 15.15

$$\begin{aligned}\mathrm{E}\left[\max\left\{S(T)-K,0\right\}\right] &= \int_{-\infty}^{\infty} \max\left\{S-K,0\right\} p(S)dS \\ &= \int_{K}^{\infty} (S-K)p(S)dS \neq \max\left\{\mathrm{E}\left[S(T)\right]-K,0\right\}\end{aligned}$$

15.5.2 Money Market Futures

The above example of a LIBOR-In-Arrears FRA may appear a bit academic to the reader. In reality though, *money market futures* are among the most actively traded interest derivatives and are nothing other than LIBOR-In-Arrears FRAs traded on an exchange, most commonly based on the 3-month LIBOR or the EURIBOR.

A money market future with a nominal N, a fixed rate K over a period from T to $T+\tau$ yields (theoretically) at maturity T a compensation payment calculated using simple compounding as in Equation 15.22, thus

$$V(T) = N\tau\left[r(T,T+\tau)-K\right]$$

Since this instrument has a *future-styled* payment mode, the changes in the position's value do not remain unrealized until maturity T, but are *immediately* realized on a margin account. As explained in Section 7.1.4, this has the effect that today's value of a futures position is directly equal to the future expectation of the payoff profile *without discounting*. This expectation of the future payoff profile is given by Equation 15.23. The value of a money market future at time t is then, in contrast to Equation 15.24, simply:

$$V(t) = N\tau\left(\mathrm{E}_t^B\left[r(T,T+\tau)\right]-K\right) \tag{15.25}$$

This is obviously equal to zero when $K = \mathrm{E}[r(T,T+\tau)]$. The fixed rate K of a money market future is in fact always chosen so that the value of the contract is zero at the time when the contract is entered into. The fixed rate K of a money market futures contracted at time t thus gives directly the information on the opinion of the market on the value of $\mathrm{E}[r(T,T+\tau)]$, in other words, on the *future expectation* of the interest rate. Since the interest rate is not a tradable instrument, this is *not* equal to the forward rate.

Money market futures are often used in constructing spot rate curves for maturities ranging from approximately three months to two years (we refer the reader to Part V). To construct spot rate curves from such contracts, the *forward* rates associated with these transactions are needed, since we can use these to calculate the spot rates

quite easily by utilizing Equation 3.3, for example. To determine the forward rate $r_f(T, T+\tau|\,t)$ from the market's opinion on the future LIBOR *expectation* obtained from quotes on money market futures, the convexity adjustment must be *subtracted* from the expectation as in Equation 15.16 to obtain

$$\begin{aligned} r_f(T, T+\tau|\,t) &= \mathrm{E}_t^B\left[r(T, T+\tau)\right] - \text{Convexity Adjustment} \\ &\approx \mathrm{E}_t^B\left[r(T, T+\tau)\right] + \frac{1}{2} r_f(T, T+\tau|\,t)^2 \sigma_f^2\,(T-t)\frac{B_r''(T, T+\tau)|_{r=r_f}}{B_r'(T, T+\tau)|_{r=r_f}} \\ &= \mathrm{E}_t^B\left[r(T, T+\tau)\right] - r_f(T, T+\tau|\,t)^2 \sigma_f^2\,(T-t)\frac{\tau}{1+\tau r_f} \end{aligned} \tag{15.26}$$

where in the last step the approximation in Equation 15.21 for the convexity adjustment for the linear compounding convention is used. This is a non-linear equation which can be solved numerically for the unknown r_f.

As has been often emphasized, the convexity adjustments presented above are only approximations since they derive from an approximation of the theoretical value of $\mathrm{E}_t^B\left[r(T, T+\tau)\right]$, see Equation 15.20). The future expectation $\mathrm{E}_t^B\left[r(T, T+\tau)\right]$ can, however, be calculated by other means, for example, using a term structure model. We then obtain *another* expression for $\mathrm{E}_t^B\left[r(T, T+\tau)\right]$, and not Equation 15.20. Since the forward rates are determined solely from arbitrage considerations (independent of any model) as in Equation 3.3, this implies that also another expression for the convexity adjustment is obtained. The convexity adjustment is thus dependent on the method (the term structure model) being used. To ensure consistency, the convexity adjustment for the money market futures used in constructing the spot rate curves should be consistent with the term structure model applied for pricing.

Quotation for Money Market Futures

In Europe, money market futures are traded primarily on the *LIFFE*. Futures on the 3-month LIBOR in pound sterling (*short sterling future*), in US dollars (*euro dollar future*), in euros (*euro EUR future*) and in Swiss franc (*euro swiss future*) are available for trade on this exchange. Futures in euros on the 3-month EURIBOR and on the 1-month EURIBOR are available for trade on the *EUREX*.

Money market futures are quoted in a way which takes some getting used to. Not the delivery price $K = \mathrm{E}_t^B\left[r(T, T+\tau)\right]$ is quoted, i.e. the fixed interest rate for which the future has no value, but rather

$$\text{Quote}_{\text{Money Market Future}} = 100\% - \mathrm{E}_t^B\left[r(T, T+\tau)\right]$$

A quote of 96.52%, for example, for a money market future means that in the opinion of the market, the expectation (in the forward neutral measure) for the future 3-month rate is $\mathrm{E}[r(T, T+\tau)] = 3.48\%$. The value at time t of a futures position contracted at time $t = 0$ with $K = \mathrm{E}_0^B\left[r(T, T+\tau)\right]$ is given, according to Equation 15.25, by

$$\begin{aligned} V(t) &= N\tau\left(\mathrm{E}_t^B\left[r(T, T+\tau)\right] - \mathrm{E}_0^B\left[r(T, T+\tau)\right]\right) \\ &= N\tau\underbrace{\left(1 - \mathrm{E}_0^B\left[r(T, T+\tau)\right]\right)}_{\text{Quote at Time } t=0} - N\tau\underbrace{\left(1 - \mathrm{E}_t^B\left[r(T, T+\tau)\right]\right)}_{\text{Quote at Time } t} \end{aligned}$$

A money market future on a 3-month LIBOR (i.e. $\tau = 1/4$) with a nominal amount of $N = 1,000,000$ euros which was agreed to at a quoted price of 96.52% and which is currently quoted as 95.95% is thus valued at

$$\begin{aligned} V(t) &= 1,000,000\text{EUR}\frac{1}{4}\left(96.52\% - 95.95\%\right) \\ &= 1,425\text{EUR} \end{aligned}$$

This amount is deposited in a margin amount. A change in the value of this position is directly reflected by a corresponding daily adjustment in the balance of the margin account.

15.6 Arbitrage-Free Interest Rate Trees

Construction of arbitrage-free *tree models* begins with the assumption that the martingale property Equation 14.14 holds, justifying the name arbitrage free[7]. An appropriate normalizing factor (numeraire) is selected and the integrals necessary for the computation of the expectations (with respect to the chosen measure) are discretized in a tree-structure. This procedure will be demonstrated explicitly in this section for *1-factor short rate models*, i.e. for models which have the instantaneous short rate, defined in Equation 15.1, as their one and only stochastic driver.

We will use the risk-neutral measure and the associated numeraire, which is the bank account. The price of every interest rate instrument is then given by Equation 15.8. We discretize first with respect to time by partitioning the time axis in intervals of length δt taking this length to be so small that the (stochastic) short rate can be assumed to be constant over this interval. Then Equation 15.8 for $u = t + \delta t$ becomes

$$\begin{aligned} V(t) &= \mathrm{E}_t^A\left[e^{-\int_t^{t+\delta t} r(s)ds} V(t+\delta t)\right] \\ &\approx \mathrm{E}_t^A\left[e^{-r(t)\delta t} V(t+\delta t)\right] \\ &= e^{-r(t)\delta t}\mathrm{E}_t^A\left[V(t+\delta t)\right] \\ &= B(t,t+\delta t)\,\mathrm{E}_t^A\left[V(t+\delta t)\right] \end{aligned} \tag{15.27}$$

In calculating the integral, we have made use of the assumption that r is approximately constant on the interval of integration. Variables which are known at time t can be factored out of the expectation. According to the last equation, the risk-neutral price and the forward-neutral price (see Equation 15.12) cannot be distinguished from one another over the very *short* time interval δt. This is in agreement with Section 15.3, since r is taken to be constant (in particular, deterministic) over this short time interval. The equality, however, does not hold for longer time spans, since r changes randomly from one time interval δt to the next. Thus, globally, we are still within the framework of the risk-neutral measure even though the local equations may look "forward neutral".

[7] More traditional models, known as *equilibrium models*, will not be investigated here. Our discussion will be restricted to arbitrage-free pricing methods.

15.6.1 Backward Induction

What we have accomplished up to this point is to factor the numeraire (the bank account) out of the expectation. It now remains to calculate the expectation of the instrument's price. To this end, we discretize the continuous range of V after a time step of length δt into finitely many values, i.e. starting from the value of V at time t, the value should be allowed to take on only finitely many different values after the next time step. Allowing *two* different values generates a *bi*nomial tree, *three* a *tri*nomial tree, etc. Since the financial instrument V under consideration is an interest rate instrument, the different values potentially taken on by V at time $t + \delta t$ result directly from the different possible interest rate term structures which might exist at $t + \delta t$. Since one of the model assumptions was that the evolution of the entire interest rate curve is driven by the instantaneous short rate, it follows that the different values attained by V are ultimately determined by the values this short rate can take on.

We will work below with binomial trees. We assume that the short rate increases to the value r_u or decreases to the value r_d with a probability p and $1-p$, respectively. We have more than one possibility at our disposal to ensure that the martingale property is satisfied (and thus eliminating arbitrage). Either we fix the values r_u and r_d and selects the probability p accordingly so that the market is governed by an arbitrage-free measure (this, for example, is done in finite difference methods where the grid is given at the onset of the analysis), or we specify the probability p first and subsequently select appropriate values for r_u and r_d. We will take the second path in our discussion here. We set

$$p = 1/2 \tag{15.28}$$

and determine the value of the short rate (i.e. the discount factors) on all nodes of the tree so that the short rate process as described by the tree guarantees arbitrage freedom at time t. The binomial tree with $p = 1/2$ allows the expectation in Equation 15.27 to be written as

$$\begin{aligned} V(t) &\approx B(t, t+\delta t)\, \mathrm{E}_t^A \left[V(t+\delta t)\right] \\ &\approx B(t, t+\delta t)\, \left[p\, V(r_u, t+\delta t) + (1-p)\, V(r_d, t+\delta t)\right] \\ &= B(t, t+\delta t)\, \left[\frac{1}{2} V(r_u, t+\delta t) + \frac{1}{2} V(r_d, t+\delta t)\right] \\ &= B(t, t+\delta t)\, \left[\frac{1}{2} V_u + \frac{1}{2} V_d\right] \end{aligned} \tag{15.29}$$

where in the last step, the short form notation $V_u := V(r_u, t+\delta t)$, etc. has been introduced.

One time step later, the short rate branches again (and in consequence, the price of the financial instrument does as well): $r_u \to r_{u\,u}$ with probability p, and $r_u \to r_{u\,d}$ with probability $1-p$, and similarly for r_d. In addition, note that at time $t + \delta t$, two *different* discount factors $B(t+\delta t, t+2\delta t)$ appear, according to whether the short rate rose to r_u or fell to r_d in the previous step. To emphasize the difference, the discount factors are indexed with the associated short rate. V_u, for example, is then expressed

as

$$\begin{aligned} V_u &\approx B_{r_u}(t+\delta t, t+2\delta t)\, \mathrm{E}_t^A\left[V_u(t+2\delta t)\right] \\ &\approx B_{r_u}(t+1\delta t, t+2\delta t)\, \left[\frac{1}{2}V(r_{uu}, t+2\delta t) + \frac{1}{2}V(r_{ud}, t+2\delta t)\right] \\ &= B_u\, \left[\frac{1}{2}V_{uu} + \frac{1}{2}V_{ud}\right] \end{aligned}$$

where in the last step the short form notation $V_{uu} := V(r_{u\,u}, t+2\delta t)$, analogous to that introduced above, has been used, and also the short form notation $B_u := B_{r_u}(t+1\delta t, t+2\delta t)$. Analogously,

$$V_d \approx B_d\, \left[\frac{1}{2}V_{du} + \frac{1}{2}V_{dd}\right]$$

Substituting this into Equation 15.29, the price $V(t)$ given by a binomial tree with two steps can be calculated as

$$V(t) \approx B(t, t+\delta t)\, \left[\frac{1}{2}B_u\, \left[\frac{1}{2}V_{uu} + \frac{1}{2}V_{ud}\right] + \frac{1}{2}B_d\, \left[\frac{1}{2}V_{du} + \frac{1}{2}V_{dd}\right]\right] \tag{15.30}$$

Proceeding analogously, $V_{dd} \approx B_{dd}\, \left[\frac{1}{2}V_{ddu} + \frac{1}{2}V_{ddd}\right]$, etc., the price at time t calculated from the prices three time steps later is given by

$$\begin{aligned} V(t) \approx B(t, t+\delta t)\, &[\frac{1}{2}B_u\, [\frac{1}{2}B_{uu}\, [\frac{1}{2}V_{uuu} + \frac{1}{2}V_{uud}] + \frac{1}{2}B_{ud}\, [\frac{1}{2}V_{udu} + \frac{1}{2}V_{udd}]] \\ &+ \frac{1}{2}B_d\, [\frac{1}{2}B_{du}\, [\frac{1}{2}V_{du\,u} + \frac{1}{2}V_{dud}] + \frac{1}{2}B_{dd}\, [\frac{1}{2}V_{ddu} + \frac{1}{2}V_{ddd}]]] \end{aligned} \tag{15.31}$$

and so on. Calculating backwards through the tree (*backward induction*) is completely analogous to the treatment of options on stocks and exchange rates with binomial trees as described in Section 11.

In order to ensure that any arbitrage opportunity has been eliminated, the short rates $r_u, r_d, r_{u\,ud}$ (and in consequence, the discount factors $B_u, B_d, B_{u\,ud}$), etc. must be chosen so that the *prices calculated* using the tree actually agree with the *market prices* of traded instruments. In particular, *zero bonds* (and thus today's term structure) must be exactly reproduced. However, at each time point $t + i\delta t$, there are just as many unknowns as a (non-recombinant) binomial tree has nodes, namely 2^i starting with $r_{u\,u\,\ldots u}$ continuing through all permutations of up and down moves until $r_{dd\,\ldots d}$. The number of unknowns increases exponentially with the number of time steps! So many conditions cannot conceivably be generated by the market prices of tradable instruments. In particular, no such functional relation between the number of interest rate instruments on the market and the number of time steps (depending only on the numerical implementation) in a tree could possibly exist. Therefore, based on this principle, another condition must first be established preventing the exponential growth of the number of unknowns with respect to increasing i.

Requiring that the tree be *recombinant* for all financial instruments with path-independent payoff profiles presents itself as a good candidate for the above mentioned

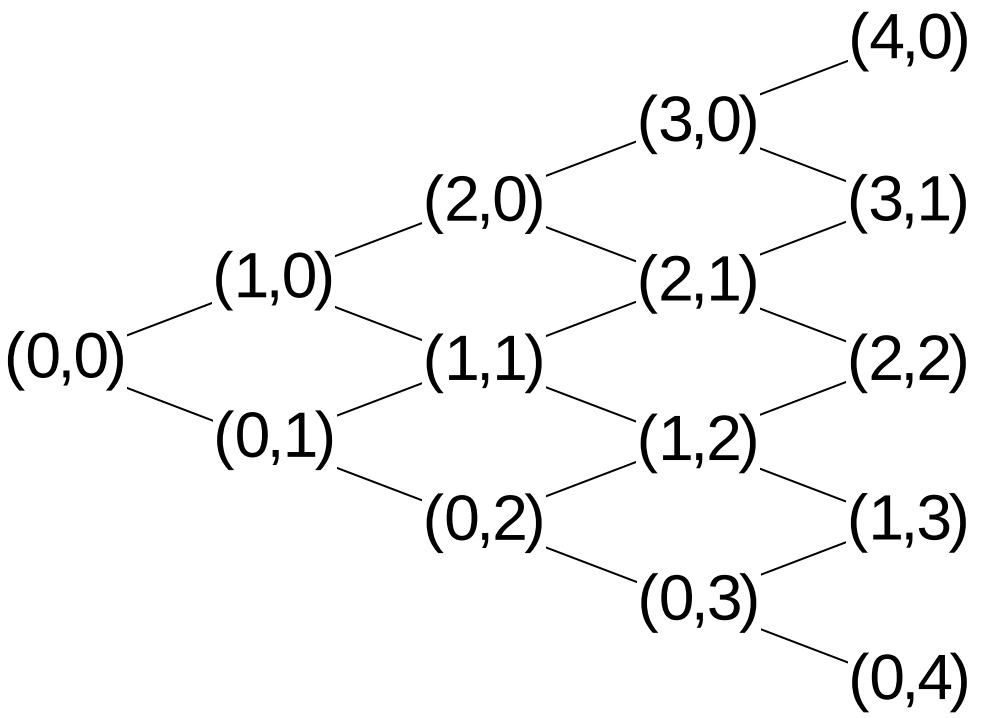


Figure 15.1: The binomial tree with the indexing showing the number of up and down moves required to get to the nodes starting from node $(0,0)$. For instance it takes two up moves and one down move to get to the node $(2,1)$.

condition, i.e. that $V_{ud} = V_{du}$, $V_{u\,ud} = V_{udu} = V_{du\,u}$, etc. This can only happen if the tree for the underlying also recombines. A recombinant binomial tree is known to have only $i+1$ nodes after i steps. This fact will be accounted for in our notation. The nodes of the tree at which we arrive having traveled upwards i times and downwards j times is uniquely determined by the ordered pair (i,j), irrespective of the order in which these upward and downward steps were taken. This is because the tree is recombinant. Therefore we denote the value of the financial instrument at this node with $V(i,j)$ and the nodes themselves with the ordered pairs (i,j) as presented in Figure 15.1. For example, $V(1,2) = V_{udd} = V_{dud} = V_{ddu}$, etc. We use the same notation for the short rate and for the zero bonds evaluated at the nodes:

$$\begin{aligned}(i,j) &= \text{Node after } i \text{ up moves and } j \text{ down moves}\\ r(i,j) &= \text{Instantaneous short rate at the node } (i,j)\\ V(i,j) &= \text{Value of an interest rate instrument at the node } (i,j)\\ B(i,j) &= \exp\{-r(i,j)\delta t\}\end{aligned} \tag{15.32}$$

In particular, $B(t, t+\delta t) = B(0,0)$. In this notation, the prices of (path independent) financial instruments can be written after one binomial step in a recombinant tree as

$$\begin{aligned} V(t) &= V(0,0) \\ &\approx V(1,0)\frac{1}{2}B(0,0) \\ &+ V(0,1)\frac{1}{2}B(0,0) \end{aligned} \tag{15.33}$$

Of course, this equation holds not only for the node $(0,0)$ but for any arbitrary node (i,j) in the tree

$$V(i,j) \approx B(i,j)\ \left[\frac{1}{2}V(i+1,j) + \frac{1}{2}V(i,j+1)\right] \tag{15.34}$$

This equation applied to $V(1,0)$ and $V(0,1)$ yields the price after two binomial steps

$$\begin{aligned} V(t) \approx\ & V(2,0)\frac{1}{4}B(0,0)\ B(1,0) \\ &+ V(1,1)\frac{1}{4}B(0,0)\ [B(1,0) + B(0,1)] \\ &+ V(0,2)\frac{1}{4}B(0,0)\ B(0,1) \end{aligned} \tag{15.35}$$

and after three

$$\begin{aligned} V(t) \approx\ & V(3,0)\frac{1}{8}B(0,0)B(1,0)B(2,0) \\ &+ V(2,1)\frac{1}{8}B(0,0)\,[B(1,0)B(2,0) + B(1,1)\,[B(1,0) + B(0,1)]] \\ &+ V(1,2)\frac{1}{8}B(0,0)\,[B(0,1)B(0,2) + B(1,1)\,[B(1,0) + B(0,1)]] \\ &+ V(0,3)\frac{1}{8}B(0,0)B(0,1)B(0,2) \end{aligned} \tag{15.36}$$

an so on. The expressions were purposely written in terms of the prices at the last node in each branch.

15.6.2 Forward Induction and Green's Functions

Although backward induction was simplified greatly by the requirement that the tree be recombinant, we are still not yet prepared to calculate an instrument's price since the instantaneous discounting factors $B(i,j)$ at the nodes $(i,j) \neq (0,0)$ remain unknown. Before we construct a tree for the instantaneous discount factors (i.e. for the short rate, see Equation 15.32) via a procedure, known as *forward induction*, on the basis of arbitrage considerations, we introduce a class of extremely useful "artificial" instruments. One such artificial instrument whose value at time t is denoted by $G(i,j)$ pays by definition one monetary unit if and only if the underlying (the short rate)

attains the tree node (i, j) at time $t + (i + j)\delta t$. This is the node at which we arrive having traveled upwards i times and downward j times, regardless of the order in which the upward and downward moves occurred as the tree is recombinant. $G(i, j)$ then is the value at time t of a single monetary unit paid out at one single node of the tree, namely at node (i, j). In this sense, G is the system's reaction to a perturbation of magnitude one at a single point in the system. This is analogous to the Green's functions in physics. We will therefore refer to G as a *Green's function*. By definition,

$$G(0, 0) \equiv 1 \tag{15.37}$$

In order to establish further values of the Green's function, we just have to set the value $V(i, j)$ in the above Equations 15.33, 15.35 and 15.36 equal to one at exactly one node and zero on all the other nodes. Equation.15.33 then yields

$$G(1, 0) = \frac{1}{2}B(0, 0) = G(0, 1) \tag{15.38}$$

From Equation 15.35 we obtain

$$\begin{aligned}
G(2, 0) &= \frac{1}{4}B(0, 0)\ B(1, 0) = \frac{1}{2}G(1, 0)B(1, 0) \\
G(0, 2) &= \frac{1}{4}B(0, 0)\ B(0, 1) = \frac{1}{2}G(0, 1)B(0, 1) \\
G(1, 1) &= \frac{1}{4}B(0, 0)\ [B(1, 0) + B(0, 1)] = \frac{1}{2}G(1, 0)B(1, 0) + \frac{1}{2}G(0, 1)B(0, 1)
\end{aligned}$$

and finally Equation 15.36 gives

$$\begin{aligned}
G(3, 0) &= \frac{1}{2}G(2, 0)B(2, 0) \\
G(0, 3) &= \frac{1}{2}G(0, 2)B(0, 2) \\
G(2, 1) &= \frac{1}{2}G(2, 0)B(2.0) + \frac{1}{2}G(1, 1)B1, 1) \\
G(1, 2) &= \frac{1}{2}G(0, 2)B(0, 2) + \frac{1}{2}G(1, 1)B(1, 1)
\end{aligned}$$

The following general recursion relation can be easily verified (this will be proven for an even more general case later)

$$\begin{aligned}
G(i, j) &= \frac{1}{2}G(i, j - 1)B(i, j - 1) + \frac{1}{2}G(i - 1, j)B(i - 1, j) \quad \text{for } \ i > 0, j > 0 \\
G(i, 0) &= \frac{1}{2}G(i - 1, 0)B(i - 1, 0) \\
G(0, j) &= \frac{1}{2}G(0, j - 1)B(0, j - 1)
\end{aligned} \tag{15.39}$$

The prices of all path-independent interest rate instruments can be represented as linear combinations of the Green's function evaluated at diverse nodes on the tree

since each payment profile of the form $f(r,T)$ can be distributed on the nodes (i,j) as appropriate:

$$f(r,T) \to f_T(i,j) := f(r(i,j),T) \quad \text{for} \quad t+(i+j)\delta t = T \quad \forall i,j$$

The value at time t of each individual payment $f_T(i,j)$ at node (i,j) is naturally equal to the Green's function belonging to this node (which has a value of exactly one monetary unit) multiplied by the number of monetary units that are to be paid, i.e. multiplied by $f_T(i,j)$. The total value of the financial instrument $V(t)$ with this payoff profile is then simply

$$V(t,T) = \sum_{(i,j)} f_T(i,j)\, G(i,j) \quad \text{for} \quad t+(i+j)\delta t = T \quad \forall i,j$$

where $\sum_{(i,j)}$ denotes "the sum over the nodes (i,j)". This can be immediately generalized to instruments with payoff profiles defined on *arbitrary* nodes (which need not all lie in the set of nodes corresponding to a time T)

$$V(t) = \sum_{(i,j)} f(i,j)\, G(i,j) \quad \text{for arbitrary payoff profiles} \quad f(i,j) \tag{15.40}$$

Path independence here is therefore *not* the restriction that the payoff profile depend only on a time point T. The payoff profile can depend on (the interest rate at) all possible nodes at all times, not however, on the path taken to arrive at these nodes (since this information is not available in the recombinant tree).

As a simple example, we consider the value of a zero bond $B(t,T)$, an instrument that pays one monetary unit at time T, regardless of the state of the underlying

$$\begin{aligned} B(t,T) &= \sum_{(i,j)} G(i,j) \quad \text{for} \quad t+(i+j)\delta t = T \\ &= \sum_{i=0}^{n} G(i,n-i) \quad \text{for} \quad n = \frac{T-t}{\delta t} \end{aligned} \tag{15.41}$$

This *theoretical* price has to exactly match the *market* price of the zero bond to prevent arbitrage. If the current term structure is available (for example, because it has been constructed on the basis of traded benchmark bonds for some maturities and interpolations in between) this term structure yields the *market* price for all zero bonds $B(t,T)$. In particular it yields the market prices for all those zero bonds maturing at times $t+i\delta t$ corresponding to the time steps of the binomial tree. Those have to be matched by Equation 15.41. Hence, we obtain one single condition for each time step $t+i\delta t$. For example, for $i=1$

$$\begin{aligned} B(t,t+1\delta t) &= \sum_{(i,j)} G(i,j) \quad \text{with} \quad i+j=1 \\ &= G(1,0)+G(0,1) \\ &= B(0,0) \end{aligned}$$

where in the last step, the Green's function as in Equation 15.38 is introduced into the equation. This simply checks for consistency: the discount factor at the first node of our tree (the right-hand side) must be equal to the market price for the zero bond with maturity $T = t + \delta t$ (left-hand side). It becomes more interesting at the next maturity date:

$$\begin{aligned} B(t, t+2\delta t) &= \sum_{(i,j)} G(i,j) \quad \text{where} \quad i+j=2 \\ &= G(2,0) + G(1,1) + G(0,2) \\ &= \frac{1}{2}G(1,0)B(1,0) + \frac{1}{2}G(1,0)B(1,0) + \frac{1}{2}G(0,1)B(1,0) + \frac{1}{2}G(0,1)B(1,0) \\ &= G(1,0)B(1,0) + G(0,1)B(0,1) \end{aligned}$$

We have used Equation 15.39 to calculate backwards from the Green's function at the time corresponding to $i + j = 2$ to the previous time step, i.e. to the time for which $i + j = 1$. The Green's function $G(1, 0)$ at the earlier time step $t + 1\delta t$ is already known (see Equation 15.38). The resulting equation is the arbitrage condition for the instantaneous discount factors $B(1, 0)$ and $B(0, 1)$ at the tree nodes $(1, 0)$ and $(0, 1)$. This procedure can be generalized to n time steps. In accordance with Equation 15.41 we write

$$B(t, t + n\delta t) = \sum_{i=0}^{n} G(i, n-i)$$

At this point we separate the boundary terms from the rest of the sum, since these obey another recursion in Equation 15.39. Now applying Equation 15.39, we see that

$$\begin{aligned} B(t, t+n\delta t) &= G(0,n) + \sum_{i=1}^{n-1} G(i, n-i) + G(n,0) \\ &= \frac{1}{2}G(0,n-1)B(0,n-1) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n-1} G(i, n-i-1)B(i, n-i-1) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n-1} G(i-1, n-i)B(i-1, n-i) \\ &\quad + \frac{1}{2}G(n-1,0)B(n-1,0) \end{aligned}$$

The boundary terms combine this expression into a simple sum. To see this, we use the index $k = i - 1$ in the second sum

$$\begin{aligned} B(t, t+n\delta t) &= \frac{1}{2}G(0,n-1)B(0,n-1) + \frac{1}{2}\sum_{i=1}^{n-1} G(i, n-i-1)B(i, n-i-1) \\ &\quad + \frac{1}{2}\sum_{k=0}^{n-2} G(k, n-k-1)B(i-1, n-k-1) + \frac{1}{2}G(n-1,0)B(n-1,0) \end{aligned}$$

Both boundary terms now have exactly the form needed to extend the index range in both sums to 0 through $n-1$. This means that the no arbitrage requirement can be represented as a simple recursion formula by means of the Green's function:

$$B(t, t+n\delta t) = \sum_{i=0}^{n-1} G(i, n-i-1)B(i, n-i-1) \tag{15.42}$$

Remember: on the left we have the market price of a zero bond, while on the right we have the values to be determined for the interest rate tree, i.e. the instantaneous discount factors at the nodes. The values on the right-hand side are all in terms of nodes corresponding to the time point

$$t + [i + (n-i-1)]\,\delta t = t + (n-1)\delta t$$

i.e. the time step prior to the maturity $t + n\delta t$ of the zero bond on the left. The *Green's function* evaluated at this earlier time point has already been calculated in the previous iteration step so that Equation 15.42 represents an arbitrage condition for the instantaneous *discount factors* at the nodes at time point $t + (n-1)\delta t$. Having determined the discount factors, we can use the recursion for the Green's function, Equation 15.39, to determine the values of the Green's function at the time point $t + n\delta t$. These are subsequently used again in the arbitrage condition, Equation 15.42, to calculate the next discount factors for the time step $t + n\delta t$. These are then used to calculate the next Green's function values for $t + (n+1)\delta t$ utilizing Equation 15.39 and so on. In this way, an arbitrage-free interest rate tree is constructed using *forward* induction from today into the future. Nevertheless, Equation 15.42 contains several (to be precise n) unknown discount factors which cannot be uniquely determined by only *one* arbitrage condition (except, of course, for the case $n = 1$). In other words, the exact reproduction of the term structure at time t is already attained singly from the fact the instantaneous discount factors and the Green's function at each node satisfy Equations 15.42 and 15.39. But this does not fix the *numerical* values of all instantaneous discount factors - or short rates $r(i, j)$. Additional information (as we will see below, the volatility) extracted from the market as well as an explicit specification of a stochastic process of the form 3.16 for the short rate is needed to fix the numerical values of $r(i, j)$. Before we start on this point however, we will first introduce a few more concepts which hold in general, i.e. for every interest rate tree, irrespective of the specification of a specific stochastic process.

15.7 Market Rates vs. Instantaneous Rates

The valuation of financial instruments using the Green's function as in Equation 15.40 above is only possible for those instruments whose payoff profiles are functions of the *instantaneous* short rate[8]. This is not usually the case for *traded* instruments, however. The underlying for most traded instruments is usually the the 3 or 6-month LIBOR,

[8] Or for financial instruments whose payoff profile is independent of the short rate, such as a zero bond, for which $f(i, j) = 1$ for $t + (i + j)\Delta t = T$ and $f(i, j) = 0$ otherwise.

for example, which is not at all equal to the short rate in the tree when $\delta t = 1$ day, for example. In contrast to the chapters on stock or FX options, the stochastic process being simulated (the short rate) does *not* describe the evolution of the underlying (the 3-month LIBOR, for example) of the instrument to be priced! This problem can be overcome in some cases, as for floaters and forward rate agreements, since, as will be seen in in Equations 16.3 or 17.1, their prices at time t can be interpreted as a combination of zero bond prices and as such can be priced *exactly* using Equation 15.42. However, the decomposition into zero bonds will not be possible for options such as caps and floors on the 3-month LIBOR whose general payoff functions are given by Equation 18.4 and, adopting the usual market practice of using simple compounding over the cap period, through Equation 18.7 (or when interpreted as bond options by Equation 18.10). In such cases, before the payoff profile at nodes (m, n) with $t + (m + n)\delta t = T$ can be calculated, the value of the *underlying* on each of these nodes must first be determined. The question now is: how do we calculate the underlying of interest (for example, a 3-month rate, a 6-month rate, a swap rate, etc.) at all nodes corresponding to the exercise date T from the stochastic process (the tree) of the instantaneous short rate?

As soon as the payoff profile at all nodes is known, the value of all (path independent) financials instruments at time t, i.e. at the node $(0, 0)$, are directly given by Equation 15.40. As with the node $(0, 0)$, the value of all instruments (in particular, of all zero bonds and thus all interest rates over arbitrary interest periods) would be known at an *arbitrary* node (m, n) if the "Green's functions" were known for *that* node.

15.7.1 Arrow-Debreu Prices

Arrow-Debreu prices (*ADP*s for short) are generalized Green's functions whose reference point is a fixed but arbitrary node (m, n) rather than the origin node $(0, 0)$. The Arrow-Debreu price $G_{m,n}(i, j)$ is the value at node (m, n) of an instrument paying one monetary unit at node (i, j). The Arrow-Debreu prices at the node $(m = 0, n = 0)$ are, of course, simply the values of the Green's function introduced above:

$$G(i, j) = G_{0,0}(i, j)$$

It follows immediately from the geometry of the tree that a monetary unit at node (i, j) can generate non-zero prices at a node (m, n) only if (i, j) is attainable when starting from the node (m, n). We know that m up moves have already occurred at node (m, n). These up moves cannot be undone, even if the following steps consist only of down moves since the index m merely counts the number of up moves having been made up to this point. This implies that for all nodes (i, j) attainable from the starting node (m, n), the condition $i \geq m$ must hold. Likewise, n down moves have already occurred at node (m, n) which cannot be reversed. As before, this means that the condition $j \geq n$ must hold for all nodes (i, j) which are attainable from the starting point (m, n). All other Arrow-Debreu prices must be zero:

$$\begin{aligned} G_{m,n}(i, j) &= 0 \quad \forall m > i \\ G_{m,n}(i, j) &= 0 \quad \forall n > j \\ G_{i,j}(i, j) &= 1 \quad \forall i, j \end{aligned} \tag{15.43}$$

where the last property is included solely for the sake of completeness, being in itself trivial: one monetary unit at node (i, j) is, of course, worth exactly one monetary unit at this node. This corresponds to property Equation 15.37 of the Green's function.

A further fundamental property of Arrow-Debreu prices follows from the backward induction, Equation 15.34, which holds for the value of any instrument at any node in the tree. Setting either $V(i+1, j) = 1$ and $V(i, j+1) = 0$ or $V(i+1, j) = 0$ and $V(i, j+1) = 1$, yields the *instantaneous* Arrow-Debreu prices, i.e. the Arrow-Debreu prices over a time step of length δt

$$G_{i,j}(i+1, j) = \frac{1}{2}B(i, j) = G_{i,j}(i, j+1) \tag{15.44}$$

The instantaneous Arrow-Debreu prices are thus half the instantaneous discount factors. This accomplishes the first step in the calculation of the Arrow-Debreu prices. We now merely require a generalization of the recursion relation given in Equation 15.39 in order to determine all subsequent Arrow-Debreu prices. To this end, consider one monetary unit at node (i, j) at time $t+(i+j)\delta t$. Because of the binomial structure of the tree, non-zero Arrow-Debreu prices are generated by this monetary unit at two nodes in the previous time slice, that is at nodes $(i, j-1)$ and $(i-1, j)$. The *sum* of the ADPs of *both* of these prices at a still earlier node (m, n) must then be equal to the Arrow-Debreu price at node (m, n) of the whole, original monetary unit at node (i, j). This is illustrated in Figure 15.2. This means that the Arrow-Debreu prices obey the following recursion:

$$G_{m,n}(i, j) = G_{m,n}(i, j-1)G_{i,j-1}(i, j) + G_{m,n}(i-1, j)G_{i-1,j}(i, j)$$

Substituting the corresponding discount factors for both of the instantaneous Arrow-Debreu prices as given in Equation 15.44, the recursion relation becomes

$$G_{m,n}(i, j) = G_{m,n}(i, j-1)\frac{1}{2}B(i, j-1) + G_{m,n}(i-1, j)\frac{1}{2}B(i-1, j) \text{ for } i \geq m,\ j \geq n \tag{15.45}$$

Together with the fundamental properties in Equation 15.43, this recursion determines uniquely all Arrow-Debreu prices. For example, for $m = i$ or $n = j$ Equation 15.43 implies that one of the Arrow-Debreu prices on the right-hand side is equal to zero[9]. This allows the recursion to be carried out explicitly for these cases:

$$\begin{aligned} G_{i,n}(i, j) = G_{i,n}(i, j-1)\frac{1}{2}B(i, j-1) &= \frac{1}{2^{j-n}}\prod_{k=1}^{j-n} B(i, j-k) \\ G_{m,j}(i, j) = G_{m,j}(i-1, j)\frac{1}{2}B(i-1, j) &= \frac{1}{2^{i-m}}\prod_{k=1}^{i-m} B(i-k, j) \end{aligned} \tag{15.46}$$

In particular, both of these equations hold at the boundary of the tree. The first equation with $i = 0$ (i.e. no up move) represents the *lower* boundary, the second with $j = 0$ (i.e. no down move) the *upper* boundary of the tree.

[9]For $m = i$ we have $G_{i,n}(i-1, j) = 0$ and for $n = j$ we have $G_{m,j}(i, j-1) = 0$.

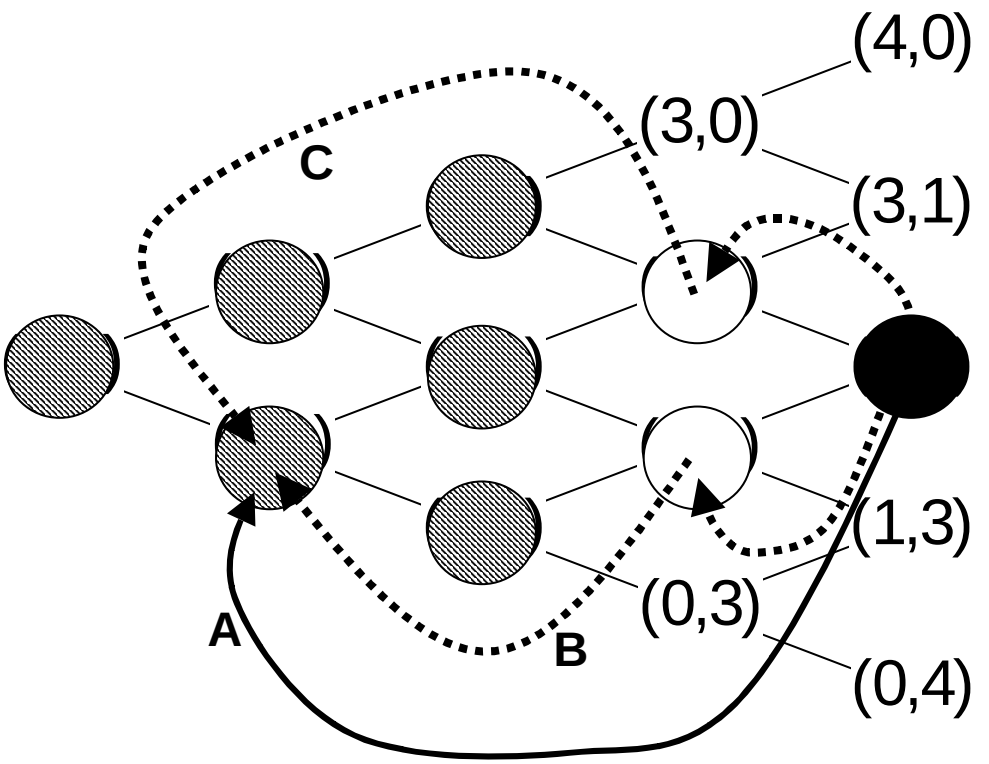


Figure 15.2: A monetary unit (black dot) at node $(2,2)$ generates ADPs at nodes $(2,1)$ and $(1,2)$ and also at all the earlier 'striped' nodes. The monetary unit must have the same influence (shown as line **A**) on a striped node, e.g. on node $(0,1)$, as both ADPs it has induced at nodes $(2,1)$ and $(1,2)$ together (shown as lines **B** and **C**).

With these Arrow-Debreu prices, the value of any arbitrary financial instrument with a payoff profile given by $f(i,j)$ at any *arbitrary* node (m,n) is, analogous to Equation 15.40, simply

$$V(m,n) = \sum_{(i,\,j)} G_{m,n}(i,j)\, f(i,j) \tag{15.47}$$

Making use of Equation 15.40, the value of this instrument at time t is then

$$V(t) = V(0,0) = \sum_{(m,n)} G(m,n) V(m,n) = \sum_{(m,n),(i,\,j)} G(m,n) G_{m,n}(i,j)\, f(i,j)$$

In particular, the value $B_\tau(m,n)$ of a zero bond at the node (m,n), whose time to maturity at this node is given by τ, is likewise given by Equation 15.47 where i and j

satisfy the following three conditions

$$\begin{aligned} i + j &= m + n + \tau/\delta t \\ i &\geq m \\ j &\geq n \end{aligned}$$

The first of the three conditions characterizes all nodes corresponding to the time $m + n + \tau/\delta t$, the maturity of the zero bond. The payoff profile $f(i, j)$ of the zero bond equals one at precisely these nodes and zero elsewhere. This condition allows j to be expressed in terms of i. The limits in the sum appearing in Equation 15.47 can be specified explicitly by the other two conditions which, as a result of Equation 15.43 must always be satisfied: the lower limit is $i \geq m$, the upper limit follows from $m + n + \tau/\delta t - i = j \geq n$ which can be rewritten as $i \leq m + \tau/\delta t$. The value of the zero bond at node (m, n) with a time to maturity of τ is thus given explicitly by

$$B_\tau(m, n) = \sum_{i=m}^{m+\tau/\delta t} G_{m,n}(i, m + n + \tau/\delta t - i) \tag{15.48}$$

Thus, at each node of the tree, a complete (future) term structure (i.e. future interest rates for arbitrary times to maturity) can be constructed from Arrow-Debreu prices, since the interest rate at node (m, n) for any arbitrary time to maturity τ is by definition (for continuous compounding) given by

$$r_\tau(m, n) = -\frac{1}{\tau} \ln B_\tau(m, n) \tag{15.49}$$

Making use of this fact, we are now in a position to price any arbitrary derivative on an underlying whose value can be derived from the term structure (as, for example, for the 3-month LIBOR rate or a swap rate).

15.7.2 Pricing Caplets Using Arrow-Debreu Prices

In anticipation of Chapter III, where caps and floors will be defined, we will demonstrate explicitly how the price of a caplet on a 3-month rate with principal N, strike rate K, exercise time (maturity) T and payment date $T' = T + \tau$ (with $\tau = 3$ months) can be expressed solely in terms of Arrow-Debreu prices. Observe that the stochastic process for the short rate need *not* be specified in order to do so. What will be shown here holds for *any arbitrary* arbitrage-free short rate model.

Consistent with Equation 18.7 for the payoff profile of a caplet, we will adopt the market convention of using *simple compounding* over a single caplet period (in this case 3 months). Equation 18.7 only holds for *this* compounding convention (for general compounding, Equation 18.4 holds). Thus, the relationship between the underlying r_τ to be substituted into the payoff profile Equation 18.7 and the associated zero bond is *not* given by Equation 15.49 but by the corresponding equation for simple compounding. The 3-month rate at which the payoff profile is to be evaluated at each

node (m, n) with $t + (m + n)\delta t = T$, or equivalently, $n = (T - t)/\delta t - m$ is thus:

$$\begin{aligned} r_\tau(m, a-m) &= \frac{1}{\tau}\left[B_\tau(m, a-m)^{-1} - 1\right] \\ &= \left(b\delta t \sum_{i=m}^{m+b} G_{m,a-m}(i, a+b-i)\right)^{-1} - \frac{1}{b\delta t} \\ &\text{for all } \; 0 \le m \le a \end{aligned} \tag{15.50}$$

where we have defined

$$a := (T - t)/\delta t \quad , \quad b := \tau/\delta t$$

This is the underlying of our caplet at all nodes relevant to the caplet's payoff profile. According to the payoff profile in Equation 18.7, the values of the caplet at exercise date T, i.e. at nodes $(m, a - m)$ with $0 \le m \le a$, are given by

$$f(m, a-m) = \tau N\, B_\tau(m, a-m) \max\left\{r_\tau(m, a-m) - K, 0\right\}$$

Both the discount factor B_τ appearing in the payoff profile as well as the underlying (the interest rate r_τ) can be expressed in terms of the Arrow-Debreu prices:

$$\begin{aligned} f(m, a-m) = N &\sum_{i=m}^{m+b} G_{m,a-m}(i, a+b-i) \\ &\times \max\left\{\left(\sum_{i=m}^{m+b} G_{m,a-m}(i, a+b-i)\right)^{-1} - 1 - \tau K, 0\right\} \end{aligned}$$

Equation 15.40 now directly yields the value of this payoff profile, i.e. the caplet value, at time t

$$\begin{aligned} c^{\text{cap}}(T, T+\tau, K\,|t) &= \sum_{m=0}^{a} G(m, a-m)\, f(m, a-m) \\ &= N \sum_{m=0}^{a} G_{0,0}(m, a-m) \sum_{i=m}^{m+b} G_{m,a-m}(i, a+b-i) \\ &\quad \times \max\left\{\left(\sum_{i=m}^{m+b} G_{m,a-m}(i, a+b-i)\right)^{-1} - 1 - \tau K, 0\right\} \end{aligned}$$

The price of the caplet has now been expressed completely in terms of the Arrow-Debreu prices. Since these prices can be determined using the recursion relations 15.45 and 15.46 with initial values given by Equations 15.43 and 15.44, this procedure can be used to price any interest rate instrument.

Practical Implementation of Arrow-Debreu Prices

Arrow-Debreu prices have four indices, two for the position of the cash flow in the tree and two for the position at which the effect of the cash flow is felt. The numerical

implementation thus requires in principle the construction of a four-dimensional field, which leads to computer memory problems for somewhat finer trees as well as performance problems in the computation of all the ADPs. It will prove to be unnecessary in most cases to compute all of the ADPs. In the above case of caplets, for example, it is only necessary to compute the ADPs of the form $G_{m,a-m}(i, a+b-i)$ with the fixed parameters a and b. Thus, only two of the four above mentioned indices are free, namely m and i. Consequently, it is entirely sufficient to determine the two-dimensional field

$$\widetilde{G}_{m,i} := G_{m,a-m}(i, a+b-i)$$

thereby reducing the numerical effort involved considerably. The effort is now only the same as that needed in the determination of the Green's function.

15.8 Explicit Specification of Short Rate Models

Up to this point, the pricing procedure has remained quite general in that no specific term structure model has been used. None of the relations introduced above can as yet be used to calculate an explicit numerical value since, as mentioned above, the arbitrage condition in Equation 15.42 is by no means sufficient to determine all the (unknown) discount factors appearing in this equation. From now on, we suppose that the term structure model is of the general form 3.16 where the stochastic variable is either the logarithm of the short rate or the short rate itself. In the second case, the short rate satisfies the equation

$$dr(t) = a\,(r,t)\,dt + b(r,t)\,dW \tag{15.51}$$

Models of this type are referred to as *normal models*. In the special case of the parameters a and b being independent of r, this process is called the *Ho-Lee Model*. Such models have the advantage that they are very easy to implement (for example, as a tree). The Ho-Lee model

$$dr(t) = a(t)dt + b(t)\,dW$$

for constant b is even analytically tractable. One of the disadvantages of this model is that it allows for negative interest rates and that the parameter b represents an absolute rather than a relative volatility, which makes it difficult to fit the model to the relative volatilities observed in the market.

The second class of short rate models does not specify a stochastic process for the short rate itself, but for its logarithm:

$$d\ln r(t) = a\,(r,t)\,dt + b(r,t)\,dW \tag{15.52}$$

These models are thus referred to as *lognormal models*. In the special case of the volatility b being independent of r, this process is called the *Black–Derman-Toy model*, although, strictly speaking, the actual Black-Derman-Toy model employs a very special form of drift (see Section 15.11.2 below). Such lognormal models have the disadvantage that they are much more difficult to implement and cannot be solved analytically. On the other hand, negative interest rates cannot occur in these models and the parameter

b represents the relative volatility, simplifying fitting this model to the relative (for example, cap) volatilities observed in the market. Such models can be transformed into models for r via the Ito formula[10], Equation 3.18:

$$dr(t) = \left[r(t)a\,(r,t) + \frac{1}{2} b(r,t)^2 r(t) \right] dt + r(t)\, b(r,t)\, dW \tag{15.53}$$

The numerical evaluation of lognormal models is easier to implement when written in this form.

15.8.1 The Effect of Volatility

According to the Girsanov Theorem (see Section 14.4), the drift $a(r,t)$ is uniquely determined by the probability measure used. We fixed this probability measure "by hand" when we required the condition specified in Equation 15.28 to be satisfied. In doing so, we implicitly fixed the drift as well. The drift can thus no longer be given as "input" into the specification of the stochastic process. Only the volatility remains as a parameter through which market information (in addition to the bond prices) may enter into our model for the determination of the discount factors in Equation 15.42. We will now show that if the volatility is given for the time step n, exactly $n-1$ conditions are generated on the n instantaneous discount factors at time n. Taken together with the arbitrage condition given by Equation 15.42, requiring that the observed market price of the zero bond with maturity $T = t + n\delta t$ be reproduced by the model, we have exactly as many conditions as unknowns and the interest rate tree can be uniquely constructed.

In general, from the viewpoint of the node (i,j), the variance of a variable x (we can think of x as representing, for example, the short rate r or its logarithm $\ln r$) is caused by its possible two different values in the next time step, either x_u or x_d. The expectation and variance for random variables of this type are given by[11]

$$\begin{aligned} \mathrm{E}\,[x] &= p x_u + (1-p) x_d \\ \mathrm{Var}\,[x] &= p(1-p)(x_u - x_d)^2 \end{aligned} \tag{15.54}$$

In particular, for the case $p = 1/2$ the variance is given by $\mathrm{Var}[x] = (x_u - x_d)^2/4$.

On the other hand, for models of the form given by Equations 15.51 and 15.52, the variance of the stochastic variable x (with $x = r$ resp. $x = \ln r$) over an interval of time of length δt is given by $b(r,t)^2 \delta t$.

[10] For this, we need to consider the following: $\ln r(t)$ in Equation 15.52 corresponds to S in Equation 3.18. For the function f we take $f(S) = e^S$ (since this is exactly r). The partial derivatives appearing in Equation 3.18 are then simply $\partial f/\partial t = 0$ and $\partial f/\partial S = \partial^2 f/\partial S^2 = f = r$.

[11] Substituting the expectation into the definition of the variance defined as the expectation of the squared deviation from the expectation gives

$$\begin{aligned} \mathrm{Var}\,[x] &= \mathrm{E}\left[(x - \mathrm{E}\,[x])^2 \right] \\ &= p\,(x_u - \mathrm{E}\,[x])^2 + (1-p)\,(x_d - \mathrm{E}\,[x])^2 \\ &= p\,(x_u - p x_u - (1-p)x_d)^2 + (1-p)\,(x_d - p x_u - (1-p)x_d)^2 \end{aligned}$$

Multiplying out and collecting terms yields the desired expression.

15.8.2 Normal Models

For models of the form in Equation 15.51 the variance of the short rate at node (i,j) must satisfy :

$$\begin{aligned} b(i,j)\sqrt{\delta t} = \sqrt{\text{Var}\left[r(i,j)\right]} &= \frac{1}{2}\left[r(i+1,j) - r(i,j+1)\right] \\ &\Longrightarrow \\ r(i+1,j) &= r(i,j+1) + 2b(i,j)\sqrt{\delta t} \end{aligned} \tag{15.55}$$

This enables us to establish a recursion formula (after performing the substitution $i+1 \to i$) for the instantaneous discount factors at all nodes corresponding to the time slice $n = i + j$

$$\begin{aligned} B(i,j) &= \exp\left\{-\, r(i,j)\delta t\right\} \\ &= \exp\left\{-\left[r(i-1,j+1) + 2b(i-1,j)\sqrt{\delta t}\right]\delta t\right\} \\ &= \exp\left\{-\, 2b(i-1,j)\sqrt{\delta t}\delta t\right\}\exp\left\{-\, r(i-1,j+1)\delta t\right\} \\ &= \exp\left\{-\, 2b(i-1,j)\sqrt{\delta t}\delta t\right\} B(i-1,j+1) \end{aligned}$$

or

$$B(i,j) \;= \alpha(i-1,j)B(i-1,j+1) \quad \text{ with } \; \alpha(i,j) = \exp\left\{-\, 2b(i,j)\delta t^{3/2}\right\} \tag{15.56}$$

Recursive substitution into this equation allows each instantaneous discount factor in the nodes corresponding to this time slice to be expressed in terms of a single discount factor at the "lowest" node $(0, i+j)$:

$$\begin{aligned} B(i,j) \;&= \alpha(i-1,j)B(i-1,j+1) \\ &= \alpha(i-1,j)\alpha(i-2,j+1)B(i-2,j+2) \\ &= \alpha(i-1,j)\alpha(i-2,j+1)\alpha(i-3,j+2)B(i-3,j+3) \\ &= \cdots \\ &= B(0,j+i)\prod_{k=1}^{i}\alpha(i-k,j+k-1) \end{aligned}$$

Observe that only α-parameters (volatility information) in the time slice $i-k+j+k-1 = i+j-1$ are required for the determination of this value, i.e. information from the *previous* time step. Volatility information for the value of the interest rate at the node actually being calculated (i.e. for the spacing of the nodes at the time slice $(i+j)$) is *not* required; these values are determined with volatility information from the immediately proceeding time! This (perhaps counter-intuitive) property is of course nothing other than the *previsibility* which we always require for the coefficients of dW and dt in all models of the form 3.16 or 14.16. At this point now, we get an intuitive picture (see also Figure 15.3) what it means that a and b in Equation 3.16

are previsible processes. This previsibility is the deeper mathematical reason for why we are able calculate anything at all.

Using this expression for $B(i,j)$ with the special choice $j = n-i-1$ and substituting it into the arbitrage condition, Equation 15.42, we obtain the arbitrage condition for the discount factor $B(0,n-1)$ at the lowest node of the time slice $(n-1)$:

$$B(t,t+n\delta t) = B(0,n-1)\sum_{i=0}^{n-1} G(i,n-i-1)\prod_{k=1}^{i}\alpha(i-k,n-i+k-2)$$

which, after a simple change of index $n \to (n+1)$, yields the arbitrage condition for the discount factor $B(0,n)$ at the lowest node of time slice n:

$$B(t,t+(n+1)\delta t) = B(0,n)\sum_{i=0}^{n} G(i,n-i)\prod_{k=1}^{i}\alpha(i-k,n-i+k-1) \qquad (15.57)$$

The left-hand side is the given *market* price of a zero bond with maturity at time slice $(n+1)$, which must be reproduced using zero bonds and the Green's function on time slice n. The α's on the right-hand side are all defined on time slice $(n-1)$ and are given by the volatility governing the process at this time point. This condition now actually contains only one unknown and can be solved easily for $B(0,n)$:

$$B(0,n) = \frac{B(t,t+(n+1)\delta t)}{\sum_{i=0}^{n} G(i,n-i)\prod_{k=1}^{i}\alpha(i-k,n-i+k-1)}$$

From this single value $B(0,n)$ we obtain all further instantaneous discount factors in this time slice by repeated application of Equation 15.56. Note that we find ourselves at time n at this point of the iteration. The instantaneous discount factors in the time slice n are determined from the price of the bond maturing at time $(n+1)$ and volatility information from the immediately preceding time step $(n-1)$. This is illustrated in Figure 15.3.

Having established the instantaneous discount factors, an expression for the instantaneous short rates results immediately from Equations 15.32 and 15.56

$$\begin{aligned} r(i,j) &= -\frac{1}{\delta t}\ln B(i,j) = -\frac{\ln B(i-1,j+1)}{\delta t} - \frac{\ln\alpha(i-1,j)}{\delta t} \\ &= r(i-1,j+1) + 2b(i-1,j)\sqrt{\delta t} \\ &= ... \\ &= r(0,j+i) + 2\sqrt{\delta t}\sum_{k=1}^{i} b(i-k,j+k-1) \end{aligned} \qquad (15.58)$$

Normal Models with r-Independent Volatility Structure

For volatility structures depending only on time (but not on the interest rate), the volatility values on time slice n are all identical. We can therefore set them equal to

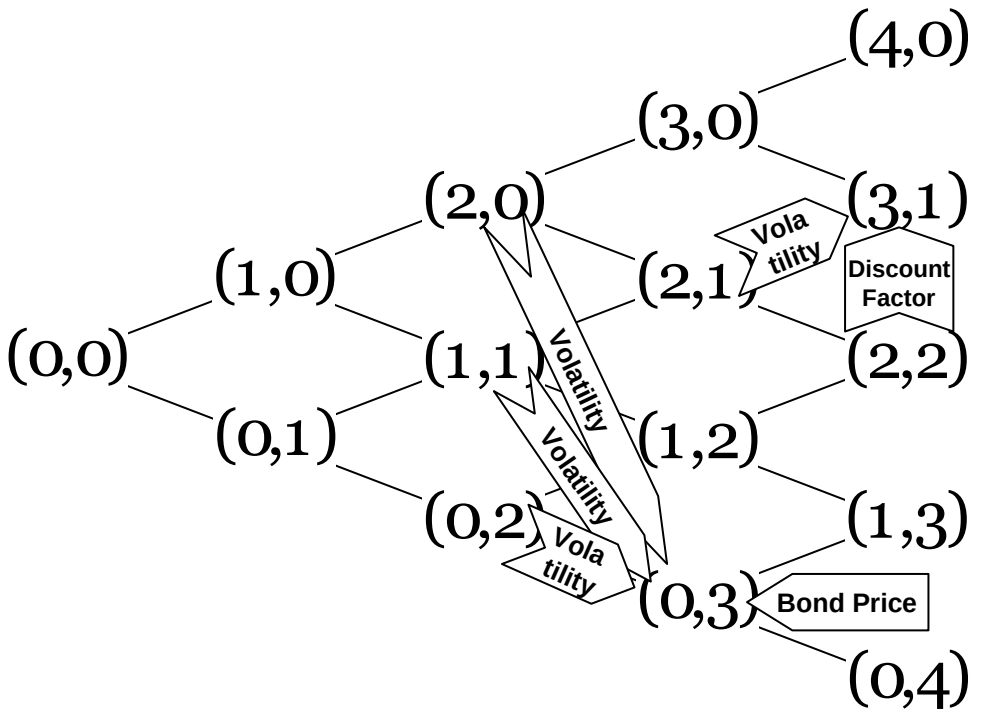


Figure 15.3: Flow of information when constructing the short rate tree. The discount factor at the lowest node needs the market price of the zero bond maturing one time step *later* and *all* volatility information from one time step *earlier*. For the other discount factors in the time slice it suffices to know the discount factors already calculated at lower nodes in the same time slice and the volatility at the neighboring node one time step earlier.

the volatility at the lowest node $(0, n)$. Thus

$$\begin{aligned} b(i,j) &= b(0,i+j) = b(0,n) \equiv \sigma(t+n\delta t) \\ \alpha(i,j) &= \alpha(0,i+j) = \alpha(0,n) \;\; \forall i,j \text{ with } \;\; i+j=n \end{aligned} \tag{15.59}$$

and Equation 15.56 reduces to

$$\begin{aligned} B(i,j) \; &= \alpha(0,n-1)B(i-1,j+1) \\ &= \alpha(0,n-1)^i \, B(0,n) \quad \text{with } \;\; i+j=n \end{aligned}$$

Equation 15.55 then implies that the short rate in the tree at time slice n changes from node to node by a constant term $2\sigma\sqrt{\delta t}$

$$\begin{aligned} r(i,j) &= r(i-1,j+1) + 2\sigma(t+(n-1)\delta t)\sqrt{\delta t} \\ &= r(0,n) + 2\,i\,\sigma(t+(n-1)\delta t)\sqrt{\delta t} \end{aligned} \tag{15.60}$$

The arbitrage condition for this discount factor at the lowest node of time slice n reduces to

$$B(t, t+(n+1)\delta t) = B(0,n)\sum_{i=0}^{n} G(i,n-i)\alpha(0,n-1)^i$$

15.8.3 Lognormal Models

An analogous procedure for lognormal models can be obtained as follows: in models of the form specified in Equation 15.52, $b(i,j)^2\delta t$ is the variance of the *logarithm* of the short rate. Therefore, as seen from the node (i,j), we have

$$b(i,j)\sqrt{\delta t} = \sqrt{\mathrm{Var}\left[\ln r(i,j)\right]} = \frac{1}{2}\left[\ln r(i+1,j) - \ln r(i,j+1)\right] = \frac{1}{2}\ln(\frac{r(i+1,j)}{r(i,j+1)}$$

$$\Longrightarrow$$

$$r(i+1,j) = r(i,j+1)\exp\left\{2b(i,j)\sqrt{\delta t}\right\} \tag{15.61}$$

This enables us to establish a recursion formula[12] (after performing the substitution $i+1 \to i$) for the instantaneous discount factors at all nodes in time slice $n = i+j$

$$\begin{aligned} B(i,j) &= \exp\left\{-r(i,j)\delta t\right\} \\ &= \exp\left\{-r(i-1,j+1)e^{2b(i-1,j)\sqrt{\delta t}}\delta t\right\} \\ &= \left[\exp\left\{-r(i-1,j+1)\delta t\right\}\right]^{e^{2b(i-1,j)\sqrt{\delta t}}} \\ &= B(i-1,j+1)^{e^{2b(i-1,j)\sqrt{\delta t}}} \end{aligned}$$

or

$$B(i,j) = B(i-1,j+1)^{\alpha(i-1,j)} \quad \text{with} \quad \alpha(i,j) = \exp\left\{+2b(i,j)\delta t^{1/2}\right\} \tag{15.62}$$

This is quite similar to the corresponding equation for normal models (Equation 15.56), but since the volatility information enters the equation as an exponent rather than a factor in the recursion, the structure is somewhat more complicated. Furthermore, a comparison with Equation 15.56 reveals a sign change in the expression for α. Nevertheless, a simple trick enables us to find a structure which is quite similar to that given in Equation 15.56. Taking logarithms in Equation 15.62 yields a recursion relation for the *logarithm* of the discount factors having the same structure as Equation 15.56:

$$\ln B(i,j) = \alpha(i-1,j)\ln B(i-1,j+1)$$

[12] Here we use the property $\exp\{ax\} = (\exp\{x\})^a$ of the exponential function with $a = e^{2b(i-1,j)\sqrt{\delta t}}$.

This recursion can be carried out explicitly giving

$$\begin{aligned}\ln B(i,j) &= \alpha(i-1,j)\ln B(i-1,j+1)\\ &= \alpha(i-1,j)\alpha(i-2,j+1)\ln B(i-2,j+2)\\ &= \alpha(i-1,j)\alpha(i-2,j+1)\alpha(i-3,j+2)\ln B(i-3,j+3)\\ &= \cdots\\ &= [\ln B(0,j+i)]\prod_{k=1}^{i}\alpha(i-k,j+k-1)\end{aligned}$$

and allowing the instantaneous discount factors in time slice n to be written as a function of the "lowest" node $(0, i+j)$:

$$\begin{aligned}B(i,j) &= \exp\left\{[\ln B(0,j+i)]\prod_{k=1}^{i}\alpha(i-k,j+k-1)\right\}\\ &= B(0,j+i)^{\prod_{k=1}^{i}\alpha(i-k,j+k-1)}\end{aligned}$$

Using this $B(i,j)$ for $j = n-i-1$ in the arbitrage condition Equation 15.42 we obtain, after performing an index transformation $n \to (n+1)$, the arbitrage condition for the discount factor $B(0,n)$ at the lowest node in time slice n.

$$\begin{aligned}B(t,t+(n+1)\delta t) &= \sum_{i=0}^{n} G(i,n-i)\exp\left\{[\ln B(0,n)]\prod_{k=1}^{i}\alpha(i-k,n-i+k-1)\right\} \qquad (15.63)\\ &= \sum_{i=0}^{n} G(i,n-i)B(0,n)^{\prod_{k=1}^{i}\alpha(i-k,n-i+k-1)}\end{aligned}$$

This can only be solved *numerically* for $B(0,n)$ using, for example, the well-known *Newton-Raphson method*[13]. From this value $B(0,n)$ we obtain all further instantaneous discount factors in this time slice by repeated application of Equation 15.62. We now have a similar situation as for normal models, see Figure 15.3: the discount factors in the time slice n are calculated from the price of the zero bond maturing in the *following*

[13]To solve a non-linear equation of the form $f(x) = 0$, the *Newton-Raphson method* uses the following iteration to find the points were the function f equals zero: having an estimate x_i for a zero of f, a better estimate is obtained from the formula

$$x_{i+1} = x_i - f(x_i)\left(\left.\frac{\partial f}{\partial x}\right|_{x=x_i}\right)^{-1}$$

We usually start the procedure with a rough estimate x_0 and iterate until the difference between x_{i+1} and x_i is sufficiently small for the required purpose. The iteration sequence converges if

$$\left| f\frac{\partial^2 f/\partial^2 x}{(\partial f/\partial x)^2}\right| < 1$$

holds in a neighborhood of the zero of f. This can always assumed to be the case in our applications.

time slice $(n+1)$ and from the volatility information from the immediately *preceding* time slice $(n-1)$.

Having established the instantaneous discount factors, the short rates follow immediately from Equations 15.32 and 15.62.

$$\begin{aligned} r(i,j) &= -\frac{1}{\delta t}\ln B(i,j) = -\alpha(i-1,j)\frac{\ln B(i-1,j+1)}{\delta t} \\ &= r(i-1,j+1)\alpha(i-1,j) \\ &= ... \\ &= r(0,j+i)\prod_{k=1}^{i}\alpha(i-k,j+k-1) \end{aligned} \tag{15.64}$$

Exact Reproduction of the Term Structure with the Lognormal Model

Alternatively, we can proceed from the process for the short rate in the lognormal model obtained from Ito's lemma, Equation 15.53, and derive all equations exactly in the form obtained for the normal model, in particular Equation 15.57, the only difference being that α is then given by

$$\alpha(i,j) = \exp\left\{-2b(i,j)r(i,j)\delta t^{3/2}\right\} \tag{15.65}$$

The $r(i,j)$ needed for this equation were calculated in the previous iteration step, so that all terms in Equation 15.57 are known. Thus, this equation can be solved *analytically* for the lowest bond, even though we are working within the context of the lognormal model.

Both of these two possible methods, i.e. the standard method of solving Equation. 15.63 by means of the Newton-Raphson method, as well as the more elegant method through Equations 15.53 and 15.57 with α as in Equation 15.65, are demonstrated in detail in the Excel workbook TermStructureModels.xls.

Lognormal Models with r-Independent Volatility Structure

Again, for a time (but not interest rate) dependent volatility structure, Equation 15.59 holds. Thus, Equation 15.56 reduces to

$$\begin{aligned} B(i,j) &= B(i-1,j+1)^{\alpha(0,n-1)} \\ &= B(0,n)^{\alpha(0,n-1)^i} \quad \text{with} \;\; i+j=n \end{aligned}$$

where $\alpha(i,j) = \exp\left\{2\sigma(t+n\delta t)\sqrt{\delta t}\right\}$. Equation 15.61 implies that the short rate is simply to be multiplied by a constant factor $\alpha(0,n-1)$ when moving from one node to the next in time slice n of the tree, thus

$$\begin{aligned} r(i,j) &= r(i-1,j+1)\alpha(0,n-1) \\ &= r(0,n)\alpha(0,n-1)^i \end{aligned} \tag{15.66}$$

and the arbitrage condition for the discount factor at the lowest node of the time slice n reduces to

$$B(t, t + (n + 1)\delta t) = \sum_{i=0}^{n} G(i, n - i)B(0, n)^{\alpha(0,n-1)^i}$$

15.9 The Example Program TermStructureModels.xls

15.9.1 Construction of Interest Rate Trees and Option Pricing

We now demonstrate explicitly how the interest rate tree is constructed for a given term structure, or equivalently, for given market prices of zero bonds,

$$B(t, t + i\delta t) \quad \text{for} \quad i = 1 \ldots n$$

and given volatilities

$$b(r, t + i\delta t) \quad \text{for} \quad i = 0 \ldots n - 1$$

The calculation of option prices from this interest rate tree – once it has been constructed – has already been discussed in Section 15.7.2 in detail for any arbitrary arbitrage-free short rate term structure model using caplets as an example.

- Time step $i = 0$
  According to Equation 15.43, or in particular according to Equation 15.37, the Green's function at the node $(0, 0)$ is simply

  $$G(0, 0) = 1$$

  From Equation 15.42, it then follows that the discount factor at node $(0, 0)$ can be obtained directly from the market price $B(t, t + 1\delta t)$ of the zero bond:

  $$B(t, t + 1\delta t) = G(0, 0)B(0, 0) = B(0, 0)$$

- Time step $i = 1$
  The discount factors just computed together with the Green's function evaluated at the node $(0, 0)$ are substituted into the recursion relation Equation 15.39 to determine the values of the Green's function at the next nodes:

  $$G(1, 0) = \frac{1}{2}G(0, 0)B(0, 0)$$
  $$G(0, 1) = \frac{1}{2}G(0, 0)B(0, 0)$$

  These values, the volatility information from the previous time slice and the price of the zero bond maturing at the next time slice are all used to obtain the discount

factor at the lower boundary which here is the node $(0,1)$. With Equation 15.57 for the normal model, the discount factor at this node is obtained as

$$B(0,1) = \frac{B(t,t+2\delta t)}{G(0,1) + G(1,0)\alpha(0,0)}$$

Application of the recursion relation in Equation 15.56 yields the other discount factor required for this step

$$B(1,0) = \alpha(0,0)B(0,1)$$

Analogously, the arbitrage condition for the discount factor on the lower boundary in the *log*normal model is, according to Equation 15.63,

$$B(t,t+2\delta t) = G(0,1)B(0,1) + G(1,0)B(0,1)^{\alpha(0,0)}$$

which can only be solved numerically (using the Newton-Raphson method, for example) for $B(0,1)$. Having solved the equation, the next discount factor for this time slice can be obtained immediately using Equation 15.62:

$$B(1,0) = B(0,1)^{\alpha(0,0)}$$

- Time step $i = 2$
  The discount factors just calculated together with the Green's function are substituted into the recursion 15.39 to obtain the values of the Green's function evaluated at the nodes corresponding to this time step

$$\begin{aligned} G(1,1) &= \frac{1}{2}G(1,0)B(1,0) + \frac{1}{2}G(0,1)G(0,1) \\ G(2,0) &= \frac{1}{2}G(1,0)B(1,0) \\ G(0,2) &= \frac{1}{2}G(0,1)B(0,1) \end{aligned}$$

  Using these values as well as the volatility information from the previous time slice and the market price of the zero bond maturing at the next time slice, Equation 15.57 for the normal model can be applied to obtain the discount factor at the lower boundary, i.e. at node $(0,2)$

$$\begin{aligned} B(t,t+3\delta t) &= B(0,2)\sum_{i=0}^{2} G(i,2-i)\prod_{k=1}^{i}\alpha(i-k,1-i+k) \\ &= B(0,2)G(0,2) \\ &+ B(0,2)G(1,1)\alpha(0,1) \\ &+ B(0,2)G(2,0)\alpha(1,0)\alpha(0,1) \end{aligned}$$

which yields for $B(0,2)$

$$B(0,2) = \frac{B(t,t+3\delta t)}{G(0,2) + G(1,1)\alpha(0,1) + G(2,0)\alpha(1,0)\alpha(0,1)}$$

Applying the recursion formula 15.56 yields the next discount factors belonging to this time step

$$\begin{aligned} B(1,1) &= \alpha(0,1)B(0,2) \\ B(2,0) &= \alpha(1,0)B(1,1) \end{aligned}$$

Analogously for the *log*normal model, the arbitrage condition Equation 15.63 for the discount factors on the lower boundary is used to obtain

$$B(t,t+3\delta t) = G(0,2)B(0,2) + G(1,1)B(0,2)^{\alpha(0,1)} + G(2,0)B(0,2)^{\alpha(1,0)\alpha(0,1)}$$

where again we can be solve for $B(0,2)$ numerically. Once $B(0,2)$ is known, the additional discount factors for this time step can be computed immediately using Equation 15.62:

$$\begin{aligned} B(1,1) &= B(0,2)^{\alpha(0,1)} \\ B(2,0) &= B(1,1)^{\alpha(1,0)} \end{aligned}$$

This procedure is repeated until the entire tree has been constructed up to maturity $T + \tau$ (maturity of the derivative to be priced plus the lifetime of the underlying). After this has been done, all required Arrow-Debreu prices can be determined using Equations 15.43, 15.44 and 15.45 and finally, these ADPs are used in pricing the derivative as demonstrated in Section 15.7.2 for caplets.

This is all demonstrated explicitly for normal models (referred to there as Ho-Lee models) and lognormal models (referred to as Black-Derman-Toy models) in the Excel workbook TermStructureModels.xls included on the CD-ROM accompanying this book. To emphasize that the interest rate tree is independent of the derivative being priced, the structure of the Visual Basic code is extremely modular: first, a short rate tree is generated. This remains in the main memory of the computer until it is used to compute the value of any desired underlying dependent on the term structure (for example a 3-month zero bond yield or a swap rate, etc.) using Arrow-Debreu prices and subsequently pricing any chosen (path-independent) derivative on that underlying with a European payoff mode. Plain vanilla caplets and floorlets on the 3-month rate serve as examples, with an explicit demonstration of their valuation being included in the workbook.

15.9.2 Absolute and Relative Volatilities

As already mentioned, the volatility input for lognormal models has to have the form of a *relative* volatility (for example, 14% of the current underlying value). For normal

models, an *absolute* volatility (for example, 0.75 percentage *points*[14]) is required. If we still wish to use a specified *relative* volatility in a *normal* model, such a relative volatility input must first be multiplied by an underlying value, i.e. an interest rate, to make it absolute. We can now ask: which interest rate should we use for this multiplication? The instantaneous short rate? The underlying (for example, the 3-month rate)? At which time? As of today? It makes most sense to use the *forward rate of the underlying at the maturity of the derivative being priced* as the multiplicative factor in moving back and forth between absolute and relative volatilities. This is especially true if we intend to compare the volatilities with the Black-76 volatility, since it is exactly this forward rate which is used as the underlying in the Black-76 model[15]. Therefore the Black-76 volatility quoted in the market belongs to this forward rate. For these reasons, this forward rate has been selected as the multiplicative factor in the Excel workbook TermStructureModels.xls.

Note that if a lognormal model is transformed into a normal model using the Ito formula as was done for Equation 15.53, the relative volatility according to Equation 15.65 on each arbitrary node (i, j) always appears in a product together with the short rate associated with this node, i.e. together with $r(i, j)$. Otherwise, everything is exactly the same as in a normal model. The product of the relative volatility and the short rate can be interpreted as an *absolute* volatility. Or in other words: in the lognormal model, the conversion factor between the relative and the absolute volatility is the short rate $r(i, j)$ at each node (i, j); in contrast to the normal model, where this conversion factor is a constant (for example, the forward rate of the underlying). In fact, the difference in the implementation of a normal and lognormal model consists of just one single line when the lognormal model is constructed according to Equation 15.53.

Both approaches, the standard approach of numerically solving Equation 15.63 by means of Newton Raphson, as well as the approach using Equations 15.53 and 15.57 with α as given in 15.65, are presented in the workbook TermStructureModels.xls.

15.9.3 Calibration of Volatilities

The current term structure and volatility structure are required as input for the construction of an interest rate tree. In practice, the volatilities $b(r, t + i\delta t)$ themselves are generally not known. Only the prices of options traded on the market (caps, floors, swaptions, etc.) can be observed directly. The tree must then be constructed, leaving the volatilities unspecified as free parameters which are then adjusted until the observed market prices of the options are reproduced by the model. Fitting the parameters $b(r, t + i\delta t)$ to the market prices in this manner is referred to as the *calibration* of the volatility. There are many ways of performing such a calibration. For instance

[14] At a current interest rate of 6% this would correspond to a relative volatility of 14%.

[15] As emphasized in Equation 15.50, the forward rate for the caplet period τ within *linear* compounding serves as the underlying of a cap in the payoff profile. This has already been accounted for in the term structure model through Equation 15.50. To obtain the correct forward rate as input for the *Black-76* model in the Excel workbook TermStructureModels.xls from the current term structure (which holds for continuous compounding), we first determine the forward *discount factors* using Equation 3.4. From those discount factors the desired forward rates for linear compounding are given by $r = \left(B^{-1} - 1\right)/\tau$.

we can – exactly as was done when reproducing the market prices of zero bonds – reproduce the market prices of the options stepwise through the tree, beginning with the shortest option lifetimes and proceeding through to the longest. This procedure is demonstrated explicitly in the Excel workbook TermStructureModels.xls where the prices of a strip of caplets at the 3-month rate are available, whose underlyings (the respective 3-month rates) cover the time span under consideration without overlap, i.e.

$$T_{k+1} = T_k + \tau$$

with T_k is the maturity of the k^{th} caplet and τ is the lifetime of the underlying. The calibration starts by assuming that the volatility in the tree between times t and $T_1 + \tau$ is constant,

$$b(r, t + i\delta t) = b(T_1) \quad \text{for all } r \text{ and all } i \text{ with } t \leq t + i\delta t \leq T_1 + \tau$$

This constant volatility is adjusted (using Newton-Raphson) until the price computed using the tree equals the known market price of the first caplet.

To reproduce the second known caplet price, we assume that the volatility remains equal to the (just calibrated) $b(T_1)$ for the time between t and $T_1 + \tau$, and is constant (although still unknown) at all nodes for times between $T_1 + \tau$ and $T_2 + \tau$:

$$b(r, t + i\delta t) = b(T_2) \quad \text{for all } r \text{ and all } i \text{ with } T_1 + \tau < t + i\delta t \leq T_2 + \tau$$

This constant volatility is then adjusted (again using Newton-Raphson) so that the price for the second caplet computed using the tree is equal to the observed market price of this caplet.

To reproduce the third caplet price we assume, as above, that the (just calibrated) volatilities computed for the time span from t to $T_2 + \tau$ continue to be valid, and adjust the volatility at all nodes associated to the times between $T_2 + \tau$ and $T_3 + \tau$ until the price of the third caplet computed using the tree matches the observed market price of this caplet, and so on[16]. By this method, we obtain a piecewise constant function for the volatility[17] as a function of time, independent of r.

Such a calibration process yields different volatility values $b(r, t + i\delta t)$ for each different term structure model. We therefore refer to the respective volatilities by the name of the model with we are working, for example *Ho-Lee volatilities*, *Black-Derman-Toy volatilities*, *Hull-White volatilities*, etc. For this reason, it is generally not possible to simply take the Black-76 volatilities as the input values for $b(r, t + i\delta t)$. The tiresome process of calibrating to the observed option *prices* is in most cases unavoidable.

Option prices on the market are usually quoted in the form of Black-76 volatilities rather than in a monetary unit. Since all other input parameters are commonly known, the Black-76 model is simply a translation algorithm for moving between the two different ways of quoting the option price. It says very little about what the bank

[16] Alternatively, the volatilities can be (simultaneously) calibrated using a least squares fit. We then minimize the sum of the quadratic differences between the calculated and the traded option prices by varying the volatilities.

[17] Naturally, this procedure can be extended if a "caplet-price-surface", i.e. caplet prices with different times to maturity *and* different strikes, is available to obtain a calibrated volatility surface as a function of time *and moneyness* (the relative difference between underlying and strike).

quoting a particular option really "believes" is the true volatility. The procedure behind the quoting is as follows: the bank calculates the price of an interest rate option which it wants to quote on the market using the term structure model of its choice (calibrated to liquid option prices). But before quoting, this price is first translated into a volatility using the Black-76 model[18]. It is this volatility that is quoted on the market. The volatility quoted is simply that which, when used as an input parameter in the Black-76 model, reproduces the price calculated with the bank's (perhaps very complicated and proprietary) term structure model.

The calibration to given (Black-76) caplet (or floorlet) prices are demonstrated in the Excel Workbook TermStructureModels.xls included on the CD-ROM accompanying this book. The calibration to other instruments such as swaption prices is considerably more complicated in its implementation, but is based on the same principles:

1. Calculation of all required Arrow-Debreu prices from the existing tree.
2. Generation of the necessary underlyings, namely the swap rate under consideration, from the Arrow-Debreu prices.
3. Calculation of the payoff profiles of the swaption at the nodes corresponding to the swaption maturity date based on the value of the underlying (swap rate) and with the help of the Arrow-Debreu prices
4. Calculation of the swaption prices at node $(0, 0)$ by discounting the payoff profile with the Green's function
5. Adjusting the volatility in the interest rate tree until the swaption price computed using the tree agrees with the price quoted on the market.

15.10 Monte Carlo on the Tree

In the discussion above, it was emphasized that only path-independent derivatives in the sense of Equation 15.40 could be priced using the methods introduced in this chapter since the trees recombine and as a result information on the history of the short rate path is lost.

However, by performing a *Monte-Carlo simulation on the tree*, the trees introduced here can still be used in pricing even such path-dependent derivatives for which the historical progression of the interest rate influences the payoff profile. This kind of Monte-Carlo simulation works as follows:

- First observe that the tree needs to be generated (and calibrated!) only once (as described above) and then stays in the memory of the computer.
- The short rate paths are then simulated by jumping randomly from node to node in the tree, always proceeding one time step further with each jump.

[18]This of course is only possible if the option concerned is one that can be priced using the Black-76 model, for example a plain vanilla option with a strike price not available on the market.

- Simulating the jumps from node to node with the appropriate transition probabilities (in our case $p = 1/2$ for up as well as for down moves, see Equation 15.28) of the chosen martingale measure (in our case the risk-neutral measure, see Equation 15.27) ensures that the simulated paths already have the correct probability weight needed for pricing financial instruments. This procedure is called *importance sampling*[19].

- For each simulated path of the short rate the corresponding path of the underlying (for instance 3-month LIBOR) must be calculated using Arrow-Debreu prices.

- At the end of each simulated path the payoff of the derivative resulting from the underlying having taken this path is calculated.

- This payoff is then discounted back to the current time t (i.e. to node $(0, 0)$) at the short rates *along the simulated path*, since after all, we still are within the risk-neutral measure, see Equation 15.27.

- After many (usually several thousand) paths have been simulated, the (several thousand) generated discounted payoff values can then simply be averaged to yield an estimation for the risk-neutral expectation of the discounted payoff. Here the very simple arithmetic average (with equal weights) of the discounted payoff values can be used without worrying about the correct probability weight of each payoff value since the payoff values (more precisely the paths which generated the payoff values) have already been *simulated* with the correct (risk-neutral) probability. This is (besides the effectiveness in sampling the phase space) another great advantage of importance sampling.

- According to Equation 15.8, the risk-neutral expectation calculated in this way is directly the desired derivative price.

15.11 The Drift in Term Structure Models

At this stage, it makes sense to reconcile the practical computations in Chapter 15 with the profound concepts from the related theory presented in Chapter 14. We first consider the relationship between the underlying instantaneous interest rates and tradable instruments such as zero bonds. We defer proceeding to the short rate models introduced above to take a brief excursion to the models in which the instantaneous *forward* rate plays the role of the underlying.

[19] Doing a move only with its associated probability ensures a very effective sampling of the *phase space* (the set of all possible values of the simulated variables): phase space regions (values of the simulated variables) which have low probabilities (and therefore contribute only little to the desired averages of whatever needs to be measured by the simulation) are only visited with low probability (i.e. rarely) while phase space regions with high probabilities (which contribute a lot to the desired averages) are visited with high probability (i.e. often). Because of this feature *importance sampling* is heavily used in thousands of Monte-Carlo applications, especially in physics, meteorology and other sciences which rely on large-scale simulations.

15.11.1 Heath-Jarrow-Morton Models

We see from the fundamental Equations 15.2 or 15.3 and 15.5 that the instantaneous *forward* rates can be used to generate the prices $B(t,T)$ of all zero bonds as well as the entire interest rate term structure $R(t,T)$. All three descriptions of the term structure, the zero bond prices $B(t,T)$, the zero bond yields $R(t,T)$ and the instantaneous forward rates $f(t,T)$ are equivalent. Of these three variables, only a single one needs to be chosen to be modeled by a general stochastic process of the form specified in Equation 3.16. We take, for example, the forward rates $f(t,T)$ to be modeled by a process of the form:

$$df(t,T) = a(t,T)\,dt + b(t,T)\,dW \qquad \text{with} \quad dW = X\sqrt{dt},\ X \sim \mathrm{N}(0,1) \tag{15.68}$$

As mentioned previously, all bond prices and thus the entire term structure can be generated from the solution of this equation. An entire class of term structure models, the *Heath-Jarrow-Morton models* (*HJM models* for short) take this approach of employing the forward rates as the driving factor of the term structure.

Like all interest rates, the instantaneous forward rates are not tradable (see Section 15.4). In Chapter 14, and in particular in Sections 14.4 and 14.5, a detailed discussion can be found on how to proceed when the underlying is not tradable; we choose a tradable instrument U whose price $U(S,t)$ is a *function* of the underlying. Then all of the results shown in Chapter 14 hold:

- The Harrison-Pliska Theorem establishes the *uniqueness* (in complete markets) of the probability measure with respect to which the prices of tradable financial instruments normalized with an arbitrarily chosen, tradable numeraire instrument Y are martingales.

- According to the Girsanov Theorem, this implies that there is only *one single underlying drift* which may be used for pricing.

- In addition, the drift of the underlying in the real world plays absolutely no role in the world governed by the martingale measure if the numeraire instrument Y satisfies the property 14.33 (which is always the case).

The existence of a function $U(S,t)$ relating the underlying S to a tradable instrument U is the deciding factor for the validity of the above results. If the underlying is the instantaneous forward rate $f(t,T)$, such a functional relation to a tradable instrument exists, namely Equation 15.3 for the zero bond price $B(t,T)$, and thus the results found in Chapter 14 hold for the Heath-Jarrow-Morton models. In particular, for every complete market, there exists for each numeraire instrument exactly one single underlying drift which may be used for pricing. Therefore, as far as pricing is concerned, the HJM model is uniquely determined by the specification of the volatility term $b(t,T)$ in Equation 15.68 (which models the process in the real world). The Girsanov Theorem implies that the transition from the real world into the world governed by the martingale measure only effects a (in this case unique) change in the drift; the volatility term $b(t,T)$ is invariant under this transformation. Indeed, the forward rate process in the

risk-neutral measure corresponding to the real world process in Equation 15.68 has the following appearance [78]:

$$df(t,T) = \left[b(t,T) \int_t^T b(t,s)ds \right] dt + b(t,T)\, d\widetilde{W}$$

Here $d\widetilde{W}$ denotes the standard Brownian motion with respect to the martingale measure. The coefficient of dt appearing in square brackets is the drift with respect to the martingale measure. This formula shows *explicitly* that for HJM models, the entire model (including the drift) is uniquely specified through the volatility $b(t,T)$. The drift to be used in the valuation is *unique,* in complete agreement with the general statements made in Chapter 14.

15.11.2 Short Rate Models

As opposed to the HJM models, the term structure models in Section 15.6 make use of the instantaneous *spot* rate defined in Equation 15.1 as the driving factor. However, a one-to-one mapping between the spot rates and the zero bond prices does *not* exist and in consequence, *no* one-to-one mapping between the spot rates and the term structure $R(t,T)$ can exist either. The instantaneous spot rates are not sufficient to generate the term structure. This is indicated by the fact that the instantaneous spot rate $r(t)$ is a function of a single time variable t in contrast to the processes $B(t,T)$, $R(t,T)$ (and $f(t,T)$ as well!), which are functions of *two* time variables. Taking the limit $dt \to 0$ in the definition of the instantaneous short rate in Equation 15.1 results in the loss of the second argument (and thus in the loss of the corresponding information). This can be seen explicitly in Equation 15.6. For this reason, there is *no* analogy to Equation 15.3 relating the instantaneous spot rates directly to the zero bond prices. The best possible alternative available is to determine the bond prices from the *expectations* of the short rates (see for example Equation 15.9), but not directly as a *function* of the short rates. This has significant consequences:

The results presented in Chapter 14, in particular those in Sections 14.4 and 14.5 (regarding the martingale measure, unique drift, etc.), can be shown for non-tradable underlyings only if there exists a tradable instrument whose price process is a *function* of the underlying. This *direct functional relationship* between the underlying (the instantaneous spot rate) and a tradable instrument is *missing* in short rate models. Or from the view point of the Harrison-Pliska Theorem: since $r(t)$ contains less information than $f(t,T)$ or $B(t,T)$, the market is not complete for short rate models. For this reason, the martingale measure in short rate models is *not uniquely determined* by fixing the numeraire instrument. Thus, the Girsanov theorem asserts that we retain the freedom of choosing from various drift terms in the model. The information lost in the transition shown in Equation 15.6, for example, must be reinserted into the model "by hand". This is accomplished by directly specifying a drift in the world governed by the *martingale measure.* This is the essential difference in the models here compared to those encountered in the previous chapters, where the drift was always specified in the *real* world. The situation for short rate models is different; the drift is specifically chosen for the world governed by the *martingale measure* rather than for the real world.

Only through this drift specification is the martingale measure uniquely determined in short rate models.

A popular motivation for selecting a drift is an effect called *mean reversion* which is observed in the evolution of interest rates but not seen in e.g. stock prices. Interest rates do not rise or fall to arbitrarily high or low levels but tend to oscillate back and forth about a long-term mean. This can be modeled with a drift in the functional form $\mu - vr$ for some $v > 0$: for values of r small enough so that $vr < \mu$ holds, the drift is positive and, consequently, r tends on average toward larger values. Conversely, for values of r large enough so that $vr > \mu$ holds, the drift is negative and r tends to drift toward smaller values on average. The interest rate thus tends to drift toward a mean value μ at a rate ν. Naturally, a stochastic component driven by $\sim dW$ is superimposed onto this deterministic movement. An example of a *mean reversion model* is the *Hull-White model*[90]. This model specifies the following stochastic process for the short rate with respect to the *risk-neutral martingale measure:*

$$dr(t) = [\mu(t) - v(t)r]\ dt + \sigma(t)\, dW \tag{15.69}$$

This is only one of many examples. Several well-known models, each named after their respective "inventors" can be distinguished from one another, after having separated them into categories of normal and lognormal models, essentially through the form of their drift. The best known representatives of these models are summarized in the following list:

- Normal models $dr(t) = a\,(r,t)dt + b(r,t)\, dW$
  - Stationary models $b(r,t) = \sigma$
    - Arbitrage-free models
      - Hull-White $a\,(r,t) = \mu(t) - vr$ (mean reverting)
      - Ho-Lee $a\,(r,t) = \mu(t)$
    - Equilibrium models (not arbitrage free because of too few degrees of freedom)
      - Vasiceck $a\,(r,t) = \mu - vr$ (mean reverting)
      - Rendleman-Barter $a\,(r,t) = \mu$
  - Non-stationary models $b(r,t) = \sigma(t)$
- Lognormal models $d\ln r(t) = a\,(r,t)dt + b(r,t)\, dW$
  - Stationary models $b(r,t) = \sigma$
  - Non-stationary models $b(r,t) = \sigma(t)$
    - Arbitrage-free models
      - Black-Karasinski $a\,(r,t) = \mu(t) - v(t)\ln r$ (mean reverting)
      - Black-Derman-Toy $a\,(r,t) = \mu(t) - \frac{\partial\sigma(t)/\partial t}{\sigma(t)} \ln r$

The models are called either stationary or non-stationary depending on whether or not the volatility is assumed to be a function of time[20]. The models allowing for neither a time dependence nor an r dependence in the drift or volatility terms (Vasiceck, Rendleman-Barter) obviously cannot reproduce the current term structure arbitrage-free. For these models, only a "best fit" can be obtained (for example by minimizing the *root mean square error*). These models are referred to as *equilibrium models*. Since equilibrium models are not arbitrage free, they will receive no further attention in this book. Nowadays they are no longer very important in practice.

The volatility of dr is independent of r in normal models and proportional to r in lognormal models (see, Equation 15.53). An intermediate scheme between these two possibilities is the *Cox-Ingersoll-Ross* model [35] for which the volatility is assumed to be proportional to $\sqrt{r}$.

$$dr = (\mu - vr)dt + \sigma\sqrt{r}dW$$

Note that all these processes are modeled in the *risk-neutral martingale measure* directly, i.e. should be used directly in the valuation of the financial instrument under consideration without first performing a drift transformation through an application of the Girsanov Theorem. The volatility term is invariant under the Girsanov transformation. This implies that the volatility taken for the valuation is the same as that observed in the real world. The form taken on by the drift is, however, a result of the particular choice of the measure (coordinate system) established for pricing through an application of the Girsanov Theorem and as such, a rather artificial construct. It is thus not readily apparent why a specific form of drift (for example mean reversion) should be modeled in a specific (for example risk-neutral) artificial world (dependent on the selection of a particular numeraire) when our intuitive conception of the drift actually pertains to the *real* world.

Or more precisely: according to the Girsanov Theorem, the process modeled with respect to the martingale measure differs from the real world process by a previsible process $\gamma(r,t)$. This previsible process is *arbitrary* (with the restriction that it must satisfy the boundedness condition $\mathrm{E}\left[\exp\left(\frac{1}{2}\int_0^T \gamma(r,t)dt\right)\right] < \infty$). Therefore the choice of model with respect to a martingale measure provides as good as no information about the drift of the short rate in the *real* world. For example, the model given by Equation 15.69 has a mean reversion in the world governed by the martingale measure, but has the form

$$dr(t) = [\mu(t) - v(t)r(t) + \gamma(r,t)]\ dt + \sigma(t)\, dW$$

in the real world with a (practically) *arbitrary* previsible process $\gamma(r,t)$. It is therefore by no means clear that this process shows any mean reversion in the real world.

After having placed such emphasis on the necessity of specifying a drift, i.e. a martingale measure, explicitly in short rate models (despite the results in Chapter 14 the martingale measure is not unique here, even when the numeraire has been fixed), the attentive reader will surely have asked why only the volatility but no drift information has entered as input into our explicit computations in Section 15.8. For

[20] Ho-Lee and Hull-White are often applied for time-dependent volatilities. Their inventors, however, originally assumed constant volatilities.

both the normal and the lognormal models, the interest rate trees were constructed in their entirety and no drift information from the specific stochastic processes was needed at any point.

This stems from the condition in Equation 15.28, which we introduced "by hand" for the sake of simplicity; the probability of an up move was simply *set* to $p = 1/2$. Through this choice, we have explicitly selected one particular measure from the family of arbitrage-free martingale measures belonging to the risk-neutral numeraire instrument (the bank account). This fixed the drift in accordance with the Girsanov Theorem. We can also see this fact explicitly since the drift can be quite simply determined from the generated tree. As was shown in Section 15.8 for the variance, it follows from the general equation 15.54 for a binomial tree with $p = 1/2$ that the expectations of the short rate as seen from the node (i, j) are

$$a(i,j)\delta t = \mathrm{E}\left[r(i,j)\right] = \frac{1}{2}\left[r(i+1,j) + r(i,j+1)\right] \qquad \text{Normal Model}$$

$$a(i,j)\delta t = \mathrm{E}\left[\ln r(i,j)\right] = \frac{1}{2}\left[\ln r(i+1,j) + \ln r(i,j+1)\right] \quad \text{Lognormal Model}$$

from which the drift at each node in the tree can be immediately determined since all the short rates have already been established (the tree has already been built).

For example, Equation 15.60 can be used to compute the drift in a normal model with r-independent volatility as

$$a(i,j)_{\text{normal}} = \frac{r(0,i+j+1)}{\delta t} + (2i+1)\frac{\sigma(t+(i+j)\delta t)}{\sqrt{\delta t}}$$

Analogously, the drift for *log*normal models with r-independent volatility can be computed explicitly, employing Equation 15.66 and α as in Equation 15.62:

$$\begin{aligned} a(i,j)_{\text{lognormal}} &= \frac{1}{2\delta t}\left[\ln r(i+1,j) + \ln r(i,j+1)\right] \\ &= \frac{1}{\delta t}\left[\ln r(0,i+j+1) + (i+\frac{1}{2})\ln \alpha(0,i+j)\right] \\ &= \frac{\ln r(0,i+j+1)}{\delta t} + (2i+1)\frac{\sigma(t+(i+j)\delta t)}{\sqrt{\delta t}} \end{aligned}$$

Had the probability of an up move not been fixed at 1/2, a free parameter p would remain unspecified in Equation 15.54 which (if dependent on time and perhaps on r as well) would allow for many different drift functions.

15.12 Short Rate Models with Discrete Compounding

The discount factor over a single time period used in this chapter was always of the form $B(t, t+\delta t) = e^{-r(t)\delta t}$ (see for example Equation 15.32). Intuitively, in view of Equation 3.2, this means that interest has been paid infinitely often in the reference

period δt, and that these payments were then immediately reinvested at the same rate. Strictly speaking, this contradicts the concept of a tree model, for which time has been *discretized* into intervals of positive length δt, implying by definition that nothing can happen in between these times. To be consistent, we should have therefore used *discrete* compounding, allowing the payment and immediate reinvestment of interest solely after each δt. Then, only in the limiting case $\delta t \to 0$ will the discount factor for continuous compounding be obtained. If we wish to be consistent, we would therefore have to write

$$B(t, t+\delta t) = \frac{1}{1 + r(t)\delta t} \underset{\delta t \to 0}{\longrightarrow} e^{-r(t)\delta t}$$

Despite the inconsistency, the discount factor for continuous compounding is commonly used in the literature. In *this* section, we will collect and present the differences caused by using discrete rather than continuous compounding and, in doing so, show how short rate models with discrete compounding can be treated. The discount factor in Equation 15.32 has the following form when adopting the convention of discrete compounding

$$B(i,j) = \frac{1}{1 + r(i,j)\delta t} \tag{15.70}$$

Formulating recursion equations as in Equations 15.56 or 15.62 for these discount factors $B(i,j)$ in a time slice or for the lowest zero bond in a time slice (see Equations 15.57 and 15.63) is quite awkward. Such conditions are more easily formulated for the *interest rate* $r(i,j)$. All recursion relations for the short rate follow from the arbitrage condition for the Green's function, Equation 15.42. This expression for the instantaneous discount factor in the form of Equation 15.70 is

$$B(t, t+n\delta t) = \sum_{i=0}^{n-1} \frac{G(i, n-i-1)}{1 + r(i, n-i-1)\delta t} \tag{15.71}$$

15.12.1 Normal Models

If the short rate $r(i,j)$ in Equation 15.70 is governed by a stochastic process of the form given in Equation 15.51, then Equation 15.55 holds. The recursion relation for the short rate in the normal model is given by Equation 15.58:

$$\begin{aligned} r(i,j) &= r(i-1, j+1) + 2b(i-1,j)\sqrt{\delta t} \\ &= r(0, j+i) + 2\sqrt{\delta t} \sum_{k=1}^{i} b(i-k, j+k-1) \end{aligned} \tag{15.72}$$

Substituting this into Equation 15.71 with $j = n-i-1$ and performing the transformation $n \to n+1$, we obtain a condition analogous to Equation 15.57 for the interest rate $r(0,n)$ at the lowest node in the time slice n

$$B(t, t+(n+1)\delta t) = \sum_{i=0}^{n} \frac{G(i, n-i)}{1 + r(0,n)\delta t + 2\delta t^{3/2} \sum_{k=1}^{i} b(i-k, n-i+k-1)}$$

This equation can only be solved numerically for $r(0,n)$. Once this value is known, Equation 15.72 provides all other $r(i,j)$ on the time slice n. From the $r(i,j)$, the discount factors can then be calculated immediately using Equation 15.70. Note that in the case of discrete compounding, the arbitrage condition, Equation 15.71, can no longer be solved analytically, not even in the context of the normal model.

15.12.2 Lognormal Models

If the short rate $r(i,j)$ in Equation 15.70 is governed by a stochastic process of the form 15.52, then Equation 15.61 holds. The recursion relation for the short rate in the lognormal model is given by Equation 15.64:

$$\begin{aligned} r(i+1,j) &= r(i,j+1)\exp\left\{2b(i,j)\sqrt{\delta t}\right\} \\ &= r(0,j+i)\prod_{k=1}^{i}\alpha(i-k,j+k-1) \end{aligned} \tag{15.73}$$

with $\alpha(i,j)$ as defined in Equation 15.62. Substituting this into Equation 15.71 with $j = n-i-1$, and performing the transformation $n \to n+1$, we obtain the condition on the interest rate $r(0,n)$ at the lowest node in the time slice n analogous to Equation 15.63

$$B(t,t+(n+1)\delta t) = \sum_{i=0}^{n}\frac{G(i,n-i)}{1+r(0,n)\delta t\prod_{k=1}^{i}\alpha(i-k,n-i+k-1)}$$

Again, this can only be solved for $r(0,n)$ numerically. Once this has been done, the other values $r(i,j)$ in the time slice n can be calculated immediately using Equation 15.73. The discount factors are finally determined from the $r(i,j)$ using Equation 15.70.

Part III

Instruments

16

Spot Transactions on Interest Rates

In Part II, a whole array of very workable methods for the valuation and hedging of financial instruments was introduced. We now continue in Part III with the explicit valuation of the most important and common financial instruments. We restrict our considerations to simple (for the most part, plain vanilla) instruments which still represent the largest proportion of all trades in financial markets today. The methods presented in Part II do in fact allow much more complicated instruments than those introduced here to be effectively priced. Seen from this point of view, the application of the material introduced in Part II is much more extensive than its restriction to the instruments defined in Part III might suggest. More involved applications (such as term structure models for Bermudan swaptions or multi-dimensional finite difference schemes for convertible bonds, etc.) often contain so many details specific to the individual implementation, that it is easy to become distracted from the essential ideas. Therefore such complicated examples are not particularly appropriate for discussion in an introductory text. However, Part II enables the reader to develop pricing techniques for quite complex products even if they don't receive specific treatment in Part III.

In the following sections, the valuation of different plain vanilla instruments will often be explicitly demonstrated on the basis of tables. All of these tables have been extracted from the Excel workbook PlainVanilla.xls on the CD-ROM accompanying this text. The computations contained in this workbook are carried out solely using cell calculations. VBA-programs will not be needed.

16.1 Zero Bonds

16.1.1 Cash Flows and Present Value

A *zero bond* is a security with a single cash flow: the *redemption* at maturity, i.e. payment of the nominal N at maturity. The cash flow C_T resulting from this payment

	Present Value	Mod. Duration = Key Rate Dur.	Convexity
General	$B_R(t,T)$	$\frac{-1}{B_R(t,T)} \frac{\partial B_R(t,T)}{\partial R}$	$\frac{1}{B_R(t,T)} \frac{\partial^2 B_R(t,T)}{\partial^2 R}$
Continuous	$e^{-R\tau}$	τ	τ^2
Discrete	$\frac{1}{(1+R)^\tau}$	$\frac{\tau}{1+R}$	$\frac{\tau(\tau+1)}{(1+R)^2}$
Simple	$\frac{1}{1+R\tau}$	$\frac{\tau}{1+R\tau}$	$\frac{\tau^2}{(1+R\tau)^2}$
Linear	$1 - R\tau$	$\frac{\tau}{1-R\tau}$	0

Table 16.1: Present value and sensitivities of a zero bond in the four most commonly used compounding methods. For the sake of simplicity, the abbreviation $\tau = T - t$ is used to represent the time from the value date t until maturity T.

is:

$$C_T(T) = N$$

The present value, defined as the sum of all discounted future cash flows, is accordingly

$$V_R(t,T) = NB_R(t,T) \tag{16.1}$$

16.1.2 Yield To Maturity and Par Rate

A zero bond does not pay coupons at a fixed rate K. At best we could say the coupon rate is $K = 0$ (hence the name). The yield of a zero bond results from the difference between its present value, Equation 16.1, and its nominal amount N and is thus equal to the current spot rate $R(t,T)$ over the lifetime of the zero bond. As explained is Section 6.3, a "par rate" only makes sense in the context of fixed-rate instruments and thus will not be defined for zero bonds. Since there is only one cash flow, the yield to maturity (YTM) is equal to the current spot rate until maturity T, which in turn is the only relevant key rate.

16.1.3 Sensitivities

Zero bonds are the basic elements which can be used to replicate all securities whose value derives from future cash flows. This is because the present value of any arbitrary future cash flow is nothing other than the value of a zero bond with a principal equal to that cash flow and maturity equal to the payment date of that cash flow. Because of the immense importance of the zero bond, its present value, modified duration and convexity for all compounding conventions are listed explicitly in Table 16.1.

Since the YTM of a zero bond is equal to the spot rate for its maturity, the modified duration and the key rate duration agree. For continuous compounding, the sensitivities have a particularly simple and intuitive form: the modified duration is exactly equal to the time to maturity and the convexity is the square of the time to maturity.

16.2 Floaters

A *floater* is a security whose cash flows consist of floating interest payments and a final payment at maturity. The variable interest payments refer to a *reference rate* observable on the market (for example, the 6-month *LIBOR*). The cash flows are therefore equal to the interest on the principal N of the floater compounded at the reference rate over the interest period under consideration (for example, 6 months). They are fixed at the beginning of each interest period (*fixing*) at the reference rate valid at that time and are payable at the *end* of the interest period. The future cash flows thus remain unspecified at time t with $t_m \leq t < t_{m+1}$, with the exception of the next payment due at $t_m + 1$ since this amount was established at the time of the last fixing t_m. All additional cash flows are determined at some point in the future. We will now introduce the indexing convention which will be used for the rest of this chapter. The date of the last cash flow *preceding* the present time t will be denoted by t_m, the maturity date T, i.e. the date on which the last cash flow is due, will be denoted by t_n. In summary:

$$\underbrace{t_1 < t_2 < \cdots t_m}_{\text{Past}} \leq t < \underbrace{t_{m+1} < t_{m+2} < \cdots t_n = T}_{\text{Future}}$$

16.2.1 Cash Flows and Present Value

From the same arbitrage arguments used to determine forward rates, it follows that a floater must be priced under the assumption that for each interest period the interest rates realized in the future are equal to today's forward rates. Using the Equation 3.4, the cash flows resulting from the interest payments C_Z and the final payment C_T are then:

$$\begin{aligned} C_Z^{\text{fix}}(t_{m+1}) &= \left[B_R^{\text{fix}}(t_m, t_{m+1})^{-1} - 1\right] N \\ C_Z(t_i) &= \left[B_R(t_{i-1}, t_i|t)^{-1} - 1\right] N = \left[\frac{B_R(t, t_{i-1})}{B_R(t, t_i)} - 1\right] N \quad \forall\, i = m+2,\, \ldots\, , n \\ C_T(t_n) &= N \end{aligned} \tag{16.2}$$

The present value defined as the sum of all discounted future cash flows is then

$$\begin{aligned} V_R(t,T) &= C_Z^{\text{fix}}(t_{m+1}) B_R(t, t_{m+1}) + N \sum_{i=m+2}^{n} B_R(t, t_i) \left[\frac{B_R(t, t_{i-1})}{B_R(t, t_i)} - 1\right] + N B_R(t, t_n) \\ &= C_Z^{\text{fix}}(t_{m+1}) B_R(t, t_{m+1}) + N \sum_{i=m+2}^{n} \left[B_R(t, t_{i-1}) - B_R(t, t_i)\right] + N B_R(t, t_n) \end{aligned}$$

Since the discount factor $B_R(t, t_i)$ appearing with a negative sign in the above formula is equal to the discount factor with a positive sign $B_R(t, t_{i-1},)$ in the next step in the sum, i.e. for $i \to i+1$, the series is telescoping, with all terms canceling except the discount factors with the smallest and the largest index. Thus, the above equation reduces to

$$V_R(t,T) = C_Z^{\text{fix}}(t_{m+1})\, B_R(t, t_{m+1}) + N\left[B_R(t, t_{m+1}) - B_R(t, t_n)\right] + N B_R(t, t_n)$$

Together with Equation 16.2 for the present value of the fixed cash flow this can be reduced to

$$V_R(t,T) = B_R(t,t_{m+1})\left[N + C_Z^{\text{fix}}(t_{m+1})\right] = \frac{B_R(t,t_{m+1})}{B_R^{\text{fix}}(t_m,t_{m+1})}N \tag{16.3}$$

This is quite a remarkable result! The present value of a floater is given by the sum of the principal N and the (already) fixed interest cash flow $C_Z^{\text{fix}}(t_{m+1})$, discounted from the next interest payment date t_{m+1} back to t. This corresponds to a case in which the fixed interest cash flow together with the entire principal is due on the next payment date $t_m + 1$, after which no further payments occur.

A floater is, in terms of pricing and risk management, equivalent to a zero bond with maturity at the next interest payment date and a principal equal to the sum of the floater's nominal and the already fixed interest payment.

Directly after an interest payment, i.e. for $t = t_m$ the next interest cash flow is fixed with respect to the *current* interest rate. This has the effect that ,according to Equation 16.3, the present value of the floater *at a fixing time* is identically equal to its principal:

$$B_R^{\text{fix}}(t_m,t_{m+1}) = B_R(t_m,t_{m+1}) \;\Rightarrow\; V_R(t=t_m,T) = N$$

At no point in the discussion of floaters were any assumptions made on the compounding methods, the length of the interest period or any other such characteristics. The results thus hold for any arbitrary floater independent of compounding convention.

Floaters are typically traded with a *spread*. This means that the cash flows are not fixed precisely at a reference rate rather, but a spread K is added to or subtracted from to the rate (for example, 3-month LIBOR plus 5 basis points). Such a floater (long) generates the same cash flows as a portfolio consisting of

- a floater without a spread, sometimes called *LIBOR flat* (long),
- a coupon bond (long) with the spread as a coupon and having the same principal, coupon dates and maturity as the floater
- and a zero bond (short) whose principal and maturity is the same as that of the floater.

Including the short zero bond is necessary to compensate for the final payment at maturity of the coupon bond. A floater with a spread can thus be priced by means of the *stripping* procedure described above.

16.2.2 Yield to Maturity, Par Rate and Sensitivities

A floater does not earn at a fixed rate K. As indicated by its name, it is a floating rate instrument. A "par rate" as described in Section 6.3 therefore makes no sense in this context. The YTM, however, can be defined for floating rate instruments. Since a floater is equivalent to a zero bond in terms of its market behavior (present value and risk), the YTM and sensitivities, i.e. the modified duration, convexity and the key rate duration are calculated exactly as for the associated zero bond.

16.3 Coupon Bonds

16.3.1 Cash Flows and Present Value

A *coupon bond*, often simply referred to as a *bond*, is a security with a *coupon,* which specifies a fixed interest rate K. The cash flows arising from the interest payments C_Z of a coupon bond is the interest the principal N of the bond, compounded at the rate K, earns over the interest period under consideration (for example, 6 months). In addition, a cash flow C_T for the redemption of the principal occurs at maturity.

$$\begin{aligned} C_Z(t_i) &= \left[B_K(t_{i-1},t_i)^{-1}-1\right] N \quad \forall\ i=1,\ \dots\ ,n \\ C_T(t_n) &= N \end{aligned} \tag{16.4}$$

The present value given as the sum of all discounted future cash flows is

$$V_R(t,T) = N\sum_{i=m+1}^{n} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] + NB_R(t,t_n) \tag{16.5}$$

The coupon periods are typically of identical length. Since the coupon rate is also the same over each interest period, the coupon yield $[1/B_K - 1]$ equals some constant c for all coupon periods. In this case, the present value can be simplified to:

$$V_R(t,T) = N\,c\sum_{i=m+1}^{n} B_R(t,t_i) + NB_R(t,t_n) \quad \text{with} \quad c = B_K(t_{i-1},t_i)^{-1}-1\ \ \forall\, i \tag{16.6}$$

For a coupon period of one year, c is exactly equal to the coupon rate itself for linear, simple and annual compounding conventions:

$$c = B_K(t_{i-1},t_i)^{-1}-1 = \begin{cases} (1+K)^1-1=K & \text{Annual Compounding} \\ 1+1\ K-1=K & \text{Simple and Linear Compounding} \end{cases}, \forall\, i$$

16.3.2 Yield To Maturity

By definition of the yield to maturity, discounting in Equation 16.5 at the current *yield to maturity* $\overline{R}$ instead of the current zero bond rates (i.e. setting $R=\overline{R}$) must result in the same present value. Assuming that the coupon periods are equally long, results in YTM discount factors which are all identical over entire interest periods (t_i, t_{i+1}). We can therefore set $B_{\overline{R}}(t_i,t_{i+1}) = b\ \forall i$ for a constant b. We then re-write the YTM discount factors in the following way

$$\begin{aligned} B_{\overline{R}}(t,t_i) &= \underbrace{B_{\overline{R}}(t,t_{m+1})}_{\substack{\text{Partial}\\ \text{Coupon Period}}}\underbrace{B_{\overline{R}}(t_{m+1},t_{m+2})}_{b}B_{\overline{R}}(t_{m+2},t_i) \\ &= B_{\overline{R}}(t,t_{m+1})b^2B_{\overline{R}}(t_{m+3},t_i) = \ \cdots\ = B_{\overline{R}}(t,t_{m+1})b^{i-m-1}\underbrace{B_{\overline{R}}(t_{m+(i-m)},t_i)}_{=1} \end{aligned}$$

Replacing the spot rate discount factors in Equation 16.6 with these YTM discount factors gives

$$B_{\overline{R}}(t,t_n) = B_{\overline{R}}(t,t_{m+1})b^{n-(m+1)}$$
$$\sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i) = \underbrace{B_{\overline{R}}(t,t_{m+1})}_{\substack{\text{Partial}\\ \text{Coupon Period}}} \sum_{i=m+1}^{n} \underbrace{b^{i-(m+1)}}_{\substack{\text{Entire}\\ \text{Coupon Periods}}} = B_{\overline{R}}(t,t_{m+1})b^{n-(m+1)}\frac{1-b^{m-n}}{1-b^{-1}} \tag{16.7}$$

where the formula for the sum of a *geometric series* has been used in the last step. Geometric series are introduced in all elementary mathematics texts and collections of mathematical formulas[1]. We will often use such series in what follows. The quintessential property of geometric series is that their finite sums are given by the following equation:

$$\sum_{k=1}^{n} c^{k-1} = \frac{c^n - 1}{c-1} \quad \text{for any constant } c \tag{16.8}$$

Application of this equation in the above sum over the powers of the constant b leads[2] to Equation 16.7. The present value of a bond with coupon periods of equal length expressed in terms of its yield to maturity is thus

$$V_{\overline{R}}(t,T) = NB_{\overline{R}}(t,t_{m+1})b^{n-(m+1)}\left[1 + c\,\frac{1-b^{m-n}}{1-b^{-1}}\right]$$
$$\text{with} \quad b = B_{\overline{R}}(t_m,t_{m+1}) \;, \quad c = B_K(t_m,t_{m+1})^{-1} - 1 \tag{16.9}$$

This is a closed form expression for the present value of a bond as a function of its yield to maturity. It contains the discount factor for the partial interest period from t (today) until the next coupon payment date t_{m+1} and two parameters b and c which are determined through the discount factors with respect to the coupon and the yield to maturity for the next complete coupon period (from t_m until t_{m+1}).

In Figure 16.1, fifteen different bonds with a 6% coupon, annual coupon payments and remaining lifetimes between 1 and 15 years were priced using Equation 16.6 with the spot rates, and additionally, using Equation 16.9 with a constant rate for each

[1] A sequence x_k, having the property that the quotient of two elements, one immediately following the other, is constant, $x_{k+1}/x_k = c$ for all k, is called a *geometric sequence*. Thus, for each of the elements of a geometric sequence, $x_k = x_1\,c^{k-1}$. The sum of the elements of a geometric sequence is called a *geometric series*. The following formula holds for series of this type:

$\sum_{k=1}^{n} x_k = x_1 \sum_{k=1}^{n} c^{k-1} = x_1 \frac{c^n-1}{c-1}$

[2] Set $k = i - m$. The geometric series is then:

$$\sum_{i=m+1}^{i=n} b^{i-(m+1)} = \sum_{k+m=m+1}^{k+m=n} b^{k+m-(m+1)} = \underbrace{\sum_{k=1}^{n-m} b^{k-1}}_{\text{Geom. Series}} = \frac{b^{n-m}-1}{b-1}$$
$$= b^{n-m-1}\frac{b(1-b^{m-n})}{b-1} = b^{n-m-1}\frac{b(1-b^{m-n})}{b(1-b^{-1})}$$

Time to Maturity	Interest Rates				6% Coupon Bond	
	Spot Rates		Par Rates		Present Values	
Years	Interest Rate for Maturity	Zero-Coupon Bond	Interest Rate for Maturity	Discount for one Period	Using Spot Rates	Using Par Rates
1	2.20%	0.97847	2.200%	0.97847	103.72%	103.72%
2	2.50%	0.95181	2.496%	0.97565	106.76%	106.75%
3	2.90%	0.91781	2.886%	0.97195	108.87%	108.83%
4	3.30%	0.87821	3.268%	0.96835	110.18%	110.09%
5	3.70%	0.83389	3.643%	0.96485	110.75%	110.60%
6	4.00%	0.79031	3.919%	0.96229	111.13%	110.94%
7	4.30%	0.74475	4.188%	0.95981	111.05%	110.80%
8	4.50%	0.70319	4.366%	0.95817	111.11%	110.84%
9	4.70%	0.66142	4.539%	0.95658	110.90%	110.60%
10	4.90%	0.61979	4.706%	0.95506	110.46%	110.14%
11	5.00%	0.58468	4.793%	0.95426	110.45%	110.13%
12	5.10%	0.55051	4.878%	0.95349	110.34%	110.01%
13	5.20%	0.51736	4.959%	0.95275	110.13%	109.80%
14	5.25%	0.48853	5.004%	0.95234	110.18%	109.85%
15	5.30%	0.46087	5.047%	0.95195	110.18%	109.86%

Figure 16.1: Valuation of bonds with annual coupon payments using spot rates and par rates with annual compounding.

bond. The YTMs were not used for the constant rate in Equation 16.9 (this would have by definition given the exact same price as the spot rates), but we used the par rates corresponding to each particular time to maturity. The par rates were calculated previously using 16.11 (see the following section). It is clear that spot and par rates lead to somewhat different prices. This illustrates the fact already commented upon in Section 6.3: par rates and the yield to maturity are two different interest rates.

16.3.3 Par Rates

Bonds are fixed-rate instruments. Therefore, the at-par condition 6.6 can be used in calculating the par rate. Since plain vanilla bonds before maturity pay only interest (amortization payments being made only upon maturity) the "residual debt" during the entire lifetime is equal to the principal N compounded at respect to the coupon rate as of the last coupon payment. Inserting the cash flows in Equation 16.4 into the general at-par condition 6.6, the principals N cancel and the at-par condition for bonds becomes

$$B_K(t_m,t)^{-1} = \sum_{i=m+1}^{n} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] + B_R(t,T)$$
$$\text{for} \quad t_m \le t < t_{m+1} < t_{m+2} < \cdots < t_n = T \quad , \quad n \ge m+1 \tag{16.10}$$

The coupon compounding factor (multiplied by the nominal value of one monetary unit) from the last coupon payment date until time t appears on the left hand side.

On the right, we have the present value at time t of a bond with coupon K and a principal of 1 monetary unit. The at-par condition states that if the coupon K of a bond corresponds to its par rate then its present value is equal to its principal compounded since the last coupon payment date at the par rate.

If a coupon bond is quoted at-par, then its YTM is equal to the par rate (see Equation 6.7). We again derive this result specifically for bonds. Consider a bond whose coupon rate is equal to the YTM. The present value of such a bond, calculated by discounting the future cash flows at the YTM, is then[3]

$$
\begin{aligned}
&V_{\overline{R}}(t,T)\\
&= N \sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i)[B_{\overline{R}}(t_{i-1},t_i)^{-1} - 1] + NB_{\overline{R}}(t,T)\\
&= N \sum_{i=m+1}^{n} \left[\frac{B_{\overline{R}}(t,t_i)}{B_{\overline{R}}(t_{i-1},t_i)} - B_{\overline{R}}(t,t_i)\right] + NB_{\overline{R}}(t,t_n)\\
&= \underbrace{NB_{\overline{R}}(t_m,t)^{-1} - NB_{\overline{R}}(t,t_{m+1})}_{\text{Summand for } i=m+1} + N \sum_{i=m+2}^{n} [B_{\overline{R}}(t,t_{i-1}) - B_{\overline{R}}(t,t_i)] + NB_{\overline{R}}(t,t_n)\\
&= NB_{\overline{R}}(t_m,t)^{-1} - NB_{\overline{R}}(t,t_{m+1}) + NB_{\overline{R}}(t,t_{m+1}) - B_{\overline{R}}(t,t_n) + NB_{\overline{R}}(t,t_n)\\
&= NB_{\overline{R}}(t_m,t)^{-1}
\end{aligned}
$$

In fact, the present value equals the principal compounded as of the last coupon payment date. A coupon bond with the YTM as the coupon compounding rate thus satisfies the at-par condition. This implies that its coupon rate (the YTM) is equal to the par rate.

The graph of the par rates as a function of the time to maturity of the coupon bond is called the *par yield curve*. The par yield curve is then the yield to maturity of at-par quoted coupon bonds as a function of the maturity.

The special case where only one future cash flow payment date remains ($n = m+1$), namely the maturity, is quite interesting. The sum in Equation 16.10 has only one remaining term in this case. With $i = n$, we obtain

$$
\begin{aligned}
B_K(t_m,t)^{-1} = B_R(t,t_n)B_K(t_m,t_n)^{-1} \Rightarrow\\
B_R(t,t_n) = \frac{B_K(t_m,t_n)}{B_K(t_m,t)} = B_K(t,t_n) \quad \text{for} \quad n = m+1
\end{aligned}
$$

Thus, the par rate of the coupon bond after the last coupon payment before maturity is equal to the spot rate for this maturity.

For coupon periods of equal length, the interest payments from the coupon are also

[3] To do so, Equation 3.4 for constant interest rates is used:

$$
\frac{B_{\bar{R}}(t,t_i)}{B_{\bar{R}}(t_{i-1},t_i)} = \begin{cases} B_{\bar{R}}(t,t_{i-1}) & i > m+1 \\ B_{\bar{R}}(t_m,t)^{-1} & i = m+1 \end{cases}
$$

equal for each period. With c as defined in Equation 16.9, the at-par condition is then

$$B_K(t_m,t)^{-1} = c \sum_{i=m+1}^{n} B_R(t,t_i) + B_R(t,T) \tag{16.11}$$

If t coincides with the beginning of a coupon period, i.e. for $t = t_m$, then $B_K(t_m,t) = 1$. The par rate discount factor for one coupon period can then be written as a function of the spot rates:

$$c = \frac{1 - B_R(t,t_n)}{\sum_{j=m+1}^{n} B_R(t,t_j)} \text{ for } t = t_m$$

$$\Rightarrow$$

$$B_K(t_m,t_{m+1}) = \frac{\sum_{j=m+1}^{n} B_R(t_m,t_j)}{1 + \sum_{k=m+1}^{n-1} B_R(t_m,t_k)}$$

Finally, with Equation 16.7, we obtain the relation between the par rate and the YTM of a coupon bond with coupon periods of equal length:

$$B_K(t,t_m)^{-1} = B_{\overline{R}}(t,t_{m+1}) b^{n-(m+1)} \left[1 + c\frac{1-b^{m-n}}{1-b^{-1}}\right] \tag{16.12}$$

16.3.4 Sensitivities

The sensitivities with respect to the yield to maturity, namely the modified duration and the convexity, are calculated using Equation 16.5:

$$D_{\text{mod}} = -\frac{1}{V}\frac{\partial V}{\partial \overline{R}} = \frac{N}{V}\sum_{i=m+1}^{n}\left[1 - B_K(t_{i-1},t_i)^{-1}\right]\frac{\partial B_{\overline{R}}(t,t_i)}{\partial \overline{R}} - \frac{N}{V}\frac{\partial B_{\overline{R}}(t,t_n)}{\partial \overline{R}} \tag{16.13}$$

$$\text{Convexity} = \frac{1}{V}\frac{\partial^2 V}{\partial \overline{R}^2} = \frac{N}{V}\sum_{i=m+1}^{n}\left[B_K(t_{i-1},t_i)^{-1} - 1\right]\frac{\partial^2 B_{\overline{R}}(t,t_i)}{\partial \overline{R}^2} + \frac{N}{V}\frac{\partial^2 B_{\overline{R}}(t,t_n)}{\partial \overline{R}^2}$$

If all coupon periods are of equal length, the closed form expression in Equation 16.9 can be differentiated explicitly with respect to the yield to maturity to obtain closed form expressions for the modified duration and the convexity of the bond. These expressions are not listed here as they are quite lengthy and illustrate no new concepts.

The sensitivities with respect to the spot rate, i.e. the key rate durations are

$$D_{\text{key}}(t,t_i) = \frac{N}{V}\left[1 - B_K(t_{i-1},t_i)^{-1}\right]\frac{\partial B_R(t,t_i)}{\partial R} \quad \forall \; i = m+1, \ldots, n-1$$

$$D_{\text{key}}(t,t_n) = \frac{N}{V} B_K(t_{n-1},t_n)^{-1}\frac{\partial B_R(t,t_n)}{\partial R} \tag{16.14}$$

Time to Maturity	Interest Rates		6% Coupon Bond				Zero Bonds	
	Spot Rates		PV	Sensitivities			Sensitivities	
Years	Interest Rate for Maturity	Zero Bonds	Using Spot Rates	Modified Duration in Years	Key Rate Duration for Maturity	Convexity in Years^2	Modified Duration x PV	Convexity x PV
1	2.20%	0.97847	103.72%	0.978	0.978	1.915	0.957	1.874
2	2.50%	0.95181	106.76%	1.898	1.844	5.503	1.857	5.436
3	2.90%	0.91781	108.87%	2.761	2.605	10.533	2.676	10.405
4	3.30%	0.87821	110.18%	3.572	3.272	16.810	3.402	16.470
5	3.70%	0.83389	110.75%	4.332	3.848	24.142	4.023	23.289
6	4.00%	0.79031	111.13%	5.050	4.349	32.420	4.563	30.737
7	4.30%	0.74475	111.05%	5.721	4.771	41.441	5.004	38.421
8	4.50%	0.70319	111.11%	6.356	5.136	51.183	5.390	46.482
9	4.70%	0.66142	110.90%	6.951	5.434	61.430	5.694	54.471
10	4.90%	0.61979	110.46%	7.504	5.670	72.040	5.919	62.187
11	5.00%	0.58468	110.45%	8.035	5.878	83.185	6.137	70.279
12	5.10%	0.55051	110.34%	8.532	6.038	94.584	6.299	78.077
13	5.20%	0.51736	110.13%	8.996	6.154	106.137	6.408	85.472
14	5.25%	0.48853	110.18%	9.443	6.252	118.032	6.513	93.046
15	5.30%	0.46087	110.18%	9.863	6.316	130.016	6.581	100.234

Figure 16.2: The sensitivities of the coupon bonds from Figure 16.1. Annualized compounding was used throughout.

The expressions for the sensitivities in Equations 16.13 and 16.14 appear more complicated than they actually are. The derivatives can be interpreted as the modified duration, key rate duration, and convexity of zero bonds, each multiplied by the present value of the respective zero bond (see Equation 6.15). For each zero bond, we must use the interest rate R with respect to which the derivative was taken, explicitly: the YTM must be used for the modified duration and convexity, the spot rate of the appropriate interest period for the key rate duration. The present value and sensitivities of the zero bond for all of the usual compounding conventions can be taken from Table 16.1. The calculation of the sensitivities of the fifteen coupon bonds in Figure 16.1 is demonstrated explicitly in Figure 16.2.

16.4 Swaps

Swaps are contracts used to exchange interest payments. In their simplest form, a plain vanilla swap, fixed cash flows are exchanged for floating cash flows.

A bank, which holds a coupon bond, for example, but which wishes to reduce the interest rate risk of this fixed rate instrument can initiate a Swap according to which the counterparty of the bank receives fixed interest payments from the bank while the bank in return receives floating interest payments from the counterparty (see Figure 16.3). If the swap's negotiated fixed rate is equal to the coupon of the bond then the fixed payments from the bond are transferred entirely to the counterparty and the

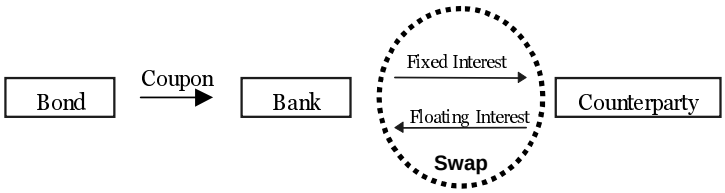


Figure 16.3: Reducing the interest rate risk of a bond by using a plain vanilla interest rate swap.

Short Name	**Complete Name**	**Corresponds to**	**Meaning**
Long Swap	Long Payer Swap	Short Receiver Swap	pay fixed interest and receive floating interest
Short Swap	Short Payer Swap	Long Receiver Swap	receive fixed interest and pay floating interest
Long Call	Long Call on a Payer Swap	Long Put on a Receiver Swap	right to pay fixed interest and receive floating interest
Long Put	Long Put on a Payer Swap	Long Call on a Receiver Swap	right to receive fixed and pay floating interest
Short Call	Short Call on a Payer Swap	Short Put on a Receiver Swap	obligation to receive fixed and pay floating interest
Short Put	Short Put on a Payer Swap	Short Call on a Receiver Swap	obligation to pay fixed and receive floating interest

Table 16.2: Long/short conventions for swaps and swaptions.

bank's net return is the floating interest received from the counterparty.

The meaning of the terms "long" and "short" are a bit ambiguous for swaps. We will therefore refer to swaps more precisely as payer or receiver swaps. In this book, the usual market convention will be adopted: "long" means "long payer", i.e. one *pays* the fixed rates. This is the exact opposite as was the case for bonds: one is long in a bond when one *receives* its fixed rate (i.e. coupon) cash flows. The long/short conventions for swaps extend to options on swaps, called *swaptions* (see Section 18.4.4). All conventions for swaps and swaptions are summarized in Table 16.2.

16.4.1 Cash Flows and Present Value

A *plain vanilla interest rate swap* (long payer) is equivalent to a portfolio consisting of a long floater and a short bond paying a coupon at the same rate as the fixed side of the swap. If the floater has a spread over the reference rate, it can be added to the coupon of the fixed side for the purpose of pricing the swap if the payment dates of the fixed and floating sides of the swap coincide. Otherwise, the floating side can, as mentioned in Section 16.2, be interpreted as a portfolio consisting of a floater, a bond and a zero bond. The present value of a swap can be computed directly from the price of a floater and that of a bond. For example, for the case of a long payer swap

$$\begin{aligned} V_R(t,T) &= V^{\text{Floater}} - V^{\text{Bond}} \\ &= N\frac{B_R(t,t_{m+1})}{B_R^{\text{fix}}(t_m,t_{m+1})} - N\sum_{i=m+1}^{n} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] - NB_R(t,t_n) \end{aligned} \tag{16.15}$$

If Equation 16.9 is used for the bond instead, we obtain the present value of the swap expressed in terms of the yield to maturity of the swap, assuming that the length of the interest periods are equal (where b and c are as defined in Equation 16.9):

$$V_{\overline{R}}(t,T) = NB_{\overline{R}}(t,t_{m+1})\left(B_R^{\text{fix}}(t_m,t_{m+1})^{-1} - b^{n-m-1}\left[1+c\,\frac{1-b^{m-n}}{1-b^{-1}}\right]\right) \tag{16.16}$$

In Figure 16.4, we consider a swap with a notional of 250,000 euros in which the 1-year floating rate is exchanged for a fixed interest rate of 6% over a period of 15 years. This swap can be priced in two different ways: in the first approach, a cash flow table of the swap is constructed, i.e. the expected cash flows as seen from today are calculated for both the fixed and floating sides (called *legs*) and the resulting net cash flow is determined. The present value is the sum of all net cash flows discounted at the current spot rates corresponding to the respective payment dates. The cash flows on the fixed side have already been established through the fixed interest rate and the principal agreed upon in the swap contract. The expected cash flows on the floating side are obtained from the forward rates for 1-year deposits. Calculation of the forward rates is demonstrated in Figure 3.6. The second approach is to interpret the swap as a floater minus a bond in accordance with Equation 16.15. In order to price the fixed side as a bond and the floating side as a floater, the repayment of the principal at maturity must be incorporated on each side. Since the amortization of the bond and the floater cancel each other out, we obtain exactly the same present value for the swap as before.

Furthermore, the YTM of the swap is calculated in this example (see Section 16.4.2). In doing so, all net cash flows were discounted at respect to the same interest rate. This interest rate parameter was adjusted until the sum of the discounted net cash flows was equal to the present value of the swap.

16.4.2 Yield to Maturity and Swap Rate

Since a swap has floating interest components, a "par rate" makes no sense at first. However, a similar concept for the fixed side of the swap can be defined. The *swap rate*

Interest Rates				Swap Fixed versus 1-Year Floating						
				6.00%	Notional Principal	250,000		YTM 1.48%		
Spot		1-Year Forward		Fixed Leg		Floating Leg		Net Cash Flow		
Time to Maturity	Spot Rate	Starting in Years	Forward Rate	Cash Flow	PV Using Spot Rates	Cash Flow	Present Value	Cash Flow	PV Using Spot Rates	PV Using YTM
1	2.20%	0	2.20%	15,000	14,677	5,500	5,382	-9,500	-9,295	-9,361
2	2.50%	1	2.80%	15,000	14,277	7,002	6,665	-7,998	-7,612	-7,766
3	2.90%	2	3.70%	15,000	13,767	9,262	8,501	-5,738	-5,267	-5,490
4	3.30%	3	4.51%	15,000	13,173	11,273	9,900	-3,727	-3,273	-3,513
5	3.70%	4	5.32%	15,000	12,508	13,289	11,081	-1,711	-1,427	-1,590
6	4.00%	5	5.51%	15,000	11,855	13,783	10,893	-1,217	-962	-1,114
7	4.30%	6	6.12%	15,000	11,171	15,296	11,391	296	220	267
8	4.50%	7	5.91%	15,000	10,548	14,777	10,391	-223	-157	-198
9	4.70%	8	6.31%	15,000	9,921	15,785	10,440	785	519	687
10	4.90%	9	6.72%	15,000	9,297	16,793	10,408	1,793	1,111	1,548
11	5.00%	10	6.01%	15,000	8,770	15,013	8,778	13	8	11
12	5.10%	11	6.21%	15,000	8,258	15,516	8,542	516	284	432
13	5.20%	12	6.41%	15,000	7,760	16,019	8,287	1,019	527	841
14	5.25%	13	5.90%	15,000	7,328	14,755	7,208	-245	-119	-199
15	5.30%	14	6.00%	15,000	6,913	15,006	6,916	6	3	5
Sum over Present Values					**160,224**		**134,784**		**-25,441**	**-25,441**
Amortization Bond and Floater				250,000	115,216	250,000	115,216			
				PV Bond	275,441	PV Floater	250,000	Floater - Bond	**-25,441**	

Figure 16.4: Cash flow table of a swap. The forward rates for 1-year periods starting in 1 to 14 years were calculated in Figure 3.6. Annual compounding was used. The present value of the swap is - 25,441 euros and its YTM is 1.48% p.a.

is defined as the coupon on the swap's fixed leg which makes the current present value, Equation 16.15, of the swap equal to zero. This is similar to the at-par condition on a coupon bond, Equation 16.10. The only difference is that the left hand side is not the residual debt of the bond but the present value of the swap's floating side (i.e. the present value of the floater):

$$N\frac{B_R(t,t_{m+1})}{B_R^{\text{fix}}(t_m,t_{m+1})} = N\sum_{i=m+1}^{n} B_R(t,t_i)\left[B_{K_S}(t_{i-1},t_i)^{-1}-1\right] + NB_R(t,t_n)$$

At the start of an interest period, $t = t_m$, the value of the floater is equal to its principal and the condition given by Equation 16.10 for the swap rate reduces to that for the par rate of a coupon bond. Thus,

At the start of an interest period, the swap rate is equal to the par rate of the bond on the fixed side of the swap.

For interest periods of equal length, the swap rate (for $t = t_m$) can be computed from the current spot rates using Equation 16.11, or from the yield to maturity using

Equation 16.12. Between coupon payment dates, however, there is no such simple way to determine the swap rates. In general, it can only be computed from the above condition using numerical iteration procedures.

In contrast to the par rate, the *yield to maturity* is also defined for floating rate instruments. The YTM of a swap can in principle be calculated from its present value, Equation 16.15. By definition, the interest rate to be determined (for example, by numerical iteration) is that which, when substituted for the spot rates in Equation 16.15, yields the current present value of the swap. This rate is then the swap's yield to maturity. Over identically long interest periods, Equation 16.16 can be used. At a fixing time $t = t_m$, the present value of the floater equals N and the remaining discount factor with respect to the YTM is equal to b. Consequently, Equation 16.16 reduces to an equation which can be easily solved for the YTM-discount factor b if the present value of the swap is known:

$$V_{\overline{R}}(t = t_m, T) = N\left(1 - b^{n-m}\left[1 + c\,\frac{1 - b^{m-n}}{1 - b^{-1}}\right]\right) \tag{16.17}$$
$$\text{where}\quad b = B_{\overline{R}}(t_m, t_{m+1})\,,\quad c = B_K(t_m, t_{m+1})^{-1} - 1$$

16.4.3 Sensitivities

The sensitivities with respect to the yield to maturity, namely the modified duration and the convexity, are

$$\begin{aligned}
D_{\text{mod}} &= -\frac{N}{V} B_R^{\text{fix}}(t_m, t_{m+1})^{-1} \frac{\partial B_{\overline{R}}(t, t_{m+1})}{\partial \overline{R}} \\
&\quad + \frac{N}{V} \sum_{i=m+1}^{n} [B_K(t_{i-1}, t_i)^{-1} - 1] \frac{\partial B_{\overline{R}}(t, t_i)}{\partial \overline{R}} + \frac{N}{V} \frac{\partial B_{\overline{R}}(t, t_n)}{\partial \overline{R}} \\
\text{Convexity} &= \frac{N}{V} B_R^{\text{fix}}(t_m, t_{m+1})^{-1} \frac{\partial^2 B_{\overline{R}}(t, t_{m+1})}{\partial \overline{R}^2} \\
&\quad - \frac{N}{V} \sum_{i=m+1}^{n} [B_K(t_{i-1}, t_i)^{-1} - 1] \frac{\partial^2 B_{\overline{R}}(t, t_i)}{\partial \overline{R}^2} - \frac{N}{V} \frac{\partial^2 B_{\overline{R}}(t, t_n)}{\partial \overline{R}}
\end{aligned} \tag{16.18}$$

The sensitivities with respect to the spot rates, i.e. the key rate durations, are

$$\begin{aligned}
D_{\text{key}}(t, t_{m+1}) &= \frac{N}{V}\left[B_K(t_m, t_{m+1})^{-1} - 1 - B_R^{\text{fix}}(t_m, t_{m+1})^{-1}\right] \frac{\partial B_R(t, t_{m+1})}{\partial R} \\
D_{\text{key}}(t, t_i) &= \frac{N}{V}\left[B_K(t_{i-1}, t_i)^{-1} - 1\right] \frac{\partial B_R(t, t_i)}{\partial R} \quad \forall\ i = m+2, \ldots, n-1 \\
D_{\text{key}}(t, t_n) &= -\frac{N}{V} B_K(t_{n-1}, t_n)^{-1} \frac{\partial B_R(t, t_n)}{\partial R}
\end{aligned} \tag{16.19}$$

The derivatives can again be interpreted as sensitivities of zero bonds, each multiplied by the present value of the corresponding zero bond as indicated in Equation 6.15. In doing so, the interest rate used for each zero bond must be the rate with respect to

which the derivative was taken. This means the YTM of the *swap* for the modified duration and the convexity, and the spot rate corresponding to the appropriate time to maturity for the key rate duration.

16.5 Annuity Loans

In light of the growing importance of credit risk management and the rise in popularity of credit derivatives, the necessity of evaluating the current value of a bank's loan portfolio has increased as well. The leading banking houses have already made the transition towards marking their loans to market. The valuation of the loan portfolio under current market conditions is a conditio sine qua non for the modern management of credit risks. The argument that loans cannot, as a rule, be traded like other securities and that the market value is therefore not realizable fails to hold water since the advent of *credit derivatives* such as *credit linked notes, total return swaps*, etc. which were conceived for the purpose of trading credit risks. The calculation of the mark-to-market of a loan will now be demonstrated for an annuity loan.

16.5.1 Cash Flows and Residual Debt

Like a bond, an *annuity loan* is determined by its cash flows. These cash flows are the constant installments C paid during the lifetime of the loan and the payment of the residual debt at maturity. The fundamental difference between an annuity loan and a coupon bond is that payments on the principal are made during the lifetime of the loan, i.e. before maturity. Thus, the *residual debt* $N(t)$ by reference to which the interest payments are determined, changes with time. The amount C paid at the end of each interest period covers the interest accrued in that period. The amount of C exceeding the interest payment due goes toward repaying the principal on the loan. The contractually agreed upon interest rate is denoted by K. The payment C can be decomposed into an interest cash flow C_Z and an amortization cash flow C_T as follows:

$$C = C_Z(t_i) + C_T(t_i) \quad \forall \ i = 1, \ \dots \ , \ n \tag{16.20}$$

where

$$C_Z(t_i) = N(t_{i-1}) \left[B_K(t_{i-1}, t_i)^{-1} - 1 \right]$$
$$C_T(t_i) = C - N(t_{i-1}) \left[B_K(t_{i-1}, t_i)^{-1} - 1 \right]$$

The residual debt $N(t_i)$ at time t_i is the amount on which interest must still be paid after the time t_i. An expression for this residual debt can be derived recursively: the residual debt $N(t_i)$ at time t_i equals the residual debt at time t_{i-1} less the amortization payment at time t_i

$$N(t_i) = N(t_{i-1}) - C_T(t_i) = N(t_{i-1}) B_K(t_{i-1}, t_i)^{-1} - C$$

This is a recursion relation. We can now use the same relation to express the residual debt at time t_{i-1} in terms of the residual debt at time t_{i-2}, repeating this procedure i

times until reaching the "residual debt" $N(t_0)$ at the time t_0, when the loan was made. This is just the original amount of the loan N.

$$\begin{aligned}
N(t_i) &= N(t_{i-1})B_K(t_{i-1},t_i)^{-1} - C \\
&= [N(t_{i-2})B_K(t_{i-2},t_{i-1})^{-1} - C]B_K(t_{i-1},t_i)^{-1} - C \\
&= N(t_{i-2})B_K(t_{i-2},t_i)^{-1} - CB_K(t_{i-1},t_i)^{-1} - C \\
&= [N(t_{i-2})B_K(t_{i-3},t_{i-2})^{-1} - C]B_K(t_{i-2},t_i)^{-1} - CB_K(t_{i-1},t_i)^{-1} - C \\
&= N(t_{i-3})B_K(t_{i-3},t_i)^{-1} - CB_K(t_{i-2},t_i)^{-1} - CB_K(t_{i-1},t_i)^{-1} - C \\
&= \dots \\
&= N(t_{i-i})B_K(t_{i-i},t_i)^{-1} - C\sum_{j=0}^{i-1} B_K(t_{i-j},t_i)^{-1}
\end{aligned}$$

and thus

$$\begin{aligned}
N(t_i) &= NB_K(t_0,t_i)^{-1} - C\sum_{j=0}^{i-1} B_K(t_{i-j},t_i)^{-1} \\
&\underset{l=i-j}{=} NB_K(t_0,t_i)^{-1} - C\sum_{l=1}^{i} B_K(t_l,t_i)^{-1}
\end{aligned}$$

Equation 6.5 implies now that (for $i = m$) the residual debt of an annuity loan at an *arbitrary* time t between the last (at time t_m) and next payment can be written as

$$N(t) = NB_K(t_0,t)^{-1} - C\sum_{l=1}^{m} B_K(t_l,t)^{-1} \tag{16.21}$$

This equation has an intuitive interpretation: the residual debt at time t is the original amount N of the loan compounded from the time t_0 when the loan was made until t at the fixed rate K less the installments C already paid, where each of the installments is also compounded from the payment date t_l until t.

In many cases, the installment periods are assumed to be of equal length. Since the fixed rate is the same for all periods, this implies that the discount factors $B_k(t_l,t_{l+1})$ for each entire payment period are also equal. We denote this value by a. Equation 3.3 can now be used to express the residual debt for a constant rate K as

$$\begin{aligned}
B_K(t_l,t) &= B_K(t_l,t_m)\underbrace{B_K(t_m,t)}_{\substack{\text{Partial}\\ \text{Period}}} = \underbrace{B_K(t_l,t_{l+1})}_{a}B_K(t_{l+1},t_m)B_K(t_m,t) \\
&= \dots = a^{m-l}\underbrace{B_K(t_{l+m-l},t_m)}_{1}B_K(t_m,t)
\end{aligned}$$

Summing the series using the Equation 16.8

$$\sum_{l=1}^{m} B_K(t_l,t_m)^{-1} = B_K(t_m,t)^{-1}\sum_{l=1}^{m} a^{l-m} = B_K(t_m,t)^{-1}a^{1-m}\underbrace{\sum_{l=1}^{m} a^{l-1}}_{\text{Geom. Series}}$$
$$= B_K(t_m,t)^{-1}a^{1-m}\frac{a^m-1}{a-1} = B_K(t_m,t)^{-1}\frac{a(1-a^{-m})}{a(1-a^{-1})}$$

reduces the equation for the residual debt for installment periods of equal length to

$$N(t) = \left[Na^{-m} - C\frac{1-a^{-m}}{1-a^{-1}}\right] B_K(t_m,t)^{-1} \tag{16.22}$$
$$\text{where } a = B_K(t_m,t_{m+1}) \;\;,\quad t_m \le t < t_{m+1}$$

16.5.2 Present Value

In addition to the installment payments, the final amortization of the residual debt $N(t_n)$ occurs at maturity, $T = t_n$. This is treated as a cash flow for the purpose of calculating the present value. The amount of this cash flow is given by Equation 16.21 with $t = t_n$. The present value as the sum of all discounted future cash flows thus becomes:

$$V(t,T) = \sum_{i=m+1}^{n} B_R(t,t_i)\underbrace{C}_{\text{Installments}} + B_R(t,t_n)\underbrace{N(t_n)}_{\text{Residual Debt}} \quad \text{for } t_m \le t < t_{m+1} \tag{16.23}$$
$$= C\sum_{i=m+1}^{n} B_R(t,t_i) + B_R(t,t_n)\left[NB_K(t_0,t_n)^{-1} - C\sum_{l=1}^{n} B_K(t_l,t_n)^{-1}\right]$$

Note that the residual debt and the installment payments are fully determined by the terms of the loan contract. The only variable parameters in the present value are the discount factors obtained from the current interest rate curve. In the corresponding equation for a bond (Equation 16.5), the spot rate discount factors also appear as the only parameters not fixed and are thus the only variables which are influenced by the market. Thus, by equating the coefficient of $B_R(t,t_i)$ for $i = m+1,\dots n$ in Equations 16.5 and 16.23, we obtain a bond whose behavior with respect to its present value and market risk is exactly the same as the loan, if the same spot rate curve is used as was for the loan.

$$N_{\text{Bond}}\left[B_K(t_{i-1},t_i)^{-1} - 1\right]_{\text{Bond}} = C_{\text{Loan}} \tag{16.24}$$
$$N_{\text{Bond}} = N(t_n)_{\text{Loan}}$$
$$= \left[NB_K(t_0,t)^{-1} - C\sum_{l=1}^{n} B_K(t_l,t)^{-1}\right]_{\text{Loan}}$$

An annuity loan with a principal of N, a fixed rate K and constant installments C can, in accordance with Equation 16.24, be interpreted as a coupon bond whose principal is equal to the residual debt of the loan at maturity and whose coupon is chosen such that the cash flows resulting from the coupon payments are exactly the same as the cash flows from the interest *and amortization* of the loan. All cash flows of such a bond are thus exactly those of the corresponding loan. In terms of valuation, cash management, and risk management, there are therefore no fundamental differences between bonds and annuity loans.

In general, the residual debt of the loan at maturity is not equal to zero, even when the term of the loan runs until complete amortization. At time t_n of the last installment payment, the residual debt is not exactly equal to the amortization component of the installment C, but is usually somewhat larger. The last payment date t_n of such a loan is the date after which the residual debt is smaller than C. As a rule, this residual debt is paid at t_n as well. For this type of loan, the residual debt and thus the principal of the associated bond is indeed quite small but Equation 16.24 can be applied nevertheless. If a loan happens to be completely amortized upon payment of the final installment, Equation 16.23 gives the present value as simply $\sum B_R(t, t_i)\, C$, i.e. equal to the present value of a series of zero bonds, each with face value C and whose respective maturities correspond to the installment payment dates of the loan. An interpretation as a coupon bond is thus no longer necessary.

An explicit example of these concepts is provided in Figure 16.5 for an $8\frac{1}{2}$ year loan of one million euros at a fixed rate of 8.4%, with semi-annual installment payments of exactly 50,000 euros (this corresponds to an initial annual amortization of 1.81%). The associated bond has a face value of 746,470 euros and a coupon of 13.85%. Observe that for the calculation of the bond, all relevant information can be obtained from the terms of the loan contract. The current term structure and even the current present value is not required. Loans, in principle, do not differ significantly from bonds.

The YTMs and par rates for the loan and its equivalent bond are included in Figure 16.5. A further discussion of the YTM and par rate for the loan can be found in Section 16.5.3. The interest rates for the bond are determined as follows: in the column "PV Using YTM" the *future* cash flows are discounted at the YTM. In accordance with Table 6.2, the value of the YTM was adjusted (using the Excel function "Goal Seek") until the sum of the thus discounted cash flows equaled the present value obtained by using the current spot rates. In the column "PV Using Spot Rate" under "Coupon = Par Rate", the present value of a bond with the par rate as the coupon was calculated varying the par rate until the present value thus calculated equaled the residual debt.

Note that both for the bond and the loan, the par rate is not the same as the YTM. Furthermore, the par rate of the loan and the equivalent bond are differ significantly although the future cash flows are exactly the same (only the YTMs are the same since these depend exclusively on the future cash flows)! This is because the bond and the loan have different *current* residual debts. The par rate, an extremely instrument-sensitive parameter, reveals this difference. An instrument is thus not fully characterized by its future cash flows. The open claims associated with it are needed as well. The instrument is only completely specified when both the future cash flows and the current residual debt are known.

In the above example, the accrued interest, defined as the difference between the

	Bond	Loan
Principal	788,264	1,000,000
Value Date	Feb-21-01	Feb-21-01
	Original Term	8.5
Time to Maturity	7.358	7.358
Coupon Period	0.50	0.50
Installment C	50,000	50,000
	Initial Yearly Amortization	1.81%
Fixed Rate	**13.09%**	**8.40%**
YTM	**9.35%**	**9.35%**
Par Rate	**9.45%**	**9.39%**

			Equivalent Bond				
IR Term Structure			Coupon=Fixed Rate			Coupon=Par Rate	
Payment Date	Time 30/E360	Spot Rate	Cash Flow	PV Using Spot Rates	PV Using YTM	PV Using Spot Rates	Cash Flow
Jun-30-01	0.358	2.20%	50,000	49,612	48,424	36,124	36,407
Dez-31-01	0.861	3.20%	50,000	48,662	46,297	35,432	36,407
Jun-30-02	1.358	4.50%	50,000	47,098	44,285	34,294	36,407
Dez-31-02	1.861	5.50%	50,000	45,258	42,339	32,954	36,407
Jun-30-03	2.358	6.30%	50,000	43,291	40,499	31,521	36,407
Dez-31-03	2.861	7.10%	50,000	41,090	38,720	29,919	36,407
Jun-30-04	3.358	7.90%	50,000	38,732	37,037	28,202	36,407
Dez-31-04	3.861	8.50%	50,000	36,490	35,410	26,569	36,407
Jun-30-05	4.358	9.00%	50,000	34,344	33,871	25,007	36,407
Dez-31-05	4.861	9.30%	50,000	32,451	32,383	23,629	36,407
Jun-30-06	5.358	9.50%	50,000	30,745	30,975	22,386	36,407
Dez-31-06	5.861	9.60%	50,000	29,217	29,614	21,274	36,407
Jun-30-07	6.358	9.70%	50,000	27,754	28,327	20,208	36,407
Dez-31-07	6.861	9.80%	50,000	26,326	27,083	19,169	36,407
Jun-30-08	7.358	9.90%	838,264	418,510	434,318	411,724	824,670
Values as at				**Present Value**	**949,581**	**949,581**	**798,413**
Feb-21-01				**Residual Debt**	**802,120**	**802,120**	**798,413**
				Accrued Interest	**13,856**	**13,856**	**10,149**
				Market Price	**118.71%**	**118.71%**	**100.00%**

Figure 16.5: The equivalent coupon bond for an annuity loan with fixed rate 8,40% per year and semi-annual payments.

residual debt and the face value, is also calculated, as well as the price of the bond, quoted as the clean price (present value less the accrued interest) divided by the face value. It is clear to see that the price of the bond with the par rate as the coupon is exactly 100%.

In many cases, all interest periods are assumed to be of equal length. If we then use Equation 16.22 for the residual debt at maturity, the resulting present value of the loan as given in 16.23 becomes

$$V(t,T) = \left[Na^{-n} - C\frac{1-a^{-n}}{1-a^{-1}} \right] B_R(t,t_n) + C \sum_{i=m+1}^{n} B_R(t,t_i) \tag{16.25}$$

Discounting at the current YTM of the loan rather than the current zero bond yield (i.e. setting R = yield to maturity), the discount factors with respect to R over an entire interest period are also equal (still assuming that the interest periods are of equal length): $B_{\overline{R}}(t_i,t_{i+1}) = b \ \forall i$. Then, as was done for bonds, Equation 16.7 can be applied to the sum consisting of the discount factors. The present value of a annuity loan with identical interest periods can then be expressed in terms of its yield to maturity as

$$V_{\overline{R}}(t,T) = \left[Na^{-n} - C\left(\frac{1-a^{-n}}{1-a^{-1}} - \frac{1-b^{m-n}}{1-b^{-1}} \right) \right] b^{n-m-1} B_{\overline{R}}(t,t_{m+1})$$
$$\text{with } a = B_K(t_m,t_{m+1}) \quad , \quad b = B_{\overline{R}}(t_m,t_{m+1}) \tag{16.26}$$

For a loan *without* initial amortization, this equation reverts to Equation 16.9[4].

16.5.3 Yield To Maturity and Par Rates

Substituting the cash flows 16.20 into the at-par condition, Equation 6.6, this condition reads for annuity loans

$$N(t) = C \sum_{i=m+1}^{n} B_R(t,t_i) + B_R(t,t_n)N(t_n) \quad \text{with}$$
$$t_m \le t < t_{m+1} < t_{m+2} < \cdots < t_n = T\,, \quad n \ge m+1 \tag{16.27}$$

Finally, with Equation 16.21 for the residual debt, we have

$$\begin{aligned} & NB_K(t_0,t)^{-1} - C\sum_{l=1}^{m} B_K(t_l,t)^{-1} \\ & = C \sum_{i=m+1}^{n} B_R(t,t_i) + B_R(t,t_n)\left[NB_K(t_0,t_n)^{-1} - C\sum_{l=1}^{m} B_K(t_l,t_n)^{-1}\right] \end{aligned} \tag{16.28}$$

The fixed rate K satisfying this condition is the par rate. Comparison with Equation 16.10 reveals that, because of the amortization, the par rates of an annuity loan are quite different from those for bonds. This even holds for the equivalent bond associated with the loan as is quite evident from Figure 16.5. Thus, the yield curve (par rates of coupon bonds) cannot be utilized for the valuation of loans!

As stated in general in Equation 6.7, the YTM of a financial instrument quoted at par is equal to the par rate of that instrument. We now derive this result again specifically for loans. The present value (calculated by discounting the future cash flows with the YTM) of a loan whose fixed rate is equal to the YTM is

$$\begin{aligned} V(t,T) &= C \sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i) + B_{\overline{R}}(t,t_n)\left[NB_{\overline{R}}(t_0,t_n)^{-1} - C\sum_{l=1}^{n} B_{\overline{R}}(t_l,t_n)^{-1}\right] \\ &= C \sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i) + NB_{\overline{R}}(t_0,t)^{-1} \\ &\quad - CB_{\overline{R}}(t,t_n)\left(\sum_{l=1}^{m} B_{\overline{R}}(t_l,t_n)^{-1} + \sum_{l=m+1}^{n} B_{\overline{R}}(t_l,t_n)^{-1}]\right) \\ &= C \sum_{i=m+1}^{n} B_{\overline{R}}(t,t_i) + NB_{\overline{R}}(t_0,t)^{-1} - C\sum_{l=1}^{m} B_{\overline{R}}(t_l,t)^{-1} - C \sum_{l=m+1}^{n} B_{\overline{R}}(t,t_l) \end{aligned}$$

[4]Without amortization, $C = C_Z(t_i) = \underbrace{N(t_{i-1})}_{N}\left[B_K(t_{i-1},t_i)^{-1} - 1\right] = N(a^{-1} - 1) = Nc$ holds, where c has been defined as for bonds.

$$
\begin{aligned}
&= NB_{\overline{R}}(t_0,t)^{-1} - C\sum_{k=1}^{m} B_{\overline{R}}(t_k,t)^{-1} \\
&= B_{\overline{R}}(t_m,t)^{-1} \underbrace{\left[NB_{\overline{R}}(t_0,t_m)^{-1} - C\sum_{k=1}^{m} B_{\overline{R}}(t_k,t_m)^{-1} \right]}_{N(t_m)}
\end{aligned}
$$

The present value is in fact equal to the residual debt compounded since the last payment date. The at-par condition is thus satisfied for a loan whose fixed interest rate is equal to the YTM.

Of particular interest is the special case where only one future cash flow remains, namely that paid at maturity ($n = m + 1$). The at-par condition for the loan at such times is then

$$
\begin{aligned}
B_K(t_m,t)^{-1}N(t_m) &= B_R(t,t_{m+1})[C + N(t_{m+1})] \\
&= B_R(t,t_{m+1})[C + B_K(t_m,t_{m+1})^{-1}N(t_m) - C] \\
&= B_R(t,t_{m+1})B_K(t_m,t_{m+1})^{-1}N(t_m) \qquad \text{for} \quad n = m+1
\end{aligned}
$$

This means that the par rate is then equal to the spot rate until maturity:

$$
B_R(t,t_{m+1}) = \frac{B_K(t_m,t_{m+1})}{B_K(t_m,t)} = B_K(t,t_{m+1}) \qquad \text{for} \quad n = m+1
$$

For interest periods of equal length, the fixed-rate discount factors in each period are also equal. Consequently, the residual debt is of the form given in Equation 16.22 and the at-par condition 16.28 becomes

$$
\begin{aligned}
&B_K(t_m,t)^{-1}\left(Na^{-m} - C\frac{1-a^{-m}}{1-a^{-1}} \right) \\
&= C\sum_{i=m+1}^{n} B_R(t,t_i) + B_R(t,t_n)\left(Na^{-n} - C\frac{1-a^{-n}}{1-a^{-1}} \right)
\end{aligned} \tag{16.29}
$$

At the start of a period we have $t = t_m$ and therefore $B_K(t_m,t) = 1$. Then, this equation can be solved (for example, using numerical iteration) for the par rate discount factor a as a function of the spot rate discount factors.

With Equation 16.7 for the spot rate discount factors, we finally obtain the relation between the par rate and the YTM of a loan with interest periods of equal length:

$$
\begin{aligned}
&B_K(t_m,t)^{-1}\left(Na^{-m} - C\frac{1-a^{-m}}{1-a^{-1}} \right) \\
&= B_{\overline{R}}(t,t_{m+1})b^{n-(m+1)}\left[Na^{-n} - C\frac{1-a^{-n}}{1-a^{-1}} + C\frac{1-b^{m-n}}{1-a^{-1}} \right]
\end{aligned} \tag{16.30}
$$

Comparing this with Equation 16.12 clearly shows how different the par rate of an annuity loan is from that of a bond.

	IR Term Structure			Annuity Loan						
i	Payment Date	Time in Years 30/E360	Spot Rate	Cash Flow	Interest	Amortization	Residual Debt	Discounted at Spot Rates	Discounted at YTM	Residual Debt if Par Rate were Fixed Rate
0	Dez-31-99	-1.142					1,000,000			1,000,000
1	Jun-30-00	-0.642		50,000	41,153	8,847	991,153			995,905
2	Dez-31-00	-0.142		50,000	40,789	9,211	981,942			991,622
3	**Jun-30-01**	**0.358**	**2.20%**	**50,000**	**40,410**	**9,590**	**972,352**	**49,612**	**48,424**	**987,143**
4	Dez-31-01	0.861	3.20%	50,000	40,015	9,985	962,368	48,662	46,297	982,458
5	Jun-30-02	1.358	4.50%	50,000	39,605	10,395	951,972	47,098	44,285	977,558
6	Dez-31-02	1.861	5.50%	50,000	39,177	10,823	941,149	45,258	42,339	972,433
7	Jun-30-03	2.358	6.30%	50,000	38,731	11,269	929,880	43,291	40,499	967,073
8	Dez-31-03	2.861	7.10%	50,000	38,268	11,732	918,148	41,090	38,720	961,467
9	Jun-30-04	3.358	7.90%	50,000	37,785	12,215	905,933	38,732	37,037	955,603
10	Dez-31-04	3.861	8.50%	50,000	37,282	12,718	893,215	36,490	35,410	949,470
11	Jun-30-05	4.358	9.00%	50,000	36,759	13,241	879,973	34,344	33,871	943,056
12	Dez-31-05	4.861	9.30%	50,000	36,214	13,786	866,187	32,451	32,383	936,347
13	Jun-30-06	5.358	9.50%	50,000	35,646	14,354	851,833	30,745	30,975	929,330
14	Dez-31-06	5.861	9.60%	50,000	35,056	14,944	836,889	29,217	29,614	921,991
15	Jun-30-07	6.358	9.70%	50,000	34,441	15,559	821,330	27,754	28,327	914,316
16	Dez-31-07	6.861	9.80%	50,000	33,800	16,200	805,130	26,326	27,083	906,288
17	Jun-30-08	7.358	9.90%	50,000	33,134	16,866	788,264	24,963	25,906	897,891
							PV Final Residual Debt	393,547	408,412	448,279
	Values as of						**Present Value Loan**	**949,581**	**949,581**	**1,004,313**
	Feb-21-01						**Present Residual Debt**	**993,227**	**993,227**	**1,004,313**

Figure 16.6: Cash flow table of the annuity loan from Figure 16.5. The next payment date is Jun–30-01.Thus, in our notation we have $m = 2$ and $n = 17$. Par rate and YTM were determined via numerical iteration. The results for those yields are displayed in Figure 16.5.

In Figure 16.6, all cash flows of the loan from Figure 16.5 are listed and priced with respect to the spot rates belonging to the debtors credit worthiness. The two columns "Interest" and "Amortization" show explicitly how the constant installment (cash flow) can be decomposed into interest (at the fixed rate of 8.40% specified in the contract) and amortization payments. The residual debt remaining after each payment is shown in the column "Residual Debt". All of the values for the residual debt listed in the example were computed using Equation 16.21 or Equation 16.22. At the valuation date Feb-21-01, the present value of the loan is 949,581.00 euros, the current residual debt (i.e. the residual debt as of the last payment date compounded up to February 21st), however, is equal to 993,227.00 euros. The present value is thus less than the remaining debt! This means that the interest paid by the debtor on the loan is too low considering his or her credit worthiness. This example demonstrates the mark to market of a loan.

In the column "Residual Debt Using Par Rate as Fixed Rate", the residual debt was calculated as if the par rate were the fixed interest rate instead of the rate agreed to in the contract. Using Equation 16.27 and Equation 16.29, the par rate was adjusted (using the Excel function "Goal Seek") until the present value of the loan (calculated as

before with the spot rates) was equal to the current residual debt. The par rate is thus the rate which the debtor should pay if the transaction is to be "fair". At 9.49%, it is substantially higher than the agreed upon 8.40% rate. In the column "Present Value with YTM" the cash flows and the final residual debt (calculated at the contractually agreed upon rate) were discounted at the YTM rather than with the spot rates and the YTM was adjusted until the sum of these discounted cash flows equaled the actual present value. All of these yields are listed in Figure 16.5.

16.5.4 Sensitivities

The sensitivities of an annuity loan with respect to the yield to maturity, i.e. the modified duration and the convexity are given in their full generality (i.e. without the assumption that the interest periods are of equal length) by Equation 16.21 as

$$\begin{aligned} D_{\text{mod}} &= -\frac{1}{V}\frac{\partial V}{\partial \overline{R}} = -\frac{C}{V}\sum_{i=m+1}^{n}\frac{\partial B_{\overline{R}}(t,t_i)}{\partial \overline{R}} - \frac{N(t_n)}{V}\frac{\partial B_{\overline{R}}(t,t_n)}{\partial \overline{R}} \\ \text{Convexity} &= \frac{1}{V}\frac{\partial^2 V}{\partial \overline{R}^2} = \frac{C}{V}\sum_{i=m+1}^{n}\frac{\partial^2 B_{\overline{R}}(t,t_i)}{\partial \overline{R}^2} + \frac{N(t_n)}{V}\frac{\partial^2 B_{\overline{R}}(t,t_n)}{\partial \overline{R}^2} \end{aligned} \tag{16.31}$$

If the interest periods are of equal length, the closed form Equation 16.26 can be differentiated explicitly with respect to the yield to maturity to obtain a closed form expression for the modified duration and the convexity of an annuity loan. These expressions are not given here as they are quite lengthy and provide no new insights.

The sensitivities with respect to the spot rates, namely the key rate durations are

$$\begin{aligned} D_{\text{key}}(t,t_i) &= -\frac{C}{V}\frac{\partial B_R(t,t_i)}{\partial R} \quad \forall\ i = m+1,\ \ldots\ ,n-1 \\ D_{\text{key}}(t,t_n) &= -\frac{C+N(t_n)}{V}\frac{\partial B_R(t,t_n)}{\partial R} \end{aligned} \tag{16.32}$$

The derivatives can, again using Equation 6.15, be interpreted as the sensitivities of zero bonds, each multiplied by the present value of the corresponding zero bond. In doing so, the rate R with respect to which the derivative was taken must be used in each of the zero bonds. This means that for the modified duration and the convexity, the YTM of the *loan* is taken, while for the key rate duration the spot rates, each with the appropriate time to maturity, is used.

Alternatively, the YTM and sensitivities can be calculated for the equivalent bond using Equation 16.24. Since the sensitivities depend exclusively on the future cash flows and not on the residual debt, the sensitivities for the bond listed in Figure 16.5 are equal to those of the associated annuity loan as well. In this way, mark-to-market and risk management of a credit portfolio can be performed using the traditional sensitivities.

16.5.5 Remarks on Credit Risk

The modified duration and convexity are well suited for an estimation of the credit risk posed by a loan (interpreted as *spread risk*). The only variables, i.e. parameters

not established in the loan contract, appearing in Equation 16.23 are the discount factors with respect to the spot rate curve. Only when the spot rate curve used in the valuation of the loan changes does the present value of the loan change. In contrast to a government bond, for example, the *credit worthiness* of the debtor as well as the market (market risk) plays a role: if the debtor becomes less credit worthy, the *spread* between the market curve (government bonds) and the interest rate curve associated with the credit worthiness of the debtor increases. This means that the level of the debtor's interest rate curve increases and the present value decreases according to Equation 16.23. As with market risks, credit risk can therefore be traced back to the sensitivity of a loan with respect to a change in the interest rate curve. But such changes can now have two different causes:

- A change in the market (market risk)
- A change in the credit worthiness effecting a change in the spread (credit risk)

Thus, credit risk in this context can be considered as a spread risk. Note that a *default* alone is not the only credit risk. A decrease in credit worthiness, as reflected by a decrease in the *rating* from AA to B, for example, would increase the obligor's spread significantly and thus result in a decrease in the present value of the loan; even if the obligor continues to pay all of his or her obligations. The *default risk* often quoted in Germany is an extreme special case of decrease in credit worthiness and therefore only a subset of the area of credit risk.

We could argue that a loss in present value as a result of a deterioration of credit rating (a decline from AA to B, for instance) is irrelevant as long as the debtor continues to make his or her interest and amortization payments and that therefore only a default represents an actual risk. This argumentation corresponds to the situation on the capital market a few decades ago when a bond was an instrument purchased with the intention of holding it until maturity (*buy and hold strategy*). Investors were then not interested in price fluctuations before maturity. Attitudes have since changed fundamentally and they will also change in the credit market. Loans will in future be treated as liquid tradable "instruments" and their present value will be realizable on the market. But even if a loan is held until maturity, a loss in present value in consequence of an increase in the interest rate spread represents *on average over many loans and over a longer time span* an actual loss in value since a higher interest spread is nothing other than a reflection of an increased *default probability*.

An increase in interest rate spreads represents an upward shift in the entire term structure. The assumption that the interest rates increase uniformly over all times to maturity by the same amount, i.e. a parallel shift of the entire curve, is a relatively good reflection of this situation (in contrast to market risk). The credit risk (as a spread risk) of an instrument is thus well quantified through the sensitivity of the present value with respect to parallel shifts of the term structure. This means that the parameters modified duration and convexity are well suited to quantify the credit risk (as opposed to the analogous situation of market risks).

17

Forward Transactions on Interest Rates

17.1 Forward Rate Agreements

A *forward rate agreement*, abbreviated as *FRA*, is a contract in which both contract parties agree to a fixed rate K on a principal N to be paid for some future interest period between T and T'. An FRA can be interpreted as an agreement loan to be made in the future with an interest rate already fixed today. The party receiving the loan makes the fixed interest payments. In contrast to bonds, we will refer to this party's position as a *long* position in the FRA, whereas the counterparty receiving the interest payments is *short* in the FRA. A (long) FRA can thus be interpreted as an agreement on two future cash flows: a receipt of the principal N at time T (the loan is made) and a payment at maturity T' of the FRA in the amount of the principal N compounded at the agreed rate K over the period $T'-T$ (the loan plus interest is paid back). Both of these cash flows discounted back to time t yield the present value of the FRA at time t:

$$\begin{aligned} F_R(T,T',K|t) &= B_R(t,T) \underbrace{N}_{\text{Cash Flow at Time } T} - B_R(t,T') \underbrace{\frac{N}{B_K(T,T')}}_{\text{Cash Flow at Time } T'} \\ &= NB_R(t,T)\left[1-\frac{B_R(T,T'|t)}{B_K(T,T')}\right] \end{aligned} \tag{17.1}$$

where the definition of the forward rate, Equation 3.3, was used in the last step. It is common practice in the market for forward transactions, the interest rate K is chosen so that the contract is worthless at the time it is concluded. As can be deduced from Equation 17.1, this condition is satisfied when the fixed rate K agreed upon in the FRA equals the *forward rate* at time t corresponding to the FRA period from T until T'. The last form of Equation 17.1 can be interpreted as follows: the present value of an FRA is the sum of its discounted future cash flows where these cash flows are discounted in two steps: first from end T' back to the beginning T of the FRA period

using the forward rate belonging to that period, then from T back to the current time t using the current spot rate for maturity T. FRAs are, for the purpose of valuation, interpreted as two future cash flows but are always settled in one net payment either at the beginning T or at the end T' of the FRA period. These settlement payments have of course to be equal to the value of the FRA at those times:

$$F_R(T,T',K|T) = N\left[1 - \frac{B_R(T,T')}{B_K(T,T')}\right] \qquad \text{Settlement at } T$$

$$F_R(T,T',K|T') = N\left[\frac{1}{B_R(T,T')} - \frac{1}{B_K(T,T')}\right] \quad \text{Settlement at } T'$$

17.2 Interest Rate Futures

17.2.1 Futures on Zero Bonds

The holder of a zero bond receives no interest payments before maturity. This corresponds to an underlying "without dividends". The forward price as seen from today's date t of a zero bond at time T having a face value of one monetary unit and maturing at $T' > T$ is given by Equation 7.1 as

$$S(t,T) = \underbrace{B_R(t,T)^{-1}}_{\text{Compounded Until } T} \underbrace{B_R(t,T')}_{\text{Present Value at } t} = \frac{B_R(t,T)B_R(T,T'|t)}{B_R(t,T)} = B_R(T,T'|t) \tag{17.2}$$

The forward price as seen from time t of a zero bond at T is thus the price which the bond would have at T if the forward rate as seen at t were the actual spot rate at time T. With Equation 7.5 we see that the present value of a future with delivery price K on such a zero bond is this forward price less the delivery price.

$$F_S^{\text{Zero Bond}}(t,T,T',K) = B_R(T,T'|t) - K$$

17.2.2 Futures on Coupon Bonds

We will determine the forward price of a coupon bond from the cash & carry arbitrage which led to Equation 7.1. Regardless of the underlying instrument, only two questions were of importance for the cash & carry arbitrage:

- What amounts must actually be paid in order to purchase the underlying today and at an agreed upon future date?
- Which cash flows arise for the holder of the underlying up to maturity which do *not* arise for the holder of the future?

For futures on coupon bonds, often just called *bond futures* for short, we therefore have to apply the cash & carry arbitrage for forward contracts on underlyings with dividend *payments*, where the dividend payments of course correspond to the coupon payments of the underlying bond. But *only* those coupon payments which arise during

the lifetime of the forward contract are to be taken into consideration[1]. Denoting the coupon payments at t_i by $C(t_i)$, the following holds for the dirty forward price of the bond

$$S(t,T) = \frac{S(t) - D(t,T)}{B_R(t,T)} \quad \text{for } D(t,T) = \sum_{t<t_i\leq T} C(t_i)B_R(t,t_i) \tag{17.3}$$

In this equation

- $D(t,T)$ is the present value of all coupon payments to be made up to the maturity of the future;
- $S(t,T)$ is the dirty price (including accrued interest) to be paid for the bond at time T as seen from time t;
- $S(t)$ is the dirty price (including accrued interest) at which the bond is available for purchase at time t.

According to Equation 7.5, the value of a future with a delivery price K on such a bond is this dirty forward price less the dirty delivery price. In the market, however, *clean* prices are quoted. Therefore the interest accrued up to time T must be added to the quoted (clean) delivery price. For forward and option contracts, all calculations must generally be made on the basis of dirty prices, i.e. the accrued interest should always be taken into consideration.

The bond futures traded on exchanges do not normally refer to real bonds as their underlying but rather to *synthetic bonds* which are defined by their coupon and time to maturity. Usually there are several real bonds which are quite similar to such a synthetic bond and which are therefore admissible as an underlying for the future under consideration. A future very popular in Germany, the *Euro Bund Future* (previously referred to as *Bund Future*) has as its underlying a synthetic German government bond with 10 years to maturity and a coupon of 6%, the underlying for the *Euro Bobl Future* (previously referred to as *Bobl Future*) is a synthetic bond with a 6% coupon and 5 years to maturity and the underlying of the *Euro Schatz Future* (previously referred to as *Schatz Future*) is a synthetic bond with a 6% coupon and 2 years to maturity. The face value of the synthetic underlying, called the *contract size*, is 100,000 euros

[1]Bonds *cannot* be interpreted as an underlying with a dividend *yield* q when pricing options and forward contracts on bonds. This is because the bond holder is *not* paid the yield if no coupon payment dates occur during the lifetime of the derivative.The accrued interest on the underlying is exchanged directly between the counterparties through the dirty price. In cash & carry arbitrage, only the "extra" cash flows which are not exchanged by the counterparties of the forward contract must be taken into consideration. These are just the coupon *payments* during the lifetime of the forward contract. These are received by the holder of the underlying, but not by the holder of the forward contract. Hence, the only consistent treatment of a bond is to interpret it as an underlying with dividend payments during the lifetime of the forward contract and to perform each calculation (for the the spot price, the forward price and option strikes) with the dirty price (= clean price plus accrued interest). This becomes particularly clear when *no* coupon payments occur up to time T : despite accrued interest, the bond compounds not at its yield to maturity but at the market rate holding for a time span up to T. Otherwise arbitrage would be possible. Thus, for a forward contract, such a bond should be considered as an underlying without dividends.

for all these futures. The admissible real underlyings that the holder of a short future positon is allowed to deliver at the maturity of such futures are

- for *Euro Bund Futures:* German federal bonds (*Bundesanleihen*) with a remaining term upon delivery of $8\frac{1}{2}$to $10\frac{1}{2}$years.
- for *Euro Bobl Futures:* German federal bonds (*Bundesanleihen*) and German federal debt obligations (*Bundesobligationen*) with a remaining term upon delivery of $4\frac{1}{2}$to $5\frac{1}{2}$years.
- for *Euro Schatz Futures:* German federal treasury notes (*Bundesschatzanweisungen*), German federal debt obligations (*Bundesobligationen*) and German federal bonds (*Bundesanleihen*), all with a remaining term upon delivery of $1\frac{3}{4}$to $2\frac{1}{4}$years.

In order to enable the investor to compare the prices of all the different instruments allowed to be delivered at maturity of a future, the futures exchange *EUREX* (previously called the *DTB*) calculates a *conversion factor* for each real deliverable bond which converts the clean price of the synthetic bond into the clean price of the deliverable bond. The EUREX calculates this conversion factor P and the accrued interest Z as follows:

$$P = \frac{1}{(1+K)^{m/12}} \left[\frac{K_L}{K} (1+K) + \frac{1}{(1+K)^n} \left(1 - \frac{K_L}{K} \right) \right] - K_L(1 - m/12)$$

$$Z = \frac{i}{360} N K_L$$

where

$K_L =$ coupon of the deliverable bond
$K =$ coupon of the synthetic bond $= 0,06$
$n =$ full years until the maturity of the deliverable bond
$m =$ full months until the next coupon payment of the deliverable bond
$i =$ interest days since the last coupon payment of the deliverable bond
$N =$ contract size = 100,000 euros

The same conversion factor is used for Bund, Bobl and Schatz futures (as well as for options on these instruments). These futures differ only in the lifetime of their underlying synthetic bond.

At the maturity of the future, the holder of the long position in this future must purchase the bond delivered by the holder of the short position at the future price $S(T,T) = S(T)$, multiplied by a conversion factor P, in addition to the accrued interest Z which is always paid upon purchase of a bond[2]. For a contract size N the total amount (denoted by g) payable by the holder of the long future is

$$g = N\, S(T)\, P + Z$$

[2]The actual price $S(T)$ and not the originally agreed upon delivery price $S(t_0, T)$ – where t_0 denotes the date the future contract was entered into – must be paid since the differences between the actual forward price of the underlying and the delivery price $S(t_0, T)$ have already been settled on a daily basis because of the variation margin system.

The holder of the short position delivers a deliverable bond in exchange for this amount g. This bond costs him or her the face value N multiplied by the quoted (clean) price $S_L(T)$ of the deliverable bond at time T plus the accrued interest. Subtracting from these costs the amount received from the holder of the long future, the net costs are

$$N\,S_L(T) + Z - g = N\,[S_L(T) - P\,S(T)]$$

These costs vary from bond to bond. The holder of the short position will therefore choose from all the deliverable bonds exactly that bond, which minimizes the holder's costs. In other words, the bond which minimizes the value of $S_L(T) - PS(T)$ will be chosen. This bond is called the *cheapest to deliver* or *CTD* for short. In fact, futures (and options as well) on synthetic bonds are priced as if they were written directly on the cheapest to deliver bond instead of on a synthetic bond. This means that the future's price is that given by Equation 17.3 using the forward price of the CTD bond (quoted as the clean price, i.e. less the interest accrued up to T and divided by the conversion factor). Without this procedure, Equation 17.3 could not be applied in practice since the coupon dates of the synthetic bond are not defined and thus the present value of the coupon payments of synthetic bonds cannot be determined.

The CTD bond is generally known on the market. From this bond, the forward price of the synthetic underlying (which is the one and only thing quoted in such futures markets) can be calculated in five steps

- Determine the dirty price at time t of the CTD from its quoted (clean) price plus the accrued interest.
- Determine the present value of the CTD's coupon payments due between times t and T.
- Calculate the dirty price of the CTD for the date T using Equation 17.3.
- Determine the (clean) price of the CTD for the date T from the dirty price less the interest accrued up to T.
- Divide this price by the conversion factor for the CTD to obtain the (clean) forward price of the synthetic bond.

17.3 Forward Swaps

A *forward swap*, also called a *deferred swap* is a swap with a lifetime over a *future* time span from T to T', for example in 3 years ($T = 3$) over 2 years ($T' = 5$). The cash flow dates are denoted by $T_0, T_1, ..., T_n = T'$ where T_0 denotes the last fixing before T. The times $T_1, ..., T_n$ are thus, as seen from T, still future cash flow dates. The relation between these times can be summarized by

$$t < T_0 \leq T < T_1 < T_2 < \cdots < T_n = T' \tag{17.4}$$

17.3.1 Present Value

Equation 16.15 for the present value of a swap holds for the future time T as well. According to Equation 16.3, the present value of the floating side of a forward swap at time T equals the principal compounded over the period ranging from the time of the last fixing (here T_0) until the next payment date at the interest rate fixed at time T_0. In contrast to Equation 16.3, this fixing also lies in the future, not in the past. Since the market is arbitrage free, the rate for the future time span from T_0 to T_1 has to be set equal to today's forward rate $R(T_0, T_1|t)$ for valuation purposes. Similiarly, to discount from the future payment date T_1 back to the *future* value date T (this happens in the numerator in Equation 16.3) today's forward rate corresponding to this time span must be used. The present value of this "Forward Floater" is thus the ratio of two forward rates. With Equation 3.4 we finally obtain for the present value of the floating side

$$\begin{aligned}\frac{B_R(T,T_1)}{B_R^{\text{fix}}(T_0,T_1)}N &= \frac{B_R(T,T_1|t)}{B_R(T_0,T_1|t)}N = \frac{B_R(t,T_1)}{B_R(t,T)}\frac{B_R(t,T_0)}{B_R(t,T_1)}N \\ &= \frac{B_R(t,T_0)}{B_R(t,T)}N = B_R(T_0,T|t)^{-1}N\end{aligned}$$

Consider now the present value of the swap's fixed side. Again, to ensure arbitrage freedom, the current forward rates at time t must be used in the discount factors with respect to the future time T. Thus, the generalization of Equation 16.15 for the present value of a swap at a *future* value date T paying an interest rate K on the fixed side is

$$\begin{aligned}V_R(T,T',K) &= NB_R(T_0,T|t)^{-1} \\ &\quad - N\sum_{i=1}^{n} B_R(T,T_i|t)\left[B_K(T_{i-1},T_i)^{-1}-1\right] - NB_R(T,T'|t)\end{aligned} \tag{17.5}$$

Finally, multiplying both sides by $B_R(t,T)$ and exploiting Equation 3.4 yields[3] the present value of the forward swap at time t as a function of the current spot rates:

$$\begin{aligned}V_R(T,T',K|t) &= NB_R(t,T_0) - NB_R(t,T') \\ &\quad - N\sum_{i=1}^{n} B_R(t,T_i)\left[B_K(T_{i-1},T_i)^{-1}-1\right]\end{aligned} \tag{17.6}$$

Note that T does not appear on the right-hand side of Equation 17.6! The value of a forward swap is therefore not directly dependent on T. As long as T lies between T_0 and T_1 with $t < T_0$, all future fixing and payment dates in Equation 17.4 remain the same, and therefore the value of the forward swap also remains the same. Thus, the exact point in time, at which a forward swap is entered into, is irrelevant for its value as seen from the time point t. The "date of entry" T can thus be arbitrarily chosen within the individual forward swap period. For this reason, forward swaps are generally written to begin at the start of one complete period, i.e. at $T = T_0$.

[3] Where we note that $B_R(T_0,T|t)^{-1} = \frac{B_R(t,T_0)}{B_R(t,T)}$

17.3.2 Forward Swap Rates and Forward Yield to Maturity

The present value of a swap whose fixed rate is equal to the swap rate is, by the definition of the swap rate, equal to zero. A swap of this type is called a par swap. The swap rate of a par swap in the *future* is referred to as the *forward swap rate*. This is the interest rate on the fixed side of a (forward) swap which makes the value of the swap at a *future* time T equal to zero (thus making today's value equal to zero as well). This rate will be denoted by $K_S(T, T'|t)$. Setting Equation 17.5 equal to zero gives the condition for the forward swap rate:

$$\sum_{i=1}^{n} B_R(T, T_i|t) \left[B_{K_S(T,T'|t)}(T_{i-1}, T_i)^{-1} - 1 \right] = B_R(T_0, T|t)^{-1} - B_R(T, T'|t) \quad (17.7)$$

Solving Equation 17.7 (for example, using numerical iteration) for the forward swap rate, we obtain this value as a function of the forward (zero) rates as of the future date T. Multiplying both sides by $B_R(t, T)$ and exploiting Equation 3.4 or, equivalently, setting Equation 17.6 equal to zero finally yields the forward swap rate as a function of the *current* zero rates (i.e. the spot rates):

$$\sum_{i=1}^{n} B_R(t, T_i) \left[B_{K_S(T,T|t)}(T_{i-1}, T_i)^{-1} - 1 \right] = B_R(t, T_0) - B_R(t, T') \quad (17.8)$$

The forward swap rate is the interest rate on the fixed side of a forward swap which has a present value (at time t) of zero. Comparing Equations 17.8 and 17.7 immediately shows that the forward swap rate converges to the swap rate as $t \longmapsto T$.

Like the present value, the forward swap rate at a time $t < T$ (see Equation 17.8), does not depend directly on time T. The forward swap rate does not change as long as T remains between T_0 and T_1 with $t < T_0$, in other words, as long as the relevant future fixing and payment dates in Equation 17.4 do not change. The time point within one period at which the swap is entered into is equally irrelevant for the forward swap rate as for the present value. T can be chosen arbitrarily within the same period of the forward swap and is always set at T_0.

For the special case that the interest periods of the forward swap are all of equal length, the compounding with respect to the forward swap rate over one period can be isolated:

$$B_{K_S(T,T'|t)}(T_0, T_1)^{-1} - 1 = \frac{B_R(t, T_0) - B_R(t, T')}{\sum_{i=1}^{n} B_R(t, T_i)} \quad (17.9)$$

Again, without loss of generality, the forward swap may begin at the start of an entire period and we set $T_0 = T$.

The present value of a forward swap, Equation 17.6, can also be interpreted as the difference to a forward *par* swap. Subtracting the value of the forward par swap (which is zero by definition) from Equation 17.6, the present value of the forward swap becomes

$$V_R(T, T', K|t) = N \sum_{i=1}^{n} B_R(t, T_i)[B_{K_s(T,T'|t)}(T_{i-1,}T_i)^{-1} - B_K(T_{i-1,}T_i)^{-1}] \quad (17.10)$$

This is precisely the discounted difference between the cash flows of the forward swap under consideration (with a fixed rate K) and those of the forward *par* swap. The interpretation of this equation, which represents the present value of a forward *payer* swap, is that the holder of the forward payer swap with a fixed rate K can at time t enter into a forward par *receiver* swap whose fixed rate is the current forward swap rate K_S (and with the same fixing dates as those of the forward payer swap) without incurring any additional costs. At the future fixing dates, the holder then receives interest at rate K_S from the receiver swap and must pay interest at rate K because of the payer swap. The net cash flow at each fixing date is just the difference between these two interest payments. As seen from today, all these cash flows are certain, i.e. risk free (no market risk is involved). They must merely be discounted back from the appropriate dates to today to yield the present value of the forward payer swap. If the current forward swap rate K_s is larger than the contractually agreed upon fixed rate K, the present value of this forward payer swap is positive, which makes sense since the holder *pays* the fixed rate K.

The *forward YTM* is the constant rate which, when used for the forward discount factors, yields the present value of a forward swap at time T. The general condition for the forward YTM thus follows from Equation 17.5

$$\begin{aligned}
&B_{\overline{R}}(T_0,T|t)^{-1} - \sum_{i=1}^{n} B_{\overline{R}}(T,T_i|t)[B_K(T_{i-1},T_i)^{-1}-1] - B_{\overline{R}}(T,T'|t) \\
&= B_R(T_0,T|t)^{-1} - \sum_{i=1}^{n} B_R(T,T_i|t)[B_K(T_{i-1},T_i)^{-1}-1] - B_R(T,T'|t) \\
&= B_R(t,T)^{-1}\left[B_R(t,T_0) - \sum_{i=1}^{n} B_R(t,T_i)[B_K(T_{i-1},T_i)^{-1}-1] - B_R(T,T_i')\right] \qquad (17.11)
\end{aligned}$$

The first equation gives the forward YTM as a function of the forward (zero) rates with respect to time T, the second as a function of the current spot rates. A similar relation holds between the forward swap rate and the forward YTM as for the corresponding spot interest rates, namely that the forward YTM of a forward par swap is equal to its forward swap rate. This can be shown quite easily: calculating the present value using the forward YTM and substituting the forward YTM for the fixed rate K does indeed yield a present value of zero (we perform the calculation here with $T = T_0$ for the sake of simplicity):

$$\begin{aligned}
&1 - \sum_{i=1}^{n} B_{\bar{R}}(T,T_i|t)\left[B_{\bar{R}}(T_{i-1},T_i|t)^{-1}-1\right] - B_{\bar{R}}(T,T'|t) \\
&= 1 - \sum_{i=1}^{n}\left[B_{\bar{R}}(T,T_{i-1}|t) - B_{\bar{R}}(T,T_i|t)\right] - B_{\bar{R}}(T,T'|t) \\
&= 1 - \underbrace{B_{\bar{R}}(T,T_0|t)}_{1} + B_{\bar{R}}(T,\underbrace{T_n}_{T'}|t) - B_{\bar{R}}(T,T'|t) = 0
\end{aligned}$$

This implies that, if the forward YTM is the fixed rate, the condition for the forward swap rate is satisfied and thus, the forward YTM is indeed the forward swap rate.

The time T appears in several terms in Equation 17.11 for the forward YTM. This means that the forward YTM depends strongly on the exact point in time at which the forward swap with a fixed rate K is entered into. The rather complicated general form Equation 17.11 becomes much simpler in concrete situations. As an example, we present the situation in which the interest periods are of equal length and the swap is entered into at the beginning of a complete period, at $T = T_0$. Similar to the analogous situation for bonds, Equation 16.7, the YTM discount factors can be written in the form of a geometric series and Equation 17.11 reduces to

$$\begin{aligned} b^n \left[\frac{1 - b^{-n}}{1 - b^{-1}} + 1 \right] &= \sum_{i=1}^{n} B_R(T, T_i|t) + c^{-1} B_R(T, T'|t) \\ &= B_R(t,T)^{-1} \left[\sum_{i=1}^{n} B_R(t, T_i) + c^{-1} B_R(t, T') \right] \end{aligned} \tag{17.12}$$

where we have introduced the constants $c = B_K(T_0, T_1)^{-1} - 1$ and $b = B_{\overline{R}}(T_0, T_1|t)$. Again, the first equation delivers the forward YTM as a function of the forward (zero) rates with respect to T, the second as a function of the *current* spot rate.

In Figure 17.1 the valuation of a forward swap and the calculation of the forward yield to maturity is demonstrated. The present values were first calculated as of the start date T of the forward swap and then (in the last row) discounted back to the current value date t. The forward swap on a principal of 250,000 euros has at time t a negative present value of -3,431 euros and a forward YTM of 6.69%. Note that both the forward rates for all interest periods (each one year long) and the complete forward term structure from the start of the forward swap (in 3 years) is required for this computation. These forward rates were calculated in Figure 3.6. Finally, Figure 17.2 demonstrates the computation of the forward swap rate for the lifetime and term structure given in Figure 17.1.

17.4 Forward Bonds

17.4.1 Present Value

Equation 16.5 for the present value of a bond also holds for future times T. For the discount factors with respect to the future time T, the current forward rates must be used in place of the spot rates to ensure arbitrage freedom. Thus, the value of a forward bond with an interest rate K as of a future value date T becomes

$$B_R(T, T', K) = N \sum_{i=1}^{n} B_R(T, T_i|t) \left[B_K(T_{i-1}, T_i)^{-1} - 1 \right] + B_R(T, T'|t) \tag{17.13}$$

Multiplying both sides by $B_R(t, T)$ and exploiting Equation 3.4 finally gives the present value of the forward bond at time t as a function of the current spot rate:

$$V_R(T, T', K|t) = N \sum_{i=1}^{n} B_R(t, T_i) \left[B_K(T_{i-1}, T_i)^{-1} - 1 \right] - N B_R(t, T') \tag{17.14}$$

Interest Rates						Swap Fixed vs. 1-Year Floating for 10 Years in 3 Years						
						6.00% Notional Principal 250,000				Forward YTM 6.69%		
Spot		For 1 Year		In 3 Years		Fixed Leg		Floating Leg		Net Cash Flow		
Time to Maturity	Spot Rate	in Years	Forward Rate	for Years	Forward Rate	Cash Flow	Present Value at T	Cash Flow	Present Value at T	Cash Flow	PV at T	PV Using YTM
1	2.20%	0	2.20%									
2	2.50%	1	2.80%									
3	2.90%	2	3.70%									
4	3.30%	3	4.51%	1	4.51%	15,000	14,353	11,273	10,787	-3,727	-3,566	-3,493
5	3.70%	4	5.32%	2	4.91%	15,000	13,628	13,289	12,074	-1,711	-1,555	-1,503
6	4.00%	5	5.51%	3	5.11%	15,000	12,916	13,783	11,868	-1,217	-1,048	-1,002
7	4.30%	6	6.12%	4	5.36%	15,000	12,172	15,296	12,411	296	240	228
8	4.50%	7	5.91%	5	5.47%	15,000	11,492	14,777	11,321	-223	-171	-161
9	4.70%	8	6.31%	6	5.61%	15,000	10,810	15,785	11,375	785	565	532
10	4.90%	9	6.72%	7	5.77%	15,000	10,129	16,793	11,340	1,793	1,211	1,140
11	5.00%	10	6.01%	8	5.80%	15,000	9,556	15,013	9,564	13	8	8
12	5.10%	11	6.21%	9	5.84%	15,000	8,997	15,516	9,307	516	309	288
13	5.20%	12	6.41%	10	5.90%	15,000	8,455	16,019	9,030	1,019	574	533
14	5.25%	13	5.90%	11	5.90%							
15	5.30%	14	6.00%	12	5.91%							
Sum of PVs as at Starting Date T							**112,509**		**109,077**		**-3,431**	**-3,431**
Sum of PVs as at Value Date t (today)							**103,262**		**100,112**		**-3,149**	**-3,149**

Figure 17.1: Forward swap fix vs. 12-month floating, starting in three years for ten years.

Interest Rates								Forward Par Swap			
						Notional Principal 250,000		Forward Swap Rate 5.82%			
Spot		For 1 Year		In 3 Years		Floating Leg		Fixed Leg		Net Cash Flow	
Time to Maturity	Spot Rate	in Years	Forward Rate	for Years	Forward Rate	Cash Flow	PV at T	Cash Flow	PV at T	Cash Flow	PV at T
1	2.20%	0	2.20%								
2	2.50%	1	2.80%								
3	2.90%	2	3.70%								
4	3.30%	3	4.51%	1	4.51%	11,273	10,787	14,543	13,915	-3,269	-3,128
5	3.70%	4	5.32%	2	4.91%	13,289	12,074	14,543	13,213	-1,254	-1,139
6	4.00%	5	5.51%	3	5.11%	13,783	11,868	14,543	12,522	-760	-654
7	4.30%	6	6.12%	4	5.36%	15,296	12,411	14,543	11,800	753	611
8	4.50%	7	5.91%	5	5.47%	14,777	11,321	14,543	11,142	234	180
9	4.70%	8	6.31%	6	5.61%	15,785	11,375	14,543	10,480	1,242	895
10	4.90%	9	6.72%	7	5.77%	16,793	11,340	14,543	9,820	2,251	1,520
11	5.00%	10	6.01%	8	5.80%	15,013	9,564	14,543	9,264	471	300
12	5.10%	11	6.21%	9	5.84%	15,516	9,307	14,543	8,723	973	584
13	5.20%	12	6.41%	10	5.90%	16,019	9,030	14,543	8,197	1,476	832
14	5.25%	13	5.90%	11	5.90%						
15	5.30%	14	6.00%	12	5.91%						
Sum of PVs as at Starting Date T							**109,077**		**109,077**		**0**
Sum of PVs as at Value Date t (today)							**100,112**		**100,112**		**0**

Figure 17.2: The forward *par* swap corresponding to Figure 17.1. Only the coupon (and therefore the cash flows) of the fixed leg have been changed. The forward swap rate which makes the present value vanish is 5.82%.

Note that T does not appear on the right-hand side of the equation. The value of the forward bond thus does not depend directly on time T as long as T remains between T_0 and T_1 with $t < T_0$, (i.e., as long as the future payment dates in Equation 17.4 do not change). The exact point in time within one period at which a forward bond is entered into is irrelevant for its value at an earlier time t. The maturity T of the forward contract can thus be chosen arbitrarily within the same period of the forward bond. For this reason, forward bond contracts are usually written to begin at the start of a complete period, i.e. with $T = T_0$.

17.4.2 Forward Par Rate and Forward Yield to Maturity

The present value of a bond whose fixed rate equals the par rate is, by definition of the par rate, equal to the residual debt. A bond of this type is referred to as a par bond. The par rate of a forward par bond is called the *forward par rate.* It is the fixed rate for which the value of the forward bond at time T equals its residual debt. This rate will be denoted by $K(T, T'|t)$. The residual debt of a bond is, according to Equation 16.10, equal to the face value compounded since the last payment date at the coupon rate. Equating Equation 17.13 with the residual debt at time T yields a condition for the forward par rate:

$$\underbrace{B_{K(T,T'|t)}(T_0,T)^{-1}}_{\text{Residual Debt}} = \sum_{i=1}^{n} B_R(T,T_i|t)\left[B_{K(T,T'|t)}(T_{i-1},T_i)^{-1}-1\right] + B_R(T,T'|t) \tag{17.15}$$

By solving Equation 17.15 (using numerical iteration, for example) we obtain the forward par rate as a function of the forward (zero) rates as of the future date T. Multiplying both sides by $B_R(t,T)$ and exploiting Equation 3.4 yields:

$$B_R(t,T)\underbrace{B_{K(T,T'|t)}(T_0,T)^{-1}}_{\text{Residual Debt}} = \sum_{i=1}^{n} B_R(t,T_i)\left[B_{K(T,T'|t)}(T_{i-1},T_i)^{-1}-1\right] + B_R(t,T') \tag{17.16}$$

By solving this equation (again by numerical iteration) we obtain the forward par rate as a function of the current *spot* rates. Equation 17.16 means that the forward par rate can be interpreted as the fixed rate of a forward bond whose present value (at time t) is equal to the discounted residual debt the bond will have at time T. It follows immediately that the forward par rate converges to the current par rate as $t \to T$.

Equation 17.15 is quite similar to Equation 17.7 for the forward swap rate, with the exception that in the time between T_0 and T the floater in Equation 17.7 generates the compounding factor with respect to the forward (*zero*) rate, while the residual debt in Equation 17.15 generates a compounding factor with respect to the forward *par* rate. For $T = T_0$, both of these factors equal one and thus Equations 17.7 and 17.15 are the same. This means that:

If the maturity of the forward contract T falls at the beginning of an interest period, the forward swap rate equals the forward par rate of the bond on the fixed side of the swap.

Probably all the transactions taking place in practice are concluded with $T = T_0$, i.e. the maturity of the forward contract falls at the start of the first interest period of the forward bond, resp. the forward swap. The forward par rate and the forward swap rate are then equal. Consequently, if the interest periods are all of equal length, Equation 17.9 for the forward par rate can be used. If T is not equal to T_0, however, either Equation 17.15 or Equation 17.16 must be solved numerically for the forward par rate.

The YTM of the forward bond (Forward YTM) is the constant rate which, when substituted into the forward discount factors in Equation 17.13, yields the present value of the forward bond as of the future time T. Thus, with Equation 17.13, we obtain a general condition for the forward YTM

$$\begin{aligned}
&\sum_{i=1}^{n} B_{\overline{R}}(T, T_i|t)[B_K(T_{i-1}, T_i)^{-1} - 1] + B_{\overline{R}}(T, T'|t) \\
&= \sum_{i=1}^{n} B_R(T, T_i|t)[B_K(T_{i-1}, T_i)^{-1} - 1] + B_R(T, T'|t) \\
&= B_R(t, T)^{-1} \left[\sum_{i=1}^{n} B_R(t, T_i)[B_K(T_{i-1}, T_i)^{-1} - 1] + B_R(T, T_i') \right] \qquad (17.17)
\end{aligned}$$

The first equation gives the forward YTM as a function of the forward (zero) rates valid as of time T, the second as a function of the current spot rates. A similar relation between the *forward* par rate and the *forward* YTM holds as for the corresponding spot yields, namely that the YTM of a forward *par* bond is equal to the forward par rate. Setting $T = T_0$ here as well as in the corresponding Equation 17.11 for forward swaps, we obtain the exact same condition for the forward YTM of a forward swap, i.e.

If the maturity T of a forward contract falls on the start of an interest period, the forward YTM of a swap equals the forward YTM of the bond on the fixed side of the swap.

The general equation given in 17.17 is rather complicated but becomes much simpler in most concrete situations. For example, the general form reduces to Equation 17.12 if the interest periods are of equal length and if the bond is entered into at the beginning of a complete period, i.e. for $T = T_0$.

18

Plain Vanilla Options

In this chapter, the pricing of options on different underlying types using the Black-Scholes model and the Black-76 model is presented. The values listed in Figure 18.1 will repeatedly be used in the following examples.

For simplicity's sake, we adopt the *Act/365* day count convention, for which $T - t$ is equal to the number of calendar days between T and t divided by 365. Furthermore, the examples are calculated using the continuous compounding convention, with the following implication holding for the discount factor $B_r(t, T)$:

$$B_r(t,T) = e^{-r(T-t)} \;\Rightarrow$$
$$\frac{\partial B_r(t,T)}{\partial t} = r\, B_r(t,T)\;, \quad \frac{\partial B_r(t,T)}{\partial r} = -(T-t)B_r(t,T)$$

We could have likewise used annual compounding, for example. The discount factor and its derivatives in this case have the following form:

$$B_r(t,T) = \frac{1}{(1+r)^{(T-t)}} \;\Rightarrow$$
$$\frac{\partial B_r(t,T)}{\partial\, t} = \ln(1+r)B_r(t,T)\;, \quad \frac{\partial B_r(t,T)}{\partial r} = -\frac{T-t}{1+r}\, B_r(t,T)$$

Annualized Refinancing Rate	*r*	3.20%
Annualized Dividend Yield	*q*	4.80%
Annualized Volatility	*σ*	12.50%
Spot Price	*S(t)*	$80.00
Value Date	*t*	Jan-07-01
Expiration Date of Forward/Future	*T'*	Apr-07-01
Expiration Date of Option	*T*	Mrz-08-01
Strike Price of Option	*K*	$79.00

Figure 18.1: The input parameters for all of the following examples.

Interim Results			
$T-t$	0.16438	$T'-t$	0.2466
$Br(t,T)$	0.99475	$Br(t,T')$	0.9921
$Bq(t,T)$	0.99214	$Bq(t,T')$	0.9882
$S(t,T)$	$79.79	$S(t,T')$	$79.69
$\sigma\sqrt{(T-t)}$	0.05068	$\sigma\sqrt{(T'-t)}$	0.0621
x	0.22164	x'	0.1701
$N(x)$	0.5877	$N(x')$	0.5675
Value of Future	$0.685		
Value of Forward	$0.680		

	Option on Spot		Option on Future	
	Call	Put	Call	Put
Price	$2.020	$1.234	$1.959	$1.277
Delta	0.5831	-0.4091	0.5745	-0.4202
Gamma	0.0953	0.0953	0.0964	0.0964
Omega	23.0905	-26.5094	23.3657	-26.2009
Vega	0.1253	0.1253	0.1258	0.1258
Theta	-0.0108	-0.0144	-0.0129	-0.0130
Rho	0.0734	-0.0558	0.0720	-0.0571
Rho_q	-0.0767	0.0538	-0.0753	0.0550

Figure 18.2: Valuation of European plain vanilla spot and futures options using the Black-Scholes model.

All of the following examples have been extracted from the Excel workbook PLAIN-VANILLA.XLS on the CD-ROM accompanying this book.

18.1 Options on Spot and Forward Prices

18.1.1 European Options

European plain vanilla options are usually priced on the market with the Black-Scholes model. In doing so, Equations 9.7 for spot options and 18.4.2 for options on futures are applied. This is demonstrated in Figure 18.2. The forward prices themselves are determined using the arbitrage Equation 7.1. They are also presented in Figure 18.2 among the "Interim Results". The values of the futures and forward positions, each with a delivery price K were calculated from the forward price $S(t,T')$ using Equations 7.4 and 7.5. The sensitivities for options on the spots and futures as given in Table 13.2 are included in Figure 18.2 as well. The sensitivities with respect to variables commonly given in %, i.e. vega, rho, and rho_q are divided by one hundred. This is common practice since the sensitivities should indicate the magnitude of a price change if the corresponding parameter changes by 1% (and not by 1=100%). For similar reasons, Theta was divided by 365 since it then indicates the price change resulting from the passing of a day (and not a year).

The price of the futures option was calculated using Equation 18.4.2. In order to calculate the sensitivities, an additional procedure was implemented: as mentioned in the vicinity of Equations 18.4.2 and 13.10, an option on a future is the same as an option on the spot price of an underlying whose dividend yield is exactly the risk-free rate ($q = r$), and whose spot price corresponds to the forward price $S(t,T')$. Therefore, the same procedure as used for the spot option was applied again for the futures option but with the appropriately modified input parameters.

The gamma and delta of a spot option refer to the spot price of the underlying. In the case of the futures option, on the other hand, the forward price is the reference price. This means, for example, that the delta of the futures option indicates how

		Interim Results		Option Price	
dt=(T-t)/n	0.001644	u	1.005081	Call	$2.02
Br	0.999947	d	0.994945	Put	$1.24
Bq	0.999921	p	0.496138		
T-t	0.164384	p'	0.498672		
Br(t,T)	0.994754	y	49		
Bq(t,T)	0.992141	Bp(j>=y)	0.587978		
		Bp'(j>=y)	0.60763		

Figure 18.3: Vaulation of European plain vanilla options using the binomial distribution.

many *futures* are required to (delta) hedge the option, whereas the delta of the spot option gives the number of *underlyings* required for the hedge. The required number of instruments for other combinations can be calculated using Equation 13.16 and Table 13.1 with either of these deltas. For example, to hedge the spot option requires either 0.5831 underlyings or $0.5831 B_r(t,T')/B_q(t,T') = 0.5854$ futures.

The price of options on the corresponding forward contract can be calculated with Equation 9.9. It yields a price of \$1.954 for the call and \$1.274 for the put. Options on forward contracts are not traded nearly as much as options on futures.

Of course, the binomial model can be applied for the valuation as well. The appropriate payoff profile is to be implemented at the end of the tree but otherwise the procedure for options on a spot price, a future or a forward is always the same. For options on the spot price, the price can be determined directly from Equation 11.19 using the binomial distributions. The parameters u, d are calculated using Equation 11.31. The values for the probabilities p and $\hat{p}$ follow directly from Equation 11.9. The results of the individual steps are displayed in Figure 18.3. The price of the put was determined directly from the price of the call with the help of the put-call parity, Equation 7.7.

18.1.2 American Options

The binomial model is very effective in calculating American options. The following example will serve to demonstrate the valuation of both options and futures. The values for the futures and forward positions, each with a delivery price K listed among the final results, are computed from the forward price $S(t,T')$ making use of Equations 7.4 and 7.5.

First, a tree for the underlying price must be constructed. In Figure 18.4, we use the method prescribed by Equation 11.31 for the calculation of the up and down parameters, in the same way as was already done in Figure 18.3. The calculation is subsequently performed backwards through the entire tree:

- Time $t + 3dt = T'$
  Once the tree for the underlying has been constructed, the values of the derivative on the boundary of the tree should be inserted if these are known. In our example,

these are the forward prices at the maturity of the future T'. These are just the spot prices effective at time T'.

- Time $t + 2dt = T$
  This time is a maturity date as well, but for the options. Thus, at this point the prices of the *spot option* at maturity are set equal to its payment profile. The *future prices* at each node are determined using Equation 11.13 from the prices in the next time step. Once the future prices at the maturity of the option on the future are known, the price of the *futures option* can be set equal to its payoff profile.

- Time $t + dt$
  The *future prices* at each node are computed using Equation 11.13 from the prices holding for the next time step. The prices of both the *spot* as well as the *futures* option at each node are, in accordance with Equation 11.8, computed from the prices in the next time step. The right to early exercise is taken into account by comparing the results just obtained to the intrinsic values $S - K$ and $S' - K$ (S' denotes the forward price in this example) taking in each case the greater of the two values as the corresponding option price.

- Times $t, t - dt, t - 2dt$
  The same procedure is carried out as that described for time $t + dt$.

Once the entire tree is constructed, prices at three different nodes are available for use at time t, from which the middle node (the **du** node) contains the desired prices for the future, the option on the spot price and the option on the future. The values at both the other nodes at time t (**uu** and **dd**) will be needed for the calculation of the sensitivities using Equation 13.28. The values for delta, gamma and omega of both options listed here refer to the spot price (not to the future). The calculation of vega requires the entire tree to be recalculated for a slightly different volatility (for instance, 13.5%), taking the difference of the two prices thus obtained and dividing this value by the difference in the volatility. Similarly, rho is obtained by recalculating the tree for a somewhat modified interest rate. Only the prices at time t are necessary to perform these recalculations. The tree may thus start at t instead of $t - 2dt$.

18.2 Index Options and Index Futures

Index futures and options on an index or on an index future correspond precisely to the general case treated in Section 16, where the spot price S represents the state of the index and q the average dividend yield of the index's component stocks (*as long as these are not directly reinvested in the index!*). Examples are the (European) options on the DAX index traded on the EUREX, known as *ODAX* and the (American) Options on DAX-futures, known as *OFDX*. Figures 18.3 and 18.4 again serve as examples of the valuation of index futures, index spot options and index futures options, where the dividend yield q is the average dividend yield of the index and the underlying price S is the state of the index.

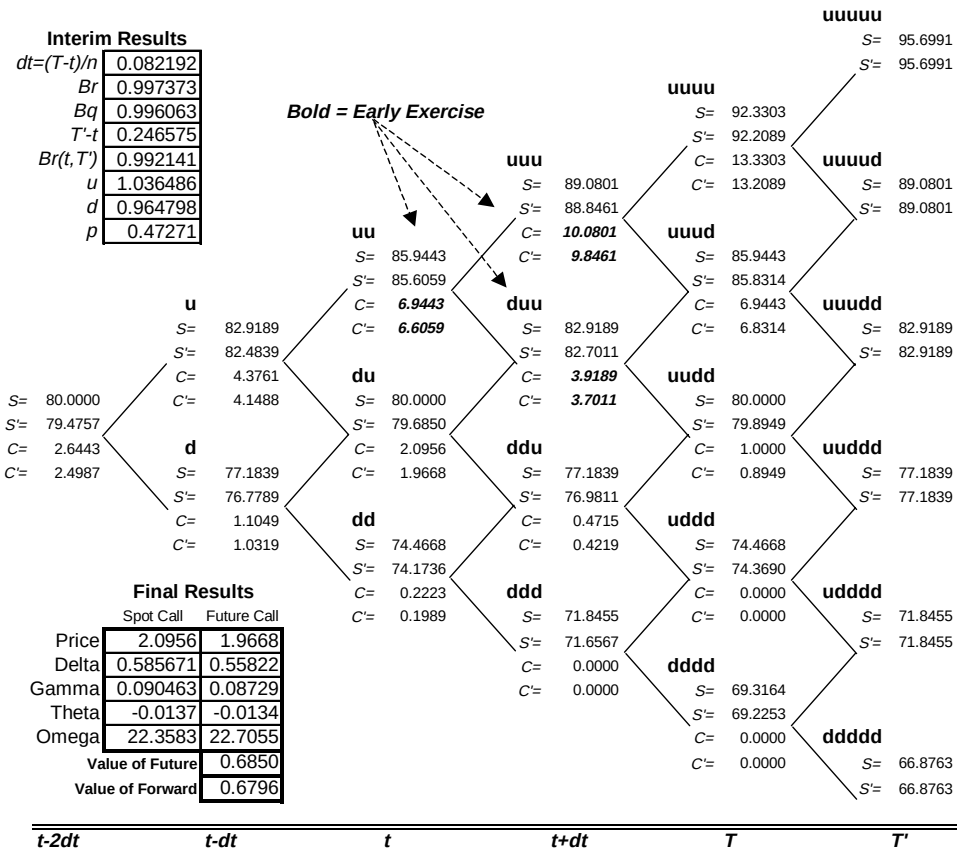


Figure 18.4: Valuation of an American call c on the spot price S, and of an American call c' on the forward price S' using a binomial tree. The tree for the options comprises two time steps into the future. The tree for the forward price consists of three time steps into the future.

The German *DAX* is a *performance index*, which means that the dividends of the component stocks are immediately reinvested in the DAX, so that in principle, no decline in price is observed when a dividend payment is made and consequently, q should be set equal to zero. The reason why q is not exactly zero in practice relates to the effects of corporate taxation. If the risk-free rate r is known, the DAX together with the DAX future can be used in Equation 7.2 to find the *implied dividend yield* which can then be used as q when calculating option prices.

18.3 Foreign Exchange Options and Futures

Foreign exchange futures ("*FX futures*") and foreign exchange options ("*FX options*") correspond exactly to those instruments treated in the previous section. Here, the *exchange rate* plays the role of the spot price S, and the risk-free rate r_f in the foreign currency under consideration plays the role of the dividend yield q. The logic behind this choice is the following: the holder of the underlying (i.e. the foreign currency) receives, in contrast to the holder of the option or the future, interest payments at the interest rate valid for the foreign currency much the same as the holder of a stock receives a dividend yield q of the stock. Again Figures 18.3 and 18.4 serve as examples for the valuation of foreign exchange futures, spot options and futures options, where here the dividend yield q is equal to the risk-free rate of the foreign currency and the underlying price S is merely the foreign exchange rate expressed in terms of the domestic currency. At an exchange rate of 1.15 euros to the US$, for example, the underlying price would be $S(t) =$1.15.

18.3.1 Put-Call-Equivalence for FX Options

An interesting relation exists between foreign exchange puts and calls, which follows immediately from the nature of these transactions: The right to *buy* 1 USD for K EUR is obviously the same as the right to *sell* K EUR for 1 USD. The EUR/USD exchange rate indicates how many euros have to be paid per dollar. The USD/EUR exchange rate, indicating how many dollars are to be paid per euro, is naturally $1/S(t)$. A *call* on the EUR/USD exchange rate with a strike price K is thus the same as K *puts* on the USD/EUR exchange rate with a strike price $1/K$. This can also be seen by considering the payoff profile. The payoff in dollars of K puts on the euro is

$$K \max\left\{0, \frac{1}{K} - \frac{1}{S(T)}\right\} = \max\left\{0, 1 - \frac{K}{S(T)}\right\} = \frac{1}{S(T)} \max\left\{0, S(T) - K\right\}$$

Here, all terms are expressed in USD units of currency since dollars are received for selling euros. The payoff of a *call* on USD, however, is more naturally expressed in terms of EUR since euros are payed when buying dollars. Before the payoffs can be compared, they both need to be expressed in terms of the same currency. The conversion of the above payoff into EUR can be performed by simply multiplying with the EUR/USD exchange rate $S(T)$. This simply yields the payoff profile of the call $\max\{0, S(T) - K\}$ on the right-hand side. If the payoff profiles agree, the options must have the same value for all earlier times t. We thus obtain the *put-call equivalence for foreign currency options*

$$c_S(t, T, K) = K \; p_{1/S}(t, T, 1/K) \tag{18.1}$$

18.3.2 FX Forward Contracts and FX Swap Rates

As far as the valuation of forward and option contracts is concerned, a currency exchange rate corresponds to an underlying with a dividend yield q, where q corresponds

to the risk-free rate r_f in the foreign currency. According to Equation 7.1 the *forward exchange rate*, i.e. the forward price of the foreign currency, is

$$S(t,T) = \frac{B_{r_f}(t,T)}{B_r(t,T)} S(t) \tag{18.2}$$

We once again present here the fundamental arbitrage argument for the special case of foreign currency transactions: An amount of N euros is used to purchase $N/S(t)$ dollars for the spot exchange rate given by $S(t)$ EUR/USD. This amount is invested in the US money market over a period of time of length $(T-t)$ at a rate r_f. Afterwards, the entire sum is changed back to euros at the exchange rate $S(T)$ EUR/USD valid at time T. The value of this capital in euros at the end of these transactions is

$$\underbrace{S(T)}_{\text{Exchange into Euros}} \qquad \underbrace{1/B_{r_f}(t,T)}_{\text{Compounding in USD}} \qquad \underbrace{N/S(t)}_{\text{Amount in USD}}$$

On the other hand, if this capital N is invested directly in the euro money market for a period of time $(T-t)$ at the rate r, it will have earned $N/B_r(t,T)$.

If no arbitrage is possible between the spot and forward market with respect to the foreign currency exchange rate, both strategies must yield the same amount *if* the exchange rate at the future date T is already fixed at time t in a foreign exchange forward contract. The forward exchange rate $S(t,T)$ for this transaction is the value for $S(T)$ established by an arbitrage-free market as seen from today. Equating both strategies immediately yields Equation 18.2.

The *basis*, defined in general as the difference between the forward price and the spot price, is referred to as a *swap rate* by FX traders because for FX transactions it results solely from the difference in the interest rates of the two currencies concerned. This swap rate $b(t,T)$ in terms of the compounding methods common to both the money and foreign exchange market, namely simple and linear compounding (see Table 3.4), is:

$$\begin{aligned} b(t,T) &\equiv S(t,T) - S(t) \\ &= \left[\frac{B_{r_f}(t,T)}{B_r(t,T)} - 1\right] S(t) \qquad \text{general} \\ &= \left[\frac{1+r(T-t)}{1+r_f(T-t)} - 1\right] S(t) \quad \text{simple compounding} \\ &= (r-r_f)(T-t)S(t) \qquad \text{linear compounding} \end{aligned} \tag{18.3}$$

For linear compounding, the swap rate is simply the spot exchange rate multiplied by the interest spread of the two currencies involved and by the length of the time period under consideration.

18.4 Interest Rate Options

In Part II the valuation of interest rate options by means of term structure models was demonstrated. In the presentation of the following examples, however, plain vanilla

interest rate options are priced using the Black-Scholes and Black-76 models still commonly used in the market for the valuation of these types of options.

18.4.1 Options on Bonds

Europeans bond options are usually priced using the Black-76 model, Equation 9.10. The parameter σ of this model is the forward volatility of the bond price under consideration. Under the assumption of the Black-76 model (forward price equals future price), this can be replaced by the volatility of the *forward* price of the underlying.

For options on zero bonds, $S(t,T)$ is the forward price as given by Equation 17.2, for options on coupon bonds, it is the (dirty) forward price according to Equation 17.3 or the present value of the forward bond at time T, Equation 17.13. As presented in Section 17.2.2, calculations made in pricing derivatives on coupon bonds should always be performed with the amounts of money which must actually be paid, i.e. accrued interest should always be included in the computations. Thus, for options on coupon bonds, the *dirty strike* of the option should be used. This is the amount of money paid by the holder of a call upon exercise in order to purchase the underlying. In general, bond options are quoted for a *clean strike*. The interest accrued up to time T must then be added to this quoted strike.

American options on bonds – as all options on underlyings with dividend payments during the lifetime of the option – generate binomial trees which *no longer* recombine after the *ex-dividend day* t_D (=coupon payment date):

$$u\,dS(t) \neq d\,u\,S(t) \quad \text{for} \quad t > t_D$$

However, the construction of the binomial tree and the backward computation of the option prices starting with the end of the tree are exactly as described in Figure 18.4 for options on underlyings without dividends. But twice as many nodes appear in each period after the ex-dividend day as would for a recombinant tree. This results in a drastic increase in computing time.

Since the coupon payment dates for the valuation of both European and American bond options must be known exactly, bond options always refer to a real bond rather than a synthetic one as was the case with futures, for example. In consequence, the markets for such instruments are much less liquid than the corresponding options on bond *futures*.

This together with the computing time required for non-recombinant binomial trees for American options is the reason why options on bonds are a good deal rarer in electronic exchanges than futures options. For example, on the EUREX, options are available on the Bund Future, the Bobl future and the Schatz future. These are all American. However, no options directly on 2, 5 or 10 year government bonds are traded on the EUREX.

18.4.2 Options on Bond Futures

European options on bond futures are priced – as are all plain vanilla European futures options on the market – using the Black-Scholes formula for options on futures given by Equation .

The following consequence of the Black-Scholes differential equations 13.5 and 13.10 are useful in the valuation of *American* options on bond futures: an option on a future is equivalent to an option on an underlying with a dividend yield equal to the risk-free rate and a spot price equal to the future's price. For such an underlying, the (recombinant) binomial tree can be constructed as described in Section 11 to price the option. This has been explicitly demonstrated in Figure 18.4 in Section 16 (the part of the example for the valuation of the option on the underlying itself is relevant to our discussion here). Thus, Figure 18.4 can serve as an explicit demonstration of pricing bond futures options if we use the clean forward bond price in place of the "spot price" of the underlying and the risk-free rate for the dividend yield. In contrast to derivatives referring *directly* to a bond (bond options, bond futures), *clean* prices have to be used for options on bond *futures* since the difference between the clean forward price and the clean strike is due when the option is exercised.

18.4.3 Caps and Floors

A *cap* (*floor*) is a series of *caplets* (*floorlets*), arranged in such a way that the periods of the individual caplets (floorlets) completely cover the time span of interest without overlap. The period length of the individual caplets (floorlets) corresponds to the lifetime of the floating interest to be bounded. A cap, for example, of a 3-month LIBOR for five years consists of a series of 19 (the first period is already fixed) caplets each having a period length of three months. To understand caps and floors it is sufficient to understand their component caplets and floorlets.

A *caplet* is an "interest rate limiting agreement" over a time span from the fixing date T until the payment date T'. The principal N of a caplet is compounded over the future time span $T' - T$ at a rate equaling the difference between the (floating) reference rate $R = R(T, T')$ and a fixed strike K *if* this rate difference is positive. The floating reference rate is fixed at the fixing date T. This construction has the effect that the holder of a caplet pays interest (for instance for a loan) on a principal N at a floating rate over a time span from T until T' but at *most* the strike rate K of the caplet.

A caplet is thus an option with the reference rate as the underlying. The option is *exercised* at the fixing date T since it is at this time that the difference between the strike and the underlying is determined. The option *payment* occurs later, namely at the payment date T'. The value of the option upon *exercise* is of course essential to its valuation. In order to obtain the value of the option upon exercise, its payoff profile must be discounted back to the exercise time T:

$$
\begin{aligned}
c_R^{\text{cap}}(T, T', K|T) &= \begin{cases} B_R(T,T)N\left[B_{R-K}(T,T')^{-1} - 1\right] & , \quad R(T,T') > K \\ 0 & , \quad R(T,T') \leq K \end{cases} \\
p_R^{\text{cap}}(T, T', K|T) &= \begin{cases} B_R(T,T')N\left[B_{K-R}(T,T')^{-1} - 1\right] & , \quad R(T,T') < K \\ 0 & , \quad R(T,T') \geq K \end{cases}
\end{aligned}
\tag{18.4}
$$

From the condition that a payment is made if the strike of the caplet is greater than the underlying, it follows that a caplet is a *call* option on the reference rate. It is

for this reason that the letter c was chosen to denote the caplet in Equation 18.4. The corresponding *put* on the reference rate is the above mentioned *floorlet*, which represents a lower limit ("floor") on the reference rate. The floorlet is denoted by the letter p in Equation 18.4. Caps and floors are also referred to as *interest rate guarantees (IRGs)*.

Since the discount factors are strictly monotone decreasing functions of the interest rate, conditions on the interest rates can be equivalently interpreted as conditions on the discount factors, i.e.

$$R(T,T) > K \Leftrightarrow B_R(T,T) < B_K(T,T), \text{ etc.} \tag{18.5}$$

Equation 18.4 holds for all compounding conventions. For the caps and floors on the market, simple compounding (common in the money market) is used almost exclusively for compounding over the cap period. This is because the time periods under consideration are usually of short duration, for example three months or six months. Thus

$$\begin{aligned} B_{R-K}(T,T') &= \left[1 + (R-K)(T'-T)\right]^{-1} \Rightarrow \\ B_{R-K}(T,T')^{-1} - 1 &= \left[R(T,T') - K\right](T'-T) \end{aligned} \tag{18.6}$$

similarly for B_{K-R}. To clarify the dependencies, the time dependence of the floating rate has been included in notation in the last line: $R(T,T')$ is the interest rate valid at time T for the period running from T until T'. Hence, the option values (the amount paid at time T' discounted over the cap period) at exercise time T in Equation 18.4 become

$$\begin{aligned} c_R^{\text{cap}}(T,T',K|T) &= B_R(T,T')N(T'-T)\max\left\{R(T,T') - K, 0\right\} \\ p_R^{\text{cap}}(T,T',K|T) &= B_R(T,T')N(T'-T)\max\left\{K - R(T,T'), 0\right\} \end{aligned} \tag{18.7}$$

By definition, the option value at exercise time is the payoff profile of the option. Thus, Equation 18.7 displays the payoff profiles of caplets and floorlets.

Valuation as Interest Rate Options (Lognormally Distributed Interest Rates)

These payoff profiles allow the option to be interpreted as a plain vanilla option with a nominal $N(T'-T)B_R(T,T')$, a strike price K and maturity T on an underlying with a price $S(T) = R(T,T')$. The forward price of this underlying as seen from today (time t) is then the forward rate $R(T,T'|t)$. The discount factor $B_R(T,T')$ appearing in the nominal is not known at time t. Market participant use the discount factor of the forward rate $B_R(T,T'|t)$ instead. It can be shown [126], that this is consistent with a term structure model. The maximum functions in Equation 18.7 correspond to payoff profiles of plain vanilla options and can therefore be priced with the Black-76 model,

Equation 9.10. The prices of caps and floors are then given by

$$
\begin{aligned}
&c_R^{\text{cap}}(T,T',K|t)\\
&= N(T'-T)\underbrace{B_R(T,T'|t)B_R(t,T)}_{B_R(t,T')}\Big[R(T,T'|t)\mathrm{N}(x) - K\mathrm{N}(x-\sigma\sqrt{T-t})\Big]\\
&p_R^{\text{cap}}(T,T',K|t)\\
&= N(T'-T)\underbrace{B_R(T,T'|t)B_R(t,T)}_{B_R(t,T')}\Big[R(T,T'|t)\mathrm{N}(-x) - K\mathrm{N}(-x+\sigma\sqrt{T-t})\Big]\\
&x = \frac{\ln\left(\frac{R(T,T'|t)}{K}\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}
\end{aligned}
\tag{18.8}
$$

According to Equation 3.4, the product of the discount factor and the forward discount factor yield precisely the discount factor from time T' back to time t.

As mentioned at the beginning of this section, a cap is nothing other than a *strip* of caplets. The value of the cap is the sum of the values of the individual caplets (with the analogous relation holding between floorlets and floors):

$$
\begin{aligned}
c_R^{\text{cap}}(T,T',K|t) &= \sum_{i=1}^{n} c_R^{\text{cap}}(T_{i-1},T_i,K|t) \qquad (18.9)\\
p_R^{\text{cap}}(T,T',K|t) &= \sum_{i=1}^{n} p_R^{\text{cap}}(T_{i-1},T_i,K|t)\\
&\text{where } T = T_0 < T_1 < T_2 < \ \ldots\ < T_n = T'
\end{aligned}
$$

As illustrated in Equation 9.10, the Black-76 model assumes that the underlying is lognormally distributed at the option's exercise time. The volatility *at that time* is given by a value σ. This means we have to assume that the reference rate $R(T_{i-1},T_i)$ of each caplet is lognormally distributed. For σ we have to take the *forward* volatility as of the fixing date T_{i-1} of the reference rate R. This forward volatility is different for each T_{i-1}. In consequence, as many volatilities are needed in pricing a cap as there are caplets in the cap. We sometimes speak of *forward-forward volatilities* in the context of this procedure.

A simpler procedure is often utilized in the market in which only *one* volatility is used for *all* caplets. These cap volatilities differ from one another depending on the lifetime of the cap. They are also referred to as *flat volatilities*. An obvious inconsistency in the flat volatilities is that the same caplet, appearing simultaneously in the strips of two different caps, can have different volatilities and thus have two different prices. For example, a caplet that limits the 6-month LIBOR to 8% in two years occurs in both a five year and a ten year, 8% cap on the 6-month LIBOR. Since the flat volatilities differ for 5 and 10-year caps, this caplet will have two different prices corresponding to the two different caps.

The valuation of caps and floors described above, i.e. the interpretation as options on the interest rate as the underlying, using of the Black-76 model naturally assumes we

are in the Black-Scholes world. In particular that the yield of the underlying (=relative change in the interest rate) is normally distributed, which implies that the underlying itself (the interest rate) is *log*normally distributed.

Valuation as Bond Options (Normally Distributed Interest Rates)

It is also possible to price caps and floors by interpreting them as options on *bonds* and performing the valuation using Black-Scholes for options on bond prices. Then the underlying is the bond price which is therefore implicitly assumed to be lognormally distributed. Thus, its relative changes (the bond *yields* which are by definition nothing other than interest rates) are assumed to be *normally* distributed.

In terms of the compounding convention most commonly used in the market, i.e. simple compounding, Equation 18.6 holds and thus $B_R(T,T') = 1/[1 + R(T' - T)]$, the corresponding form holding for $B_K(T,T')$ as well. Thus the payoff profiles in Equation 18.4 can be expressed not only as the difference between two interest rates (as in Equation 18.7) but as the difference between two *discount factors* as well. We show this here for a caplet (call):

$$\begin{aligned} B_R(T,T')\left[\frac{1}{B_{R-K}(T,T')} - 1\right] &= \frac{(R-K)(T'-T)}{1+R(T'-T)} \\ &= 1 + \frac{-1 - R(T'-T) + (R-K)(T'-T)}{1+R(T'-T)} \\ &= 1 - \frac{1+K(T'-T)}{1+R(T'-T)} \\ &= [1+K(T'-T)]\left[\frac{1}{1+K(T'-T)} - \frac{1}{1+R(T'-T)}\right] \\ &= [1+K(T'-T)]\,[B_K(T,T') - B_R(T,T')] \end{aligned}$$

Transforming the condition on the interest rates into conditions on the discount factors as in Equation 18.5, we can write Equation 18.4 (for simple compounding) using maximum functions, and the option prices as of the fixing date T are

$$\begin{aligned} c_R^{\text{cap}}(T,T',K|T) &= N\,[1+K(T'-T)]\max\{B_K(T,T') - B_R(T,T'), 0\} \\ p_R^{\text{cap}}(T,T',K|T) &= N\,[1+K(T'-T)]\max\{B_R(T,T') - B_K(T,T'), 0\} \end{aligned} \tag{18.10}$$

The factor $B_K(T,T')$ involves no unknown variables since the strike rate K and the fixing and payment dates are all specified exactly in the option contract. It is thus a constant factor known at all times t. The factor $B_R(T,T')$ is the value (with simple compounding) at the fixing time T of a zero bond maturing at payoff date T' with a face value of one monetary unit. The interpretation of the payoff profile with simple compounding is thus:

- The *caplet* is equivalent to $N\,[1+K(T'-T)]$ *puts* maturing at the fixing date T with strike $B_K(T,T')$ on zero bonds maturing at the payoff date T'. Each of these zero bonds has a face value of one monetary unit.

Parameter of the Bond Option	Corresponding Parameter for IRGs
Option	$N[1 + K(T' - T)]$ puts for caplet, exactly as many calls for floorlet
Underlying	zerobond with maturity at the payment date $T'$of the caplet
Face Value of the Underlying	one monetary unit
Forward Price $S(t, T)$	forward price of the zero bond at the fixing date T of the caplet
Volatility	volatility of the forward prices
Strike of the Bond Option	discount factor $B_K(T, T')$ over the reference period with interest rate K

Table 18.1: Caplets and floorlets as options on zero bonds.

- The *floorlet* is equivalent to $N\,[1 + K(T' - T)]$ *calls* maturing at the fixing date T with strike $B_K(T, T')$ on zero bonds maturing at the payoff date T'. Each of these zero bonds has a face value of one monetary unit.

Interpreted in this way, caplets and floorlets are equivalent to options on zero bonds and can be priced as described in Section 18.4.1. A cap and floor as a portfolio of caplets and floorlets, respectively, can thus be expressed as a portfolio consisting of puts and calls on the corresponding zero bonds. The relation to bond options priced using Equation 9.10 is summarized in Table 18.1.

Collars and Put-Call Parity for Caps and Floors

A collar is a combination of a cap and a floor having the effect that the collar's holder (if he or she also has a loan, for instance) has to pay the reference rate (for the loan) only as long as it remains within a range bounded below by the strike of the floor and above by the strike of the cap. Should the rate exit this range, the payment that the holder of the collar would receive from the cap (or pay into the floor) would compensate for the higher (or lower) interest payment on the loan. A long collar is thus a portfolio consisting of a long cap and a short floor:

$$c_R^{\text{collar}}(T, T', K_1, K_2|t) = c_R^{\text{cap}}(T, T', K_1|t) - p_R^{\text{cap}}(T, T', K_2|t)$$

Investors often buy collars to provide a safeguard against rising interest rates and to lower the price of this safeguard through the sale of the floor. Often, if the strike of the cap is given, the strike of the floor is chosen such that the floor is worth exactly as much as the cap and that thus the total collar can be acquired without cost (called *zero cost collar*).

If we require instead the strikes of the cap and the floor to be the same, i.e. $K_1 = K_2 = K$, then the holder of the collar effectively pays (for a loan, for example) exactly the strike K instead of the reference rate. This is the same as though he had a forward swap from T to T'. Thus, caps and floors satisfy the following put-call parity.

Interest Rates						Interest Rate Options			Principal 250,000		
Spot			Forward 1-Year Term			Cap 7.00%		Floor 3.24%		Collar	
Time to Maturity	Spot Rate	Zero Bond	Starting in Years	Forward Rate	Forward Vol σ	Caplets in %	Caplets in $	Floorlets in %	Floorlets in $	in %	in $
1	2.20%	0.97847	0	2.20%	18.00%						
2	2.50%	0.95181	1	2.80%	18.20%						
3	2.90%	0.91781	2	3.70%	18.40%						
4	3.30%	0.87821	3	4.51%	18.60%	0.062%	156	0.725%	1,813	-0.663%	-1,657
5	3.70%	0.83389	4	5.32%	18.80%	0.256%	640	0.463%	1,158	-0.207%	-518
6	4.00%	0.79031	5	5.51%	19.00%	0.370%	924	0.515%	1,288	-0.146%	-364
7	4.30%	0.74475	6	6.12%	19.20%	0.616%	1,540	0.433%	1,082	0.183%	458
8	4.50%	0.70319	7	5.91%	19.40%	0.585%	1,462	0.569%	1,422	0.016%	40
9	4.70%	0.66142	8	6.31%	19.60%	0.751%	1,877	0.531%	1,328	0.220%	549
10	4.90%	0.61979	9	6.72%	19.80%	0.907%	2,269	0.499%	1,248	0.408%	1,021
11	5.00%	0.58468	10	6.01%	20.00%	0.679%	1,698	0.699%	1,746	-0.019%	-48
12	5.10%	0.55051	11	6.21%	20.20%	0.750%	1,875	0.685%	1,712	0.065%	163
13	5.20%	0.51736	12	6.41%	20.40%	0.812%	2,029	0.669%	1,673	0.143%	357
14	5.25%	0.48853	13	5.90%	20.60%						
15	5.30%	0.46087	14	6.00%	20.80%						
Total Value Cap, Floor, Collar:						**5.79%**	**14,470**	**5.79%**	**14,470**	**0.00%**	**0**

Figure 18.5: Interest rate options on the 12-month floating rate. The values in dollars result from the values in % through multiplication by the nominal of $250.000.

If the cap and the floor of a collar have the same strike, the value of the collar is equal to the value of a forward swap over the lifetime of the collar with a fixed rate equal to the strike of the collar.

Or in brief

Cap minus floor equals forward swap

By definition, the forward swap rate is the coupon rate for the fixed side which makes the swap worthless. The put-call parity now states:

If a collar is worthless and if its cap and floor both have the same strike, the strike of the collar equals the forward swap rate for the collar's lifetime.

In Figure 18.5, a cap, a floor and a collar with the same lifetime as the forward swap in Figure 17.1 (thus, over ten years, starting in three years) are evaluated according to Equations. 18.8 and 18.9 using the same term structures as in Figure 17.1. The forward rates have already been calculated and are shown in Figure 3.6. The strike of the floor (3.24%) was (through numerical iteration) modified until the collar for a cap of 7% was worthless. Setting the floor strike equal to the cap strike and solving for the strike price again, to make the collar worthless does in fact yield a forward swap rate of 5.28% as seen in Figure 17.2.

18.4.4 Swaptions

A *swaption* or swap option is the right to enter into a swap upon maturity of the option, without incurring any further payment. The interest rate K on the fixed side of this underlying swap is the strike price of the option. Depending on whether the swap is a payer or receiver swap, the swaption is referred to as a *payer* or *receiver* swaption, respectively. As opposed to caps and floors, which can be interpreted as options on the floating side of a swap, swaptions are options on the *fixed* side of the swap. The lifetime of the swap begins at the maturity T of the option. Hence, the swap under consideration is a forward swap. During the lifetime of a forward swap up to its maturity T', a number n of payment dates occur at times T_i, where T_0 is defined as the last fixing date of the forward swap before the maturity of the option:

$$t < T_0 \leq T < T_i < T_{i+1} < T_n = T \;\; \forall\, i = 1, \ldots, n-1$$

Usually $T_0 = T$ of course. The payoff profile of a swaption, i.e. its value at maturity T can be determined by considering the following. If we enter into an underlying swap at time T as a consequence of exercising a receiver swaption, then – at each payment date for the entire lifetime of the swap – we receive interest on the swap's principal compounded at K in return for making interest payments on the swap's principal at the floating reference rate. If at time T we enter simultaneously into a *par* payer swap (which by definition of a par swap doesn't incur any additional costs) with the same principal and the same floating side as the underlying swap, we pay interest at the swap rate $K_S(T, T')$ valid at this time on the fixed side and receive the floating side. The cash flows on the floating sides of both swap positions (the underlying of the option and the par swap) cancel each other out, with the remaining net cash flow at each time T_i being simply the difference between the two fixed sides:

$$N\left[B_K(T_{i-1}, T_i)^{-1} - B_{K_S(T,T')}(T_{i-1}, T_i)^{-1}\right] \;\; \forall\, i = 1, \ldots, n$$

Exercise of a receiver swaption combined with a par swap obtained at no cost thus generates with certainty (i.e. without any market risk) the above cash flows over the entire lifetime of the underlying. Naturally, the receiver swaption will only be exercised if the cash flows thus generated are positive, i.e. if $K > K_S$. The option matures before any of the payments are due $T < T_i$. Similarly as for caps and floors, these payments are discounted back to T to obtain the value of the swaption at the time of exercise:

$$\begin{aligned}
&c_R^{\text{swap}}(T, T', K|T) \\
&= \begin{cases} N \sum_{i=1}^{n} B_R(T, T_i)\left[B_{K_S(T,T')}(T_{i-1}, T_i)^{-1} - B_K(T_{i-1}, T_i)^{-1}\right] & , K_S(T, T') > K \\ 0 & , K_S(T, T') \leq K \end{cases} \\
&p_R^{\text{swap}}(T, T', K|T) \\
&= \begin{cases} N \sum_{i=1}^{n} B_R(T, T_i)\left[B_K(T_{i-1}, T_i)^{-1} - B_{K_S(T,T')}(T_{i-1}, T_i)^{-1}\right] & , K_S(T, T') < K \\ 0 & , K_S(T, T') \geq K \end{cases}
\end{aligned} \tag{18.11}$$

The "call" corresponds to the payer swaption, the "put" to the receiver swaption. These equations hold in general, irrespective of the compounding convention used.

Valuation as Interest Rate Options (Lognormally Distributed Interest Rates)

As with caps and floors, simple compounding is commonly used to calculate interest over each single swap period, i.e. $B_K(T_{i-1}, T_i)^{-1} = 1 + K(T_i - T_{i-1})$, etc. As such, the option values as of the exercise date T are given by

$$c_R^{\text{swap}}(T, T', K|T) = N \max\{K_S(T, T') - K, 0\} \sum_{i=1}^{n} (T_i - T_{i-1}) B_R(T, T_i)$$
$$p_R^{\text{swap}}(T, T', K|T) = N \max\{K - K_S(T, T'), 0\} \sum_{i=1}^{n} (T_i - T_{i-1}) B_R(T, T_i)$$

where we made use of the fact that neither K nor K_S depend on the index i. These equations make it plainly clear that the *swap rate* $K_S(T, T')$ is the underlying of the swaption. The maximum functions on the right-hand side can be interpreted as the payoff profiles of plain vanilla options with strike price K and maturity T on an underlying whose price is given by $S(T) = K_S(T, T')$. The forward price of this underlying is then the *forward* swap rate $K_S(T, T'|t)$. This plain vanilla option can then be priced exactly as were caplets and floorlets with the Black-76 model, Equation 9.10.

The fact that the swaptions' payments do not occur at exercise time T but at the payment dates T_i is rectified by the introduction of the discount factors $B_R(T, T_i)$ for the (unknown) future spot rates. To avoid arbitrage opportunities, these must be set equal to the discount factors $B_R(T, T_i|t)$ for the *forward* rates holding at time t. With the definition of the forward discount factors given in Equation 3.3 and with Equation 9.10 for the maximum functions, we thus obtain the prices of the swaptions in the Black-76 model

$$\begin{aligned} &c_R^{\text{swap}}(T, T', K|t) \\ &= N\left[K_S(T, T'|t)\mathrm{N}(x) - K\mathrm{N}(x - \sigma\sqrt{T-t})\right] \sum_{i=1}^{n} (T_i - T_{i-1}) \underbrace{B_R(T, T_i|t) B_R(t, T)}_{B_R(t, T_i)} \\ &p_R^{\text{swap}}(T, T', K|t) \\ &= N\left[-K_S(T, T'|t)\mathrm{N}(-x) + K\mathrm{N}(-x + \sigma\sqrt{T-t})\right] \sum_{i=1}^{n} (T_i - T_{i-1}) \underbrace{B_R(T, T_i|t) B_R(t, T)}_{B_R(t, T_i)} \end{aligned}$$

The valuation of a swaption thus requires the calculation of only *one* option. What remains are zero bonds known at time t with maturities at the payment dates specified

Swaption				
5.82% Forward Swap Rate				
12.50% Forward Volatility				
$250,000 Principal				
	7.00% Strike		5.82% Strike	
	Call	Put	Call	Put
Value in %	1.03%	9.18%	3.45%	3.45%
Value in $	$2,582	$22,942	$8,630	$8,630

Figure 18.6: Pricing swaptions. The underlying is a swap fix vs. 12-months floating over ten years, beginning in three years.

in the contract.

$$
\begin{aligned}
c_R^{\text{swap}}(T,T',K|t) &= \widetilde{N}\left[K_S(T,T'|t)\mathrm{N}(x) - K\mathrm{N}(x-\sigma\sqrt{T-t})\right] \\
p_R^{\text{swap}}(T,T',K|t) &= \widetilde{N}\left[-K_S(T,T'|t)\mathrm{N}(-x) + K\mathrm{N}(-x+\sigma\sqrt{T-t})\right] \\
\text{where } \widetilde{N} &= N\sum_{i=1}^{n}(T_i - T_{i-1})B_R(t,T_i) \\
x &= \frac{\ln\left(K_S(T,T'|t)/K\right)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}
\end{aligned}
\tag{18.12}
$$

In Figure 18.6, the valuation of swaptions using Equation 18.12 is demonstrated for two different strikes on the forward swap in Figure 17.1. The swaptions each have the same lifetime (and the same interest rate curves were used for the valuation) as do the caps, floors and collars in Figure 18.5. The underlying, namely the forward swap rate, was calculated in Figure 17.2. For this forward swap rate, we assume a volatility of 12,5%. The sum of the spot rate discount factors from Figure 17.1 with lifetimes from 4 to 13 years yields 6.884114041. This is the factor $\widetilde{N}$ to be used in Equation 18.12. The prices of at-the-money swaptions (i.e. with strike equal to the forward swap rate of 5.82%) are the same for payer (call) and receiver (put) swaptions. This is a special case of the *put-call parity for swaptions* (see Section 18.4.4).

Valuation as Bond Options (Normally Distributed Interest Rates)

To enter into a receiver swap is the same as buying a bond whose coupons and coupon payment dates are given by the fixed side of the swap and in return selling a floater whose floating rate and payment dates are defined by the floating side of the swap. A receiver swaption can thus be interpreted as an option on a bond with a coupon K, whose payment dates and lifetime are defined by the fixed side of the underlying swap. The strike of this bond option is the amount payable for the bond at exercise. This is given by the value at time of exercise T of the floater defined by the floating side of the underlying swap. In general, the value at time T of a floater is not known at time t and as a result we are forced to price a bond option whose strike is unknown. The swaption's underlying swap, however, is usually defined so that it begins at the start of

Parameter of the Bond Option	Corresponding Parameter for Swaption
Option Type	call for receiver swaption, put for payer swaption
Underlying	bond with coupon equal to the strike K of the swaption, principal equal to the principal of the swap, maturity and coupon payment dates as on the fixed side of the swap
Forward Price $S(t,T)$	dirty forward price (including accreud interest) of the bond at maturity T of the swaption
Volatility σ	Volatility of this forward price
Strike K of the Option	Principal of the swap

Table 18.2: Swaptions as options on coupon bonds.

a complete period. This means that the floater is to be priced as if it where just fixed at time T. It's present value, according to Equation 16.3 is then identically equal to its face value. With this interpretation, swaptions are options on coupon bonds with strike equal to the principal and therefore can be priced in the same way as was done in Section 18.4.1. The correspondence to bond options is summarized in Table 18.2. This correspondence holds for all compounding methods.

Put-Call Parity for Swaptions

In a portfolio consisting of a *short* payer swaption and a *long* receiver swaption with the same underlying lifetime from T to T' and the same strike K, one of the two swaptions will *definitely* be exercised at maturity T. If the current swap rate is greater than the strike K at time T, the receiver swaption expires worthless but the counterparty exercises the payer swaption and, as a result of the short position, a receiver swap is obtained with the counterparty paying the fixed rate of K. On the other hand, if the current swap rate at time T is less than K, the counterparty allows the payer swaption to expire while the receiver swaption is exercised, and consequently a receiver swap with the counter party paying the fixed rate K enters the portfolio. Thus, in any case the portfolio at maturity results in a receiver swap at the fixed rate K. The portfolio must then have the same value as a forward receiver swap for all earlier times $t < T$ as well, since otherwise arbitrage would be possible. This is the put-call parity for swaptions. The exact same argument can be used to show that a portfolio consisting of a long payer swaption and short receiver swaption has the same value as a forward payer swap.

If a payer and a receiver swaption have the same strikes and the same lifetime then the following holds:
Short payer swaption and long receiver swaption equals long forward receiver swap
Long payer swaption and short receiver swaption equals long forward payer swap

If the common strike of payer and receiver swaptions are equal to that of the forward swap rate holding for the corresponding lifetime, then the value of the portfolio must be zero since the corresponding forward swap is worthless. This means that the receiver and payer swaptions with the forward swap rate as their strike have the same value. This was demonstrated in Figure 18.6.

19

Exotic Options

19.1 Traditional and General Definition of an Option

An *option* gives its purchaser the right to buy (*call*) or sell (*put*) a specified underlying at a fixed price (*strike*) at (*European*) or up to (*American*) a fixed date (*maturity*).

In the age of exotic options, this traditional definition is no longer sufficient. In order to include the "exotics", this definition requires generalization: an option on one or several underlyings with prices given by $S_1, S_2, ...S_m$ is characterized by its payoff profile. The *payoff* of the option is a function $F(S_1, S_2, ...S_m)$ of the underlying prices and indicates the cash flows arising for the option's holder upon exercise[1]. American options can be exercised at any time during the lifetime of the option in contrast to the European option, which can only be exercised at maturity. There are also options which can be exercised at several specific times during their lifetime. These are called *Bermuda options* or *Atlantic options* since, in a sense, they are an intermediate form between American (exercise is possible at any time during the option's lifetime) and European (exercise is possible at a *single* time, namely at maturity) options.

19.2 Payoff Profiles for Selected Exotics

In the world of exotic options payoffs are significantly more complicated than those of plain vanilla options. Selected examples are indicated below. Some more examples will be discussed in the next sections. These kinds of instruments can in principle all be priced using the methods introduced in Part II (for example, finite difference methods or Monte Carlo simulations). In the following, K always denotes the strike, T the maturity of the option and S the price of the underlying.

[1] For example, *plain vanilla* European calls and puts with maturity T and strike K on an underlying with a price S are uniquely defined by the payoff

$$c_S^{\text{plain}}(T, T, K) = \max\{0, S(T) - K\}, \quad p_S^{\text{plain}}(T, T, K) = \max\{0, K - S(T)\}$$

19.2.1 Power Options

Power options are exotics with payoff profiles similiar to those of plain vanilla options. They involve powers of the underlying, of the strike, or of their difference:

Asymmetric Power Call:	$\max\{0, (S(T))^n - K^n\}$ with $n > 0$
Asymmetric Power Put:	$\max\{0, K^n - (S(T))^n\}$ with $n > 0$
Symmetric Power Call:	0 if $S(T) \leq K$, otherwise $[S(T) - K]^n$ with $n > 0$
Symmetric Power Put:	0 if $S(T) \geq K$, otherwise $[K - S(T)]^n$ with $n > 0$
Self-Quanto Call	$S(T)\ \max\{0, S(T) - K\}$
Self-Quanto Put	$S(T)\ \max\{0, K - S(T)\}$

19.2.2 Cliquet and Coupe Options

A *cliquet option* settles periodically and resets the strike at the spot level then valid. A cliquet can therefore be thought of as a series of "pre-purchased" at-the-money options, but where the total premium is determined in advance. The payout on each option can either be paid at the final maturity, or at the end of each reset period. For one reset at a previously fixed time t' with $t < t' < T$, the payoff at the final maturity T is given as:

Cliquet Call:	$B_R(t', T) \max\{0, S(t') - K\} + \max\{0, S(T) - S(t')\}$
Cliquet Put:	$B_R(t', T) \max\{0, K - S(t')\} + \max\{0, S(t') - S(T)\}$

Here the payoff at time t' is compounded until the final maturity T. The number of reset periods is specified in the contract in advance. More resets make the option more expensive. A cliquet option is always more expensive than the corresponding plain vanilla at-the-money option with the same final maturity.

A *coupe option* settles periodically and resets the strike at the worst (from the option holder's perspective) of either the spot level then valid or the original strike set for the first period. It is a series of options where the total premium is determined in advance. The payout on each option can be paid at final maturity, or paid at the end of each reset period. For one reset at a previously fixed time t' with $t < t' < T$, the payoff is given as:

Coupe Call:	$B_R(t', T) \max\{0, S(t') - K\}$
	$+ \max\{0, \min\{S(T) - S(t'), S(T) - K\}\}$
Coupe Put:	$B_R(t', T) \max\{0, K - S(t')\}$
	$+ \max\{0, \min\{S(t') - S(T), K - S(T)\}\}$

Here the payoff at time t' is compounded until the final maturity T. The number of reset periods is determined by the buyer in advance. More reset periods make the option more expensive, but because of the "worst of" feature a coupe is always cheaper than the corresponding cliquet.

19.2.3 Look-Back Options

There are in principle two types of *look-back options* also known as *hindsight options*: Either the strike is set at the start and at maturity, the buyer can "look back" over the life of the option and choose the most favorable underlying price achieved during the option's life time to maximize profit between strike and exercise. Or the strike is set at maturity. This means at maturity the buyer can "look back" and choose the

most favorable underlying price achieved during the options life *as the option strike* to maximize profit between this strike and the underlying price at maturity. The look-back feature is thus very attractive to investors as it gives the buyer the best possible payout. However, look-back options are generally very expensive. The most common payoff profiles are

Minimum-Strike Call:	$S(T) - \min\{S(t')\}$
Maximum-Strike Put:	$\max\{S(t')\} - S(T)$
Maximum-Asset Call:	$\max\{0, \max\{S(t')\} - K\}$
Minimum-Asset Put:	$\max\{0, K - \min\{S(t')\}\}$
Look-Back Spread Call:	$\max\{0, \max\{S(t')\} - \min\{S(t')\} - K\}$
Look-Back Spread Put:	$\max\{0, K - \max\{(S(t')\} + \min\{S(t')\}\}$
Look-Back Straddle:	$\max\{S(t')\} - \min\{S(t')\}$

where each of the maxima and minima are taken over the entire lifetime of the option, i.e. for t' such that $t \leq t' \leq T$.

19.2.4 Asian Options

An *average strike Asian option* is an option where the strike rate equals the average of the underlying over the life of the option. The strike rate can therefore only be calculated at maturity of the option. Instead of using the average underlying price as the strike, it can also be used as the price which is to be compared to a fixed strike in the payoff. These kinds of Asian options are then called *average price options*. The payoff profiles are explicitly

Average-Price Call:	$\max\{\ 0, \text{average}[S(t')] - K\ \}$
Average-Price Put:	$\max\{\ 0, K - \text{average}[S(t')]\ \}$
Average-Strike Call:	$\max\{\ 0, S(T) - \text{average}[S(t')]\ \}$
Average-Strike Put:	$\max\{\ 0, \text{average}[S(t')] - S(T)\ \}$

The function "average[ ]" usually means either the geometric or arithmetic mean, depending on how the option was defined. In fact, an Asian option can use any agreed upon method for averaging. The averaging is in general taken over the entire lifetime of the option, $t \leq t' \leq T$. The average can be calculated using the daily, weekly, monthly or quarterly underlying prices noted from an agreed source at agreed upon times as specified in the contract.

19.2.5 Rainbow and Exchange Options

The family of *rainbow options* encompasses options where the payout is based upon the relationship between two or more underlyings. Also known as *multi-factor options*, rainbow options include spread options, "better-of" options, "worst-of" options etc. A rainbow option referring to two underlyings S_1 and S_2 is sometimes called a *two-color rainbow*. The most common payoff profiles are

Rainbow:	$\max\{\ S_1(T), S_2(T),\ K\ \}$
Call-On-Maximum:	$\max\{\ 0,\ \max\{S_1(T), S_2(T)\} - K\}$
Put-On-Maximum:	$\max\{\ 0,\ K - \max\{S_1(T), S_2(T)\}\}$
Call-On-Minimum:	$\max\{\ 0,\ \min\{S_1(T), S_2(T)\} - K\}$
Put-On-Minimum:	$\max\{\ 0,\ K - \min\{S_1(T), S_2(T)\}\}$
Spread-Call:	$\max\{\ 0,\ S_2(T) - S_1(T) - K\}$
Spread-Put:	$\max\{\ 0,\ K - S_2(T) - S_1(T)\}$
Dual-Strike:	$\max\{\ 0,\ S_1(T) - K_1, S_2(T) - K_2\}$
One-For-Another:	$\max\{\ 0, S_1(T) - S_2(T)\ \}$
Out-Performance:	$\max\{\ S_1(T), S_2(T)\ \}$
Under-Performance:	$\min\{S_1(T), S_2(T)\ \}$

If the payoff contains only the underlyings S_1 and S_2 as in the last three lines of this table, then the option is also called an *exchange* option.

19.2.6 Compound and Bermuda Options

Compound options are options whose underlying is also an option (usually a plain vanilla option). Upon exercise at maturity T of a compound option, the option holder obtains (or delivers) a plain vanilla option maturing at a later date $T' > T$ and strike K_2 on the same underlying in exchange for the strike K_1 of the compound option. The payoff profiles at maturity T are

Call-on-call:	$\max\{0, c_S(T, T', K_2) - K_1\}$
Call-on-put:	$\max\{0, p_S(T, T', K_2) - K_1\}$
Put-on-call:	$\max\{0, K_1 - c_S(T, T', K_2)\}$
Put-on-put:	$\max\{0, K_1 - p_S(T, T', K_2)\}$

An application of compound options are the *extendable options*. At maturity T, the holder of an extendable option has the right to extend the lifetime to a future date $T' > T$. The payoff profile for extendable options are:

Extendible call:	$\max\{0, S(T) - K_2, c_S(T, T', K_2) - K_1\}$
Extendible put:	$\max\{0, K_2 - S(T), p_S(T, T', K_2) - K_1\}$

where the strike K_1 is often chosen to be zero. With strike $K_1 = 0$, the extendable option is an example of a *Bermuda option*. A Bermuda option, also called an *Atlantic option*, is an option which can only be exercised at specific times during the lifetime of the option. It is, in a sense, an intermediate form between an American (exercise is possible at any *arbitrary* time during the lifetime of the option) and a European option (exercise is possible at a *single* time, namely at maturity). The extendable option is a Bermuda option with exactly one early-exercise date prior to maturity. Bermuda options are commonly employed as *embedded options* in structured bonds.

Another example of a compound option is the *chooser option*. At inception, the chooser option has a strike, a "chooser" date T, and a final maturity date $T' > T$. At the chooser date, the buyer can choose whether the option from the chooser date onwards to final maturity is a call or put. The payoff profile at maturity T is thus

Chooser:	$\max\{c_S(T, T', K_2), p_S(T, T', K_2)\}$

A chooser option is interesting for investors who expect strong volatility increase in the underlying but who are uncertain about the direction. It is therefore ideal mechanism to take positions on volatility.

Yet another example from the family of compound options is the *caption*. A caption is the right to buy or sell an interest rate cap at some defined point in the future for a defined premium. As an option on an option, such a right is a compound option. If at maturity of a caption representing the right to buy (sell) a cap, the underlying interest rate cap is cheaper (more expensive) in the market than the caption strike, the caption holder will let the caption expire. If, however, the underlying cap is more expensive (cheaper) in the market, the caption holder will buy (sell) the cap for the strike described in the caption agreement.

The same agreement, but with reference to a floor, is known as a *floortion*.

19.3 Black-Scholes for Exotics

A closed form Black-Scholes price can be obtained for some exotics by adeptly decomposing them into a combination of options whose prices are known analytically or can be determined using known analytical methods. This will be demonstrated for pay-later and digital options. It is sometimes possible to find Black-Scholes solutions even for options which are path dependent as will be shown for the case of barrier options.

19.3.1 Pay-Later Options

A *pay-later option* must be paid for upon exercise of the option and *only if* the option is in fact exercised. Intuitively, such an option is like an insurance premium against an accident which is payable only in event that the accident actually occurs. This may sound too good to be true but such options can actually be priced and are traded. The idea is to hide the option price in the payoff profile so that the holder of the option need pay nothing upon its acquisition. To accomplish this, the cash flows upon exercise must be reduced by a certain amount in comparison with the payoff of the corresponding plain vanilla option. The payoff profiles of the put and the call are

$$c_S^{\text{later}}(T,T,K) = \begin{cases} S(T) - K - \bar{c} & , \quad S(T) > K \\ 0 & , \quad S(T) \leq K \end{cases}$$

$$p_S^{\text{later}}(T,T,K) = \begin{cases} K - \bar{p} - S(T) & , \quad S(T) < K \\ 0 & , \quad S(T) \geq K \end{cases}$$

The option is exercised exactly when the corresponding plain vanilla would be exercised, namely when $S(T) > K$, or $S(T) < K$. The only difference is that the holder of the call at exercise has to pay a higher amount, $K + \bar{c}$, instead of the strike K, and the holder of the put receives only $K - \bar{p}$ instead of the entire strike K. To price such options, we modify the strikes ($K + \bar{c}$ for calls and $K - \bar{p}$ for puts) in the pricing formulas for plain vanilla options in, for instance, Equation 11.19 of the binomial model for European options. The (risk-neutral) *probabilities,* however, remain the same as for the plain vanilla options. In other words, x in Equation 9.7 remains the same as before. As the time steps dt become infinitesimally small, the binomial distributions converge to normal distributions and the Black-Scholes formula emerges, as shown in Section 11.4. Thus, for pay-later options we only have to modify the strike in the payout (not in the

probabilities) to directly get the Black-Scholes prices

$$c_S^{\text{later}}(t,T,K) = \tilde{S}(t,T)\mathrm{N}(x) - (K+\bar{c})\,B_r(t,T)\mathrm{N}(x-\sigma\sqrt{T-t})$$
$$p_S^{\text{later}}(t,T,K) = -\tilde{S}(t,T)\mathrm{N}(-x) + (K-\bar{p})\,B_r(t,T)\mathrm{N}(-x+\sigma\sqrt{T-t}) \tag{19.1}$$

where

$$x(t) = \frac{1}{\sigma\sqrt{T-t}}\ln\left(\frac{\tilde{S}(t,T)}{K\,B_r(t,T)}\right) + \frac{1}{2}\sigma\sqrt{T-t} \tag{19.2}$$

and $\tilde{S}(t,T)$ is the dividend corrected spot price as given in Equation 3.5. The premium to be paid at exercise, $\bar{c}$ and $\bar{p}$ for the call and put respectively, are established when the option is written, i.e. at t_0, in such a way that the option is worthless at this time. From the condition that the option price in Equation 19.1 is zero at t_0,we can derive

$$\bar{c} = \frac{\tilde{S}(t_0,T)\mathrm{N}(x_0)}{B_r(t_0,T)\,\mathrm{N}(x_0-\sigma\sqrt{T-t_0})} - K$$
$$\bar{p} = \frac{-\tilde{S}(t_0,T)\mathrm{N}(-x_0)}{B_r(t_0,T)\,\mathrm{N}(-x_0+\sigma\sqrt{T-t_0})} + K$$

where x_0 stands for the value of x at $t=t_0$.

19.3.2 Digital Options

Digital options, also referred to as *binary options*, are options for which the *amount* of the payoff is independent of the underlying. The underlying only determines whether a payoff *occurs* at all. The option holder receives the same payout irrespective of how far in the money the option closes. The payoff thus has a binary (digital) character: either a fixed amount or nothing at all is paid. We will now show how to price such options using Black-Scholes. Thus we have to assume the options to be *European*. There are however also options called *one touch digitals* in the market, which also give the buyer a fixed payout profile. Unlike ordinary digitals, one-touch digitals pay out a fixed amount if the underlying reaches the strike *at any time* during the options life time. They can therefore be considered as *American* style digital options.

Cash-or-Nothing

The holder of a *cash-or-nothing option* receives upon exercise a fixed amount of money (for example, 1 euro) if the underlying price is higher (call) or lower (put) than the strike of the option. The payoff profiles of the call and put are

$$c_S^{\text{con}}(T,T,K) = \begin{cases} 1 & ,\ S(T) > K \\ 0 & ,\ S(T) \le K \end{cases}$$
$$p_S^{\text{con}}(T,T,K) = \begin{cases} 1 & ,\ S(T) < K \\ 0 & ,\ S(T) \ge K \end{cases}$$

where "con" stands for "cash or nothing". Pricing these options is accomplished by decomposing them into options whose prices can be determined using conventional

methods (*option stripping*). The payoff profile of the *pay-later* option corresponds to that of the *plain vanilla* option where the holder pays an additional fixed amount $\bar{c}$ or $\bar{p}$ upon exercise. These fixed amounts correspond to the payoff profiles of *digital* options. A pay-later call thus has the same payoff profile as a portfolio consisting of a plain vanilla call (long) and $\bar{c}$ digital options (short). An analogous relation holds for put options:

$$c_S^{\text{later}}(t,T,K) = c_S^{\text{plain}}(t,T,K) - \bar{c}\, c_S^{\text{con}}(t,T,K)$$
$$p_S^{\text{later}}(t,T,K) = p_S^{\text{plain}}(t,T,K) - \bar{p}\, p_S^{\text{con}}(t,T,K)$$

Solving for $c_S^{\text{con}}(t,T,K)$ and $p_S^{\text{con}}(t,T,K)$ and using Equations 9.7 and 19.1 for the plain vanilla and the pay-later options yields the Black-Scholes prices for digital options (with x as in Equation 19.2):

$$c_S^{\text{con}}(t,T,K) = B_r(t,T)\mathrm{N}(x - \sigma\sqrt{T-t}) \tag{19.3}$$
$$p_S^{\text{con}}(t,T,K) = B_r(t,T)\mathrm{N}(-x + \sigma\sqrt{T-t})$$

These equations can be interpreted as follows: the value of the option is the discounted payoff upon exercise (1 euro) multiplied by the risk-neutral probability that the option is exercised.

Asset-or-Nothing

The holder of an *asset-or-nothing option* receives the underlying (or its cash value $S(T)$) if the underlying price is greater than (call) or less than (put) the strike of the option. Seen as such, the amount obtained upon exercise is certainly dependent on the underlying but remains binary (digital) in the sense that the option holder either receives the underlying or nothing at all. The payoff profiles of a call and put are

$$c_S^{\text{aon}}(T,T,K) = \begin{cases} S(T) & , \quad S(T) > K \\ 0 & , \quad S(T) \leq K \end{cases}$$
$$p_S^{\text{aon}}(T,T,K) = \begin{cases} S(T) & , \quad S(T) < K \\ 0 & , \quad S(T) \geq K \end{cases}$$

where "aon" stands for, "asset or nothing". This resembles the payoff profile of the corresponding plain vanilla options with the exception that the holder of the call, for example, need not pay the strike price K in order to obtain the underlying. Were the holder to pay the strike and simultaneously receive the amount of the strike, the payoff profile would remain the same. Thus, the asset-or-nothing call can be interpreted as a plain vanilla call and K cash-or-nothing calls. An analogous decomposition can be produced for the asset-or-nothing put. The *option stripping* of the asset-or-nothing options thus yields the expressions

$$c_S^{\text{aon}}(t,T,K) = K\, c_S^{\text{con}}(t,T,K) + c_S^{\text{plain}}(t,T,K)$$
$$p_S^{\text{aon}}(t,T,K) = K\, p_S^{\text{con}}(t,T,K) - p_S^{\text{plain}}(t,T,K)$$

Inserting the Black-Scholes equations for the already known option prices discussed above gives

$$c_S^{\text{aon}}(t,T,K) = \tilde{S}(t)\text{N}(x) \qquad (19.4)$$
$$p_S^{\text{aon}}(t,T,K) = \tilde{S}(t)\text{N}(-x)$$

where again, $\tilde{S}(t,T)$ is the spot price adjusted for dividends as given in Equation 3.5.

Range Floater

An application of digital options is the *range floater*. A range floater is a deposit or note that accrues interest on days when the underlying S is within a predefined range between S_1 and S_2 and accrues zero on days when the underlying is outside that range. Range floaters are usually principal guaranteed so that the investor is assured of at least receiving the principal back. A range floater is thus a series of daily digital options (plus a zero bond to ensure the redemption of the principal at maturity). On each day during the life of the range floater *two* digital options mature: a long digital call with strike S_1 and a short digital call with strike S_2, with $S_1 < S_2$. Each day the underlying is above S_1 the long digital call is automatically exercised. If the underlying is above S_2 both the long and the short call will automatically be exercised, resulting in zero net payout. The range floater will therefore accrue interest equal to the digital payoff each day the underlying is within the range. Range floaters can be designed with any underlying including interest rates and FX rates. Range floaters are also known as *fairway bonds* or *fairway floaters* or *daily range accruals*.

19.3.3 Barrier Options

Barrier options have become so common (primarily in foreign exchange markets) that many investors no longer think of them as exotic. Barrier options are European options and come in the form of *knock-in* and *knock-out* options. A knock-out option corresponds to a plain vanilla option with the additional condition that the option becomes worthless (*knocks out*) if the underlying price breaks through a specified barrier H. Depending on whether the barrier is smaller or larger than the current price of the underlying, we speak of *down-and-out* and *up-and-out* options. A knock-in option is completely analogous to the knock-out option. In this case however, the additional condition is that the option remains worthless as long as the underlying does *not* attain the barrier. As soon as the barrier has been attained, the option becomes plain vanilla option. The option is thus first "brought to life" (*knocks in*) when the barrier is attained. We speak of a *down-and-in* and *up-and-in* options, depending on whether the barrier has to be attained from above or below.

To determine whether the barrier has been hit or not, the underlying is usually measured once a day at a fixed time (say 12:00 London time) and with reference to an agreed upon source, for instance a Reuters, Bloomberg or Telerate page.

Sometimes a knock-out option is defined in such a way that it does not completely lose its value if the underlying breaks through the barrier, but a small previously agreed

upon amount, called the *rebate,* is paid[2]. Analogously, a rebate is sometimes paid after the maturity of a knock-in option if the option has not been activated in its lifetime. For the sake of clarity, however, we will not enter into a further discussion of these relatively minor details.

The payoff profiles at maturity T of knock-out barrier options with strike K and barrier H on an underlying with a spot price $S(t)$ are:

Up-and-Out Call:	$\max\{0, S(t) - K\}$ if $S(t') < H$
Down-and-Out Call:	$\max\{0, S(t) - K\}$ if $S(t') > H$
Up-and-Out Put:	$\max\{0, K - S(t)\}$ if $S(t') < H$
Down-and-Out Put:	$\max\{0, K - S(t)\}$ if $S(t') > H$

where the conditions must hold for *all* t' such that $t < t' \leq T$. The payoff profiles of the corresponding knock-in barrier options are

Up-and-In Call:	$\max\{0, S(t) - K\}$ if $S(t') \geq H$
Down-and-In Call:	$\max\{0, S(t) - K\}$ if $S(t') \leq H$
Up-and-In Put:	$\max\{0, K - S(t)\}$ if $S(t') \geq H$
Down-and-In Put:	$\max\{0, K - S(t)\}$ if $S(t') \leq H$

where the conditions must in each case hold for *at least one* t' with $t < t' \leq T$.

A portfolio consisting of a knock-out option and the corresponding knock-in option, for example a down-and-out call and a down-and-in call, yields the same payoff as the corresponding plain vanilla option. If the barrier is attained then the knock-out option "dies" and the knock-in becomes a plain vanilla call. Conversely, if the barrier is never attained in the lifetime of the options, the knock-out pays the same as a plain vanilla call and the knock-in is worthless. Thus, a portfolio consisting of a knock-out option and the corresponding knock-in option is worth exactly the same as the corresponding plain vanilla option for all times before and at maturity:

$$\begin{aligned} c_S^{\text{plain}}(t,T,K) &= c_S^{\text{di}}(t,T,K,H) + c_S^{\text{do}}(t,T,K,H) \\ &= c_S^{\text{ui}}(t,T,K,H) + c_S^{\text{uo}}(t,T,K,H) \\ p_S^{\text{plain}}(t,T,K) &= p_S^{\text{di}}(t,T,K,H) + p_S^{\text{do}}(t,T,K,H) \\ &= p_S^{\text{ui}}(t,T,K,H) + p_S^{\text{uo}}(t,T,K,H) \end{aligned} \tag{19.5}$$

where the superscripts "di", "do", etc. stand for "down and in", "down and out", etc. It is therefore sufficient to determine the prices of, for example, the knock-out options. The knock-in prices then follow immediately from Equation 19.5. In addition to the four different barrier option variants, we must further distinguish whether the strike is above or below the barrier. Hence, we need to find equations for 8 different cases. With an appropriate choice of notation, these 8 cases can be summarized in a single Black-Scholes prices for knock-out options (the derivation of which will not be given

[2]This concept is useful in the valuation of companies using option pricing theory.

here):

$$
\begin{aligned}
V^{\text{out}} = \delta\, B_q(t,T) S(t) \{ & \mathrm{N}\left(\frac{\beta - x_+}{\sigma\sqrt{T-t}}\right) - \mathrm{N}\left(\frac{\alpha - x_+}{\sigma\sqrt{T-t}}\right) \\
& - e^{-\lambda h}\left[\mathrm{N}\left(\frac{\beta - x_+ + 2h}{\sigma\sqrt{T-t}}\right) - \mathrm{N}\left(\frac{\alpha - x_+ + 2h}{\sigma\sqrt{T-t}}\right)\right]\} \\
& - \delta\, B_r(t,T) K \{ \mathrm{N}\left(\frac{\beta - x_-}{\sigma\sqrt{T-t}}\right) - \mathrm{N}\left(\frac{\alpha - x_-}{\sigma\sqrt{T-t}}\right) \\
& - e^{(2-\lambda)h}\left[\mathrm{N}\left(\frac{\beta - x_- + 2h}{\sigma\sqrt{T-t}}\right) - \mathrm{N}\left(\frac{\alpha - x_- + 2h}{\sigma\sqrt{T-t}}\right)\right]\}
\end{aligned}
\tag{19.6}
$$

In this equation, the following parameters are defined to hold for all knock-out options:

$$
\begin{aligned}
x_+ &= \ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right) + \frac{\sigma^2}{2}(T-t) \\
x_- &= \ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right) - \frac{\sigma^2}{2}(T-t) \\
\lambda &= 1 + \frac{\ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right)}{\frac{\sigma^2}{2}(T-t)}\,, \quad h = \ln\left(\frac{S(t)}{H}\right)
\end{aligned}
\tag{19.7}
$$

The parameters α, β and δ, on the other hand, are not identical for all knock-out options. These three parameters are presented in Table 19.1 for the different types of knock-out options. Together with the Equations 19.6 and 19.7, this table provides all the necessary information for determining the Black-Scholes price of all knock-out options[3].

The case $H \leq K$ for an up-and-out call does not make sense since for the call to pay out anything the underlying must be larger than the strike K. But for this to happen, it must cross the barrier H since $H \leq K$. But then the up-and-out call "dies". Therefore an up-and-out call with $H \leq K$ never pays out anything. Such an option is always worth zero. Inserting this result in Equation 19.5 shows that the corresponding up-and-*in* call is worth exactly as much as the plain vanilla call. This makes sense since for the call to pay out anything the underlying must be larger then the strike K and therefore cross the barrier. Thus, an up-and-in call with $H \leq K$ will be "alive" in *every* situation where the corresponding plain vanilla call pays out anything. Therefore it has the same value as the plain vanilla call.

The same reasoning leads to the result that a down-and-out put with $H > K$ never pays out anything (and is therefore worth zero) and the down-and-in put with $H > K$ is worth exactly as much as the corresponding plain vanilla put.

As an example, we present the explicit equation for the price of a down-and-out call for the case that the barrier is less than or equal to the strike. Referring to Table 19.1 for this case, we find that $\delta = +1$, $\alpha = -\ln(S(t)/K)$, $\beta = +\infty$. Substituting

[3]The following well-known properties of the norma distribution are useful for the valuation: $\mathrm{N}(\infty) = 1$, $\mathrm{N}(-\infty) = 0$, $1 - \mathrm{N}(x) = \mathrm{N}(-x)$

	$H \leq K$	$H > K$
Up-and-Out Call	pointless	$\delta = +1$ $\alpha = -\ln\left(\frac{S(t)}{K}\right)$ $\beta = -h$
Up-and-Out Put	$\delta = -1$ $\alpha = -\infty$ $\beta = -h$	$\delta = -1$ $\alpha = -\infty$ $\beta = -\ln\left(\frac{S(t)}{K}\right)$
Down-and-Out Call	$\delta = +1$ $\alpha = -\ln\left(\frac{S(t)}{K}\right)$ $\beta = +\infty$	$\delta = +1$ $\alpha = -h$ $\beta = +\infty$
Down-and-Out Put	$\delta = -1$ $\alpha = -h$ $\beta = -\ln\left(\frac{S(t)}{K}\right)$	pointless

Table 19.1: Parameters for the valuation of knock-out options using Equation 19.6.

these parameters into Equation 19.6 and making use of the properties of the normal distribution yields

$$
\begin{aligned}
&c_S^{\mathrm{do}}(t,T,H \leq K) = \\
&B_q(t,T)S(t)\left[\mathrm{N}\left(\frac{x_+ + \ln\left(\frac{S(t)}{K}\right)}{\sigma\sqrt{T-t}}\right) - e^{-\lambda h}\mathrm{N}\left(\frac{x_+ + \ln\left(\frac{S(t)}{K}\right) - 2h}{\sigma\sqrt{T-t}}\right)\right] \\
&- B_r(t,T)K\left[\mathrm{N}\left(\frac{x_- + \ln\left(\frac{S(t)}{K}\right)}{\sigma\sqrt{T-t}}\right) - e^{(2-\lambda)h}\mathrm{N}\left(\frac{x_- + \ln\left(\frac{S(t)}{K}\right) - 2h}{\sigma\sqrt{T-t}}\right)\right]
\end{aligned}
$$

Here, the argument $x_+ + \ln(S(t)/K)$ is precisely equal to the x known from the plain vanilla options (see for example Equation 9.5). Substituting this and all other parameters as defined in Equation 19.7 (with the exception of λ) finally leads to the equation for the price of a down-and-out call, in agreement with that found by Cox and Rubinstein in [34], for example:

$$
\begin{aligned}
&c_S^{\mathrm{do}}(t,T,H \leq K) = \\
&\overbrace{B_q(t,T)S(t)\mathrm{N}(x) - B_r(t,T)S(t)\mathrm{N}(x-\sigma\sqrt{T-t})}^{\text{Plain Vanilla}} \\
&- \left(\frac{S(t)}{H}\right)^{-\lambda}\left[B_q(t,T)S(t)\mathrm{N}(y) - B_r(t,T)K\left(\frac{S(t)}{H}\right)^{2}\mathrm{N}(y-\sigma\sqrt{T-t})\right] \qquad (19.8)
\end{aligned}
$$

where the following new abbreviations have been introduced

$$y = x - \frac{2}{\sigma\sqrt{T-t}} \ln\left(\frac{S(t)}{H}\right)$$
$$= \frac{1}{\sigma\sqrt{T-t}} \ln\left(\frac{B_q(t,T)\, H^2}{B_r(t,T) K\, S(t)}\right) + \frac{1}{2}\sigma\sqrt{T-t}$$

The price of the down-and-out call is thus equal to the price of the corresponding plain vanilla call less an amount generated by the barrier. It is thus always cheaper than the plain vanilla call. This makes sense, since its payoff in the most favorable case (when the barrier is not attained) equals that of the plain vanilla call. The barrier increases the risk that a payoff will not be made and thus the barrier option must always be less expensive than its plain vanilla counterpart. They are therefore the appropriate instruments for investors who have an exact opinion as to future developments in the price of an underlying and, based on this opinion, wish to decrease the costs of a hedge, for example. The speculative element should not be overlooked here: when the barrier is attained, a hedging portfolio consisting of knock-out options disappears!

The price of the corresponding down-and-in call follows immediately from Equation 19.5. The value of this call is just the difference between the price of the plain vanilla and the down-and-out call:

$$c_S^{\mathrm{di}}(t,T,H \le K) \tag{19.9}$$
$$= \left(\frac{S(t)}{H}\right)^{-\lambda} \left[B_q(t,T) S(t) \mathrm{N}(y) - B_r(t,T) K \left(\frac{S(t)}{H}\right)^{2} \mathrm{N}\left(y - \sigma\sqrt{T-t}\right) \right]$$

Double Barrier Option

In order to receive the option payout of a *double barrier option* the underlying must not breach either of *two* defined barriers at any time during the life of the transaction. Generally, the barriers are used to define a range inside which the underlying must remain for the life of the transaction in order for the option payout to be made. If the underlying as measured on any day is outside the defined range, the payout is zero, i.e. the holder has been knocked out. The payout for the double barrier option is usually a fixed amount and hence this option type is also known as a *binary knockout range*.

A double barrier option is *not* the addition of two knock-out options. If this were the case, once one barrier had been breached the other option would still be "alive". A double barrier option is the addition of two "*contingent*" Knockout Options, i.e. the "survival" of each barrier option is contingent upon the other barrier not having being breached. Should one barrier be breached, the other barrier option also dies.

Double barrier options are suitable to take advantage of markets expected to trade within a range. Most common in foreign exchange, they are available in most underlying markets, including commodities, interest rate and equity markets.

19.3.4 Ladder Options

With a *ladder option*, the strike is periodically reset when the underlying hits specified trigger levels, at the same time locking in the profit between the old and the new

strike. The trigger strikes appear as rungs on a ladder. This behavior of the option can be summarized by the following payout function: at expiry the option will pay out either the difference between the underlying spot price and the original strike, or the difference between the highest rung reached and the original strike, whichever is greater. The payoff profile in its simplest form is therefore

$$\max\left\{S(T)-K_0,\ \max_i\left\{K_i-K_0\right\},\ 0\right\}\ \text{with}\ K_0<K_1<K_2<...<K_n$$

where $K_1, K_2, ..., K_n$, represent the "rungs on the ladder" established in the contract and K_0 the original strike of the option. The maximum function for these ladder steps refers only to *those* rungs K_i attained by the underlying during the lifetime of the option. Upon exercise of such an option, the holder receives the same payoff as that of a plain vanilla option if $S(T)$ is greater than each of the previously attained steps. However, if the underlying price has declined toward the end of its lifetime, the holder receives the difference between the original strike K_0 and the highest rung attained during the lifetime of the option. Ladder options can be structured to reset the strike in either one or both directions. Also a "Ladder Put" could be constructed along the same lines (if we consider the above payoff profile as a "call"). Ladder options are also known as a *ratchet options* or *lock-in options*. A ladder option with the above payoff profile can be constructed from a plain vanilla call and as a series of long and short knock-in puts (*option stripping*). Since knock-in options can be priced using Equations 19.5 and 19.6, a Black-Scholes price for a ladder option can be obtained by pricing the corresponding portfolio of plain vanilla and barrier options. Explicitly, the portfolio which has the same payoff profile as the ladder option consists of

- one *long* plain vanilla call with strike K_0.
- for each rung K_i (with $i = 1, \ldots, n$) one *long* knock-in put with knock-in barrier $H_i = K_i$ and strike K_i.
- for each rung K_i (with $i = 1, \ldots, n$) one *short* knock-in put with knock-in barrier $H_i = K_i$ and strike K_{i-1}.

All these options have maturity T. To check that this portfolio indeed replicates a ladder option we consider a ladder option with only one rung K_1 and an initial strike $K_0 < K_1$. With one rung and one strike, five different situations can arise at maturity. These are displayed in the following table:

	K_1 reached	K_1 not reached
$S(T) > K_1$	case a)	impossible
$K_0 < S(T) < K_1$	case b)	case d)
$S(T) < K_0$	case c)	case e)

The replicating portfolio consists of the plain vanilla call with strike K_0, a long knock-in put with barrier K_1 and strike K_1 and a short knock-in put with barrier K_1 and strike K_0. We will now show that in each case the portfolio has the same payoff as the ladder option:

case a – payoff of ladder option: $S(T) - K_0$

– payoff of options portfolio: payoff from plain vanilla call is $S(T) - K_0$; the puts are both knocked in but expire worthless.

case b
– payoff of ladder option: $K_1 - K_0$
– payoff of options portfolio: payoff from plain vanilla call is $S(T) - K_0$; the puts are both knocked in; payoff from long put is $K_1 - S(T)$; short put expires worthless; altogether: $S(T) - K_0 + K_1 - S(T) = K_1 - K_0$

case c
– payoff of ladder option: $K_1 - K_0$
– payoff of options portfolio: plain vanilla call expires worthless; the puts are both knocked in; payoff from long put is $K_1 - S(T)$; payoff from short put is $-[K_0 - S(T)]$; altogether: $K_1 - S(T) - [K_0 - S(T)] = K_1 - K_0$

case d
– payoff of ladder option: $S(T) - K_0$
– payoff of options portfolio: payoff from plain vanilla call is $S(T) - K_0$; the puts are never knocked in.

case e
– payoff of ladder option: 0
– payoff of options portfolio: plain vanilla call expires worthless; the puts are never knocked in.

For more than one rung we simply repeat the above idea: each time a rung K_i is reached for the first time by the underlying, a put is knocked in which guarantees that the underlying can be sold for at least K_i. That locks in a profit $K_i - K_0$ with certainty since at maturity the underlying can be bought for K_0 because of the call and then immediately sold for K_i because of the put. However, when the next rung K_{i+1} is reached, one more put is knocked in and we would then have two puts; the "old" one with strike K_i and the "new" one with strike K_{i+1}. That is one put too many. Therefore the "old" put (the one with strike K_i) now has to be "destroyed". This is achieved by the short knock-in put with strike K_i which comes into being exactly when K_{i+1} is hit. The payoff of this short put then exactly compensates for the payoff of the "old" long put. The portfolio is then left with the payoffs of the new put and the plain vanilla call and thus, a payoff $K_{i+1} - K_0$ is locked in with certainty, just as it should be to replicate the ladder option.

19.4 Numerical Pricing Methods for Exotics

The list of exotics described above is far from complete and an analytic price determination – even under the assumptions of the Black-Scholes world – is not possible for the majority of exotics. The multitude of exotic instruments can be roughly assigned to four different classes as indicated in Table 19.2. The options are differentiated by whether the payoff profile is dependent solely on current values of the underlying or on earlier values as well and whether or not early exercise is allowed. We speak of a *path dependent* option if the option's value is dependent on the history of the underlying, i.e. if the price evolution (the *path*) of the underlying during the lifetime of the option

	European	American
Path-Independent	PDE Monte Carlo	PDE with Moving Boundary Recombining Trees
Path-Dependent	(PDE) Monte Carlo	(PDE with Moving boundary) Bushy Trees

Table 19.2: Classification and "brute force" valuation methods for exotic options.

affects the option's payoff. Assuming the availability of adequate computer capacity and performance, some common numerical pricing methods leading to reasonable prices for exotic derivatives are indicated in Table 19.2. The fundamental assumption behind these methods is that the underlying is governed by a random walk as given in Equation 3.13, or more generally by an Ito process as defined by Equation 3.16. "PDE" means that the price can be determined by solving a partial differential equation (by applying numerical methods such as the finite difference methods introduced in Part II, for example). "(PDE)" means that setting up and solving an equation (numerically) is possible for *some* path dependent options (such as barrier options), but not for all.

19.4.1 Monte Carlo for European Exotics

A method which, in principle (neglecting potential problems regarding CPU time) can be applied to all derivatives with a European payoff mode is the Monte Carlo method introduced in Chapter 12. Using a "Pseudo-Code", we will demonstrate how *Asian options* can be priced with this method. The schematic algorithm offered here explicitly shows how the chapters on Monte Carlo simulation and on random walks can be applied to option pricing. The algorithm is suitable for the valuation of all European options depending on a single underlying (with or without dividends). Multiple underlyings can be simulated using the methods discussed in Section 12.3.2 the for the case of two underlyings. The generalization of this method for more than two underlyings can be realized using Equations 21.35 and 23.1.

In the following concrete example, the algorithm has been explicitly formulated for Asian options for the purpose of demonstrating that even exotics, which have the reputation of being especially difficult to price, can be valued using the Monte Carlo method. The complete algorithm can be found in the form of an executable Visual Basic program contained in the Excel workbook MonteCarloSimulation.xls. The variable $x = \ln(S(t')/S(t))$ is simulated as a random walk where $S(t')$ is the price (denoted by "S" in the algorithm) at the current simulation time $t' = t + i\,dt$ and $S(t)$ (denoted by "S0" in the algorithm) is the price at the start of the simulation, i.e. at the value date t. Since the Monte Carlo simulations are based on random walks and, as such, on random numbers, this method will generate a somewhat different option price with each simulation run. To demonstrate the calculation of the error in accordance with Section 29.3, the statistical error (denoted by "EPrice") of the option price generated by the simulation will also be determined.

```
'* Initialization of the payoff profile:
'* ====================================
```

```
Payoff1 = 0
Payoff2 = 0
'* The loop over i is the loop over the paths:
'* ========================================
For i = 1 to Path
'* Initialization at the start of each path:
'* ======================================
    SamplePath = 0
    x = 0
'* The loop over j generates one underlying path:
'* ============================================
    For j = 1 to Steps
'* The next step on the path of the underlying's random walk:
'* ========================================================
        dx = drift * dt + vola * randomnumber() * Squareroot(dt)
        x = x + dx
        S = S0 * Exp(x)
'* The evaluation after each step in a path:
'* ======================================
        SamplePath = SamplePath + S
    Next j
'* The evaluation after each path:
'* ==============================
'* For Average-Price Call:
    SamplePath = SamplePath / Steps
    If SamplePath > Strike then
        Payoff1 = Payoff1 + (SamplePath - Strike)
        Payoff2 = Payoff2 + (SamplePath - Strike)^2
    End If
'*For Average-Price Put:
'* SamplePath = SamplePath / Steps
'* If Strike > SamplePath Then
'*     Payoff1 = Payoff1 + (Strike - SamplePath)
'*     Payoff2 = Payoff2 + (Strike - SamplePath)^2
'* End If
'*For Average-Strike Call:
'* SamplePath = SamplePath / Steps
'* If S > SamplePath Then
'*     Payoff1 = Payoff1 + (S - SamplePath)
'*     Payoff2 = Payoff2 + (S - SamplePath)^2
'* End If
'*For Average-Strike Put:
'* SamplePath = SamplePath / Steps
'* If SamplePath > S Then
'*     Payoff1 = Payoff1 + (SamplePath - S)
'*     Payoff2 = Payoff2 + (SamplePath - S)^2
```

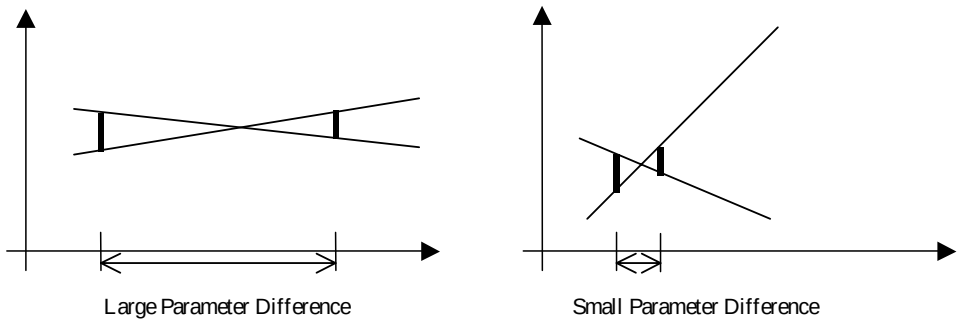


Figure 19.1: The range of possible straight lines through to points where the points are known only up to a certain error respresented by the bars.

```
'* End If
Next i
'* final evaluation of the price and its error after the simulation
'* ===================================================================
Payoff1 = Payoff1 / Paths
Payoff2 = Payoff2 / Paths
Price = Payoff1
EPrice = Squareroot(Payoff2-Payoff1^2)/Squareroot(Paths-1)
'* Discounting for Options which are not future-styled:
Price = Exp(-InterestRate * Time) * Price
EPrice = Exp(-InterestRate * Time) * EPrice
```

According to Equation 13.18, the option sensitivities ("*Greeks*") can be determined by running the simulation twice, having modified the parameter according to the sensitivity under consideration while using the *same* random numbers as in the previous simulation run. The difference in the price thus obtained, divided by the parameter change, yields the corresponding sensitivity. Intuitively, this is the slope of the line passing through two measured prices. Two effects contribute to the error of this slope induced by the error of the option prices:

- *Statistical Error*
  This error is generated by the measurement error in both prices. The error associated with the slope of the line passing through two points, whose positions are known only up to an error term, increases the closer these points lie to one another. This is illustrated in Figure 19.1. In order to minimize this error, the measured points should therefore lie as far from one another as possible. This means that the parameter change should be as large as possible.

- *Systematic Error*
  The sensitivities are defined as the derivatives of the price function with respect its different parameters, i.e. they represent the change of the price with respect to a *small* change in the parameter, divided by this change in the parameter.

Thus, choosing a large change in the parameter to reduce the statistical error means moving away from the actual definition of the sensitivities and thereby committing a *systematic* error.

Both of these effects compete with one another. The "most favorable" choice for the size of the parameter change is always a compromise between a large parameter change minimizing the effect of the statistical error and a small parameter change minimizing the systematic error.

As the commentary in the algorithm suggests, only minor modifications need to be made when other option types are priced. The locations in the code where these modifications are to be made are written in bold print[4]. The lines printed in bold are those which serve to measure the payoff profile and its error and generate the values of "`Payoff1`" and "`Payoff2`".

If, for example, a barrier option (with the additional input parameter `Barrier`) were to be priced instead of an Asian, the bold-faced lines for an up-and-in call, would read:

```
'* Initialization at the start of each path:
'* ========================================
SamplePath = 0
'* The evaluation after each step of the path:
'* ==========================================
If S > Barrier
    SamplePath = 1
End If
'* The evaluation at the end of each path:
'* ======================================
If SamplePath = 1 And S > Strike Then
    Payoff1 = Payoff1 + (S - Strike)
    Payoff2 = Payoff2 + (S - Strike) ^ 2
End If
```

That is all! By modifying these few lines measuring the payoff profile, barrier options can be priced instead of Asian options together with all their respective Greeks an error terms.

This example should serve to illustrate that any arbitrary European option, regardless of how complicated, can be priced with the Monte Carlo method if the payoff profile is known. This means that as soon as the option is defined (through its payoff) the valuation via Monte Carlo simulation is as good as complete! The maximum and minimum functions frequently appearing in the payoff profiles can be realized through simple "`If...Then`" conditions as was done in the examples above.

Some of the options listed in Section 19.2 require extensions of the above given algorithm which will not be presented here. Rainbow and exchange options, for example, can be priced through a simulation of correlated prices. An *extendable option*, for example, an extendable call, is priced by simulating its underlying's price $S(T)$ at the first maturity date T. The payoff of the extendable call is then the maximum of a

[4]In addition there may of course be changes in the input parameters.

) the payoff of a plain vanilla call with strike K_2 and maturity T and b) the value of a plain vanilla call with strike K_2 and maturity T' as of the time point T (calculated with Black-Scholes, for instance) less the strike, K_1, of the extendable option.

19.4.2 The Binomial Model for American Exotics

In contrast to the Monte Carlo method, the binomial trees introduced in Chapter 11 are suitable for the valuation of not only European but American options as well. As was done for the Monte Carlo simulation above, schematic algorithms will be provided here demonstrating explicitly how binomial trees can be applied to price options. The complete algorithm can be found in the form of an executable Visual Basic program in the Excel Workbook BinomialTree.xls.

Recombinant Trees for Path-Independent Options

As emphasized in Chapter 11, recombinant trees are only suitable for path-independent options since the path taken to arrive at a particular node is not unique if the branches of the tree recombine. The payoff profile of an (exotic) option is then defined on the boundary of the tree, in particular at maturity T of the option. We subsequently calculate backward in time applying Equation 11.8. The algorithm is actually a mere reformulation of Example 18.4. It will be applied here to price an American *self quanto option*. The payoff profile of this option, as was noted in Section 19.2.1, is $S(T) \times \max\{0, S(T) - K\}$. Naturally, any other path-dependent payoff profile could have been inserted instead. The algorithm can be generalized for options with multiple underlyings but such a generalization is quite involved as the binomial tree takes on an additional dimension with the addition of each new underlying [144].

The time to maturity $T - t$ is denoted by "t" in the algorithm, the number of binomial steps by "n". The computations are performed with continuous compounding.

```
dt = t / n
discount = Exp(-r * dt)
up = Exp(vola * Squareroot(dt))
down = 1 / up
p = (Exp((r - q) * dt) - down) / (up - down)
'* Inserting the boundary values for the option:
'* ============================================
For j = 0 to n
    S= S0 * up ^ j * down ^ (n - j)
'* Payoff Profile:
    exercise = S * ( S - Strike )
    If Optiontype = Put Then
        exercise = -exercise
    End If
    If exercise > 0 Then
        f(n; j) = exercise
    End If
Next j
```

```
'* Calculate backwards for each node in the tree:
'* ============================================
For i = n - 1 To 0 stepsize -1
    For j = i To 0 stepsize -1
        S= S0 * up ^ j * down ^ (n - j)
'* Intrinsic Value:
'* ================
    exercise = S * ( S - Strike )
    If Optiontype = Put Then
        exercise = -exercise
    End If
'* Option price:
'* ============
    f(i; j) = discount * (p * f(i + 1; j + 1)
        + (1 - p) * f(i + 1; j))
    If exercise > f(i; j) And Optionstyle = American Then
        f(i; j) = exercise
    End If
    Next j
Next i
Price(c) = f(0; 0)
```

As in the above algorithm for the Monte Carlo simulation, the lines written in bold print indicate where modifications have to be made when a different type of option is to be priced. These are the locations in the code where the payoff profile is defined in the program, i.e. where both the values at the tree's boundary (at time T) are specified and where the intrinsic value is computed for the purpose of deciding whether to exercise early. For example, a power option can be priced by simply replacing the bold-faced text above with "`exercise = (S - Strike)^P`".

Further boundary conditions, for example for specific *underlying* values (as for barrier options, for example) can be quite easily accounted for as well by setting the option's values at the nodes lying at the boundary equal to the given boundary values. For reasons of stability the grid should always be constructed so that the boundaries lie *directly on the nodes* (this applies to all grid methods, also for instance for finite differences schemes).

Binomial Bushy Trees for Path-Dependent Options The schematic algorithm presented here is suitable for the valuation of *path-dependent* American (and European) options. So-called *bushy trees* will be constructed in which each path of the underlying (denoted by "`u`") is generated without regard to recombination. Information on the entire history of the underlying is then accessible. Each individual underlying path is then evaluated to obtain the variable (denoted by "`x`" in the algorithm) determining the price of the option. For Asian options, for example, this is the average price of the underlying over the lifetime of the option. This will change from path to path, even if the paths arrive at the same end point. Once the path-dependent variables are determined, the payoff profile of the option on the boundary of the tree can be

implemented, finally using Equation 11.8 to calculate backwards in time along the tree. The procedure is demonstrated here for Asian average price options, where the lines corresponding to the Average Strike options are indicated in the commentary. The right to early exercise is accounted for in the algorithm.

```
dt = t / n
discount = Exp(-r * dt)
up = Exp(Vola * Squareroot(dt))
down = 1 / up
p = (Exp((r - q) * dt) - down) / (up - down)
If Dividendpayment Then
    divstep = Rounddown(Dividendtime * n / t; 0)
End If
'* Construction of the Underlying:
'* ================================
u(0; 0) = S0
x(0; 0) = S0
For i = 0 To n - 1
    For j = 0 To 2 ^ i - 1
        u(i + 1; 2 * j + 1) = u(i; j) * up
        u(i + 1; 2 * j) = u(i; j) * down
        If Dividendpayment And i = divstep Then
            u(i + 1; 2 * j + 1) = u(i + 1; 2 * j + 1) - Dividends
            u(i + 1; 2 * j) = u(i + 1; 2 * j) - Dividends
        End If
'* Evaluation of the underlying after each step in the path:
'* ========================================================
'* Asian:
          x(i + 1; 2 * j + 1) = u(i + 1; 2 * j + 1) + x(i; j)
          x(i + 1; 2 * j) = u(i + 1; 2 * j) + x(i; j)
    Next j
Next i
'* Inserting the boundary values for the option at maturity:
'* ========================================================
For j = 0 To 2 ^ n - 1
'* Average Price:
     exercise = x(n; j) / (n + 1) - Strike
'* Average Strike:
'* exercise = u(n; j) - x(n; j) / (n + 1)
    If Optiontype = Put Then
        exercise = -exercise
    End If
    If exercise > 0 Then
        f(n; j) = exercise
    End If
Next j
'* Valuation of the option:
```

```
'* ======================
For i = n - 1 To 0 Steplength -1
    For j = 0 To 2 ^ i - 1
'* Intrinsic value:
'* ================
'* Average Price:
         exercise = x(i; j) / (i + 1) - Strike
'* Average Strike:
'*         exercise = u(i; j) - x(i; j) / (i + 1)
        If Optiontype = Put Then
            exercise = -exercise
        End If
'* Option price:
'* ============
        f(i; j) = discount * (p * f(i + 1; 2 * j + 1)
            + (1 - p) * f(i + 1; 2 * j))
        If exercise > f(i; j) And Optionstyle = American
            f(i; j) = exercise
        End If
    Next j
Next i
Price(c) = f(0; 0)
```

Again, only those lines written in bold print need to be modified if an option of another type is to be priced. These are the locations where the payoff profile and the procedure for calculating the particular path-dependent property of the underlying enter into the code. If we are interested in pricing, for example, an American *look-back option* (minimum-strike call) with a payoff profile $S(T) - \min\{S(t')\}$, the bold print should be replaced by

```
'* Evaluation of the underlying after each step in the path:
'* ========================================================
'* Determination of the minimal price:
If x(i + 1; 2 * j + 1) > u(i + 1; 2 * j + 1) Then
    x(i + 1; 2 * j + 1) = u(i + 1; 2 * j + 1)
End
'* Inserting the boundary values for the option at maturity:
'* ========================================================
exercise = u(n;j) - x(n; j)
'* Intrinsic Value:
'* ================
exercise = u(i;j) - x(i; j)
```

These slight modifications accomplish the valuation of a look-back option rather than the original Asian option. The difference between the current price and the minimum attained up to that point is computed at maturity as well as when deciding whether to exercise early.

Naturally, the numerical methods just introduced are quite computation intensive. The computation of *bushy trees* in particular is extremely time consuming. Let n

denote the number of binomial steps. The computing time and memory requirements are then proportional to $(n+1)2^{(n+1)}$. Each additional step more than doubles the computing time and memory required! Depending on the hardware, more than 14 steps can hardly be computed. For example, 14 steps give $(14+1)2^{(14+1)} = 491,520$, for 15, we already have 1,048,576. The algorithms introduced here have not been optimized in any way with respect to reducing the computation time. They were written to ensure the clarity of the pricing algorithm in each case. Despite these difficulties, a Visual Basic Excel program based on these algorithms can price an exotic option on a Pentium PC in a few minutes. A computation time of approximately 30 seconds is sufficient to generate 1,000 Monte Carlo paths for path-dependent options with 100 steps per path or a binomial bush with 14 steps. A recombinant binomial tree with 500 steps can be generated in the same amount of time if path-independent exotics are to be priced. This illustrates quite drastically how CPU restricted the computation of binomial *bushes* is compared to the other methods.

20

Structured Products and Stripping

It is often the case that financial instruments are not traded in the form described in the previous chapters, but that combinations of these instruments are collected and traded as new financial instruments. Such constructions are called *structured products* or *structured instruments*. In order to understand the behavior, the valuation and the risk of such structured products, they must first be decomposed into their elementary instruments in a process called *stripping*. The following elementary instruments play a particularly important role in this process:

- Coupon and LIBOR strips
- Fixed rate (plain vanilla bonds) and floating rate bonds (plain vanilla floater)
- Standard swaps (plain vanilla swaps) and asset swaps
- Options which – at least under the assumptions of the Black-Scholes world – can be priced analytically (plain vanilla, digital, barrier)

The non-optional parts of the above elementary instruments can be stripped further into still smaller building blocks, the cash flows. A single *cash flow* can be interpreted as a synthetic zero bond. A *zero bond* is thus the smallest elementary instrument available on the market. It can be interpreted as an "atom" of the interest rate market. Stripping is not only to be viewed as a theoretical thought model. Bonds are actually stripped in practice into coupon and redemption payments and these can then be traded separately. This process is called *coupon stripping*.

The investment environment for fixed-rate instruments has changed dramatically in the last 20 years. The increasing yield fluctuations represent new challenges in the analysis and valuation of interest rate instruments. Limiting the analysis exclusively to yield considerations can no longer be justified. Stripping is the basis for the analysis of new financial instruments, for example through the evaluation of their sensitivities.

20.1 Coupon and LIBOR Strips vs. Bonds

Coupon strips and *LIBOR strips* are a collection of fixed and floating interest payments, respectively, whose interest periods follow one after the other continuously and without overlap. The difference between such strips and the respective fixed or floating bonds is that strips do not involve any principal redemption or amortization payments. The cash flows consist solely of interest payments which are each calculated on the basis of the compounding and day count conventions particular to the cash flow (see Section 3.1). A *coupon strip* is a series of *coupon* payments arising in regular, fixed intervals. Analogously, a *LIBOR strip* is a series of *floating* interest payment arising in regular, fixed intervals . The amount of the LIBOR payment is fixed at the beginning of the interest period for payment at maturity. As a rule, payment is made at the end of an interest period. Coupon and LIBOR strips are particularly suitable for stripping interest rate swaps since these swaps involve solely an exchange of interest payments.

A bond can be stripped into a coupon strip and (perhaps various) amortization payments as well. It is easy to see that strips (also referred to as *legs*) are the natural component factors of almost all common plain vanilla cash flow instruments and represent one of the most important building blocks for the construction of more complex payment structures (for example, structured bonds).

A simple example will demonstrate the stripping of a straight bond into a coupon strip and a principal repayment. We find this fundamental principle in the international bond market in the form of *coupon stripping* where *synthetic zero bonds* are created by stripping existing coupon bonds, and then traded separately. These synthetic zero bonds are discounted interest and principal payments of, for example, US treasuries, Australian, Canadian, or French government bonds. They are also referred to as stripped bonds. Since July 4, 1997, the German Bundesbank allows 10-year and 30-year bonds to be stripped. The stripping dates are January 4 and July 4. Examples of strippable German government bonds are the 6.25% government bond from 1994 (2024), the "New Year" government bond from 1997, and the 30-year bond with a coupon date on July 4. Examples of synthetic bonds on the international bond market are certificates of accrual on treasury securities (*CATS*), certificates on government receipts (*COUGARs*), discounted investment in negotiable government obligations (*DINGOs*) and treasury investment growth receipts (*TIGRs*). The names of such stripped bonds can be uniquely associated with the securities house which issued them, for example, CATS (Salomon Brothers), COUGARs (Paribas), LIONs (Lehmann Brothers), STRIPS (US treasury Department) or TIGRs (Merill Lynch).

The issue of such receipts frequently proceeds as follows: an investment house buys a government bond already issued (for example, a US treasury) and divides it into individual interest certificates (coupon stripping). The individual interest certificates are then issued as separately tradable and deliverable quasi-government zero bonds.

In this way, a 30 year US treasury with semi-annual coupons, has 60 *coupon issues* and one *principal issue*, each issued as a discounted zero bond. The final maturity of each coupon issue corresponds to the maturities of the underlying interest certificates; the final maturity of the principal issue corresponds to the maturity of the bond. *STRIPS* (separate trading of registered interest and principal of securities) have a special property: they are issued directly by the Federal Reserve. STRIPS therefore

Bund_Apr06	Start of Coupon Period	Maturity	Payment Date	Payment in % of Principal
Coupon	26.04.96	26.04.97	28.04.97	6.25%
Coupon	26.04.97	26.04.98	27.04.98	6.25%
Coupon	26.04.98	26.04.99	26.04.99	6.25%
Coupon	26.04.99	26.04.00	26.04.00	6.25%
Coupon	26.04.00	26.04.01	26.04.01	6.25%
Coupon	26.04.01	26.04.02	26.04.02	6.25%
Coupon	26.04.02	26.04.03	28.04.03	6.25%
Coupon	26.04.03	26.04.04	26.04.04	6.25%
Coupon	26.04.04	26.04.05	26.04.05	6.25%
Coupon	26.04.05	26.04.06	26.04.06	6.25%
Principal	26.04.96	26.04.06	26.04.06	100.00%

Table 20.1: Cash flows of the German government bond issued on April 26, 1996.

represent a direct claim on the US treasury Department.

Consider, for example, the 10-year German government bond issued on April 26, 1996. This bond pays a fixed annual coupon of 6.25% on April 26 of each year and matures on April 26, 2006. In Table 20.1, the stripping of this 10-year bond into individual cash flows is demonstrated explicitly. The bond consists of a total of eleven cash flows each of which can be interpreted as a zero bond. In the first column, the components of the 10 year bond are listed. This government bond consists of the coupon strip of 10 coupon issues and one principal issue. The third column shows the maturity dates of each of the zero bonds, the fourth column, the associated payment dates. Listed in the last column is the amount of each of the cash flows. German government bonds do not adhere to a certain business day convention (see Section 3.1); the maturity dates may fall on a weekend or on a holiday. Naturally, for *pricing* only the *future* cash flows are of interest.

20.2 Structured Bonds

Market Situation

Bonds play an important role in the capital market. Measured in terms of sales the bond market in Germany, for example, is approximately four times as large as the stock market. In the past, bond analysis was restricted to estimating the credit worthiness of the issuers and the determination of more or less exact yield numbers. Traditionally, bonds were bought with the intention of holding them until maturity and receiving the repayment at that time. The bond was purchased with the goal of earning the annual coupons and repayment at maturity. This approach is referred to as a *buy and hold strategy*. The yield, or more specifically the *interest proceeds* in this passive strategy were the criteria used to judge the return on a security. The nominal yield established upon purchase of the bond (under the assumption that it was to be held until maturity) or the total yield of a portfolio were the criteria of success for investment in fixed rate

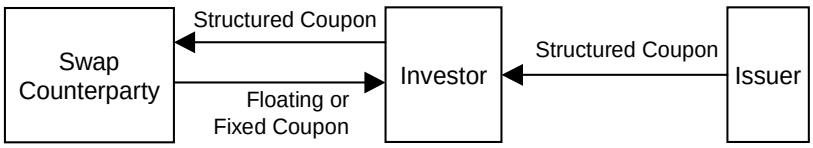


Figure 20.1: Swapping a structured bond.

instruments. The job of the securities analyst was solely to find securities earning the highest possible (nominal) yields while remaining within certain "acceptable" bounds (for example, levels of credit worthiness). Changes in the basic framework and attitudes coupled with the arrival of new types of instruments (sometimes called "financial innovations") have since changed the situation on the bond markets dramatically.

The trend toward increasingly complex financial innovations (including derivatives) will continue in the future. New and sometimes extremely risky financial constructions become available on the market almost daily. Many inexperienced investors or even investment consultants at banks can no longer see through the jungle of manifold innovations and their assessment of potential risk is frequently erroneous. It often happens that such securities are initially classified as relatively harmless and later prove to be extremely risky. This includes for example securities whose interests payments are based on the money market but whose price evolution is influenced by the capital market (for example, reverse floaters, leveraged floaters).

Such *structured bonds* primarily consist of diverse instruments, for instance derivative instruments like options, affecting the coupon, the repayment or possibly the lifetime of the bond. In the last few years, the volume in the structured bonds sector has grown significantly. In certain market constellations however, it can occur that a demand for structured bonds cannot be generated despite a sufficient supply of such instruments. In such situations, bid prices can nevertheless be established on the basis of what is known as *asset swaps.* In this way, they can then be resold as synthetic floaters, for example. Thus, finding an appropriate price for "swapping" the structured bonds becomes a central issue.

Asset Swaps with Structured Bonds

In Figure 20.1, a diagram representing the general approach taken in the purchase of a structured bond and the corresponding swapping of bond coupons is shown. The role of the swap partner is to transform the structured coupon into a fixed or floating cash flow. The original, complex structured bond is swapped for a fixed-rate or floater for the investor. Finding a price for the structured bond using an asset swap involves the valuation of the individual components making up the bond. Only in this way is it possible for the investor to remove the risk from the structured bond by making the corresponding counter-transactions.

The procedure in constructing an asset swap on a structured bond will be illustrated using a *step-up callable bond* as an example. This kind of bond typically consists of

two individual components. The first component is represented by a normal bond paying coupons until the time when the bond may be terminated by the issuer. If the issuer does not terminate the bond, the bond continues to pay coupons (but at a higher interest rate than before) until maturity. The second component of the step-up callable bond therefore consists of a short call representing the issuers right to terminate the bond. The option's underlying is a bond whose coupons agree with those of the bond in the second time period described above, starting at the time when issuer may exercise his option and maturing at the maturity of the entire bond.

As an example we consider the step-up callable bond with the following characteristics:

- 5-year lifetime, price at issue date 100%
- The issuer has the right to terminate ("call") the bond after two years
- 5% coupon in the first two years (before the termination date)
- 8% coupon in the last three years (after the termination date)

A long position in this step-up callable thus consists of

- A long position in a two-year bond with a coupon of 5%
- A long position in a three year forward bond with coupon of 8%, starting in two years
- A short call with a lifetime of two years on a three year bond with a coupon of 8% and a strike in the amount of the principal (100%).

To transform these cash flows into a normal floater, the following transactions must be carried out:

- The 5% coupon, that the investor receives in the first 2 years is transformed into a floating cash flow via a long payer swap. Thereby, the investor pays the swap partner a coupon in the amount of 5% and in return receives, for example, LIBOR plus 15 basis points (if the 2-year swap rate is 5%-0.15%=4.85%).
- Simultaneously, the coupon of 8% which the investor receives for a further three years if the bond is not terminated is turned into a floating cash flow through a long payer swaption (an option on a payer swap) with a strike of 8%. The investor thereby has the right to pass on an 8% coupon to the swap partner and in return receive, for example, LIBOR plus 15 basis points (if the 3-year forward swap rate in two years is 8%-0.15%=7.85%).

If two years later the actual 3-year swap rate (= par rate for bonds with the same credit rating) in the market is *below* 8%, the issuer could refinance more cheaply on the capital market than through the bond already issued and will therefore exercise the right to terminate. The payer swaption, however, will be exercised if and only if the 3-year swap rate in two years lies *above* 8%. As described in Section 16.4.2, the

Year	Bond Cash Flows	Received from Swap	Payment into Swap	Total Cash Flows
0	-100%			-100%
0.5		LIBOR +0.15%		LIBOR +0.15%
1	+5%	LIBOR +0.15%	-5%	LIBOR +0.15%
1.5		LIBOR +0.15%		LIBOR +0.15%
2	+5%	LIBOR +0.15%	-5%	LIBOR +0.15%
2.5		LIBOR +0.15%		LIBOR +0.15%
3	+8%	LIBOR +0.15%	-8%	LIBOR +0.15%
3.5		LIBOR +0.15%		LIBOR +0.15%
4	+8%	LIBOR +0.15%	-8%	LIBOR +0.15%
4.5		LIBOR +0.15%		LIBOR +0.15%
5	+108%	LIBOR +0.15%	-8%	LIBOR +0.15% +100%

Table 20.2: Cash flows of a portfolio consisting of a step-up callable bond, a payer swap and a swaption for the case when the bond is not called.

par rate is equal to the swap rate (assuming the same degree of credit worthiness). This means that the bond will *not* be terminated precisely then when the swaption is exercised and vice verse. If the right to terminate is not exercised, the 8% coupon from the bond is passed through the swaption (which has been exercised) on to the swap partner in return for LIBOR plus 0.15%. The cash flow structure of the entire portfolio consisting of the step-up callable bond, the 2-year payer swap and the payer swaption from the viewpoint of the investor for the case that the bond is not terminated (the call lapses and the swaption is exercised) is presented in Table 20.2. Conversely, the swaption lapses if and only if the bond is terminated. In this case, the repayment is due after two years and all transactions are concluded after two years.

In order to further clarify the procedure described above, we consider the following concrete example. On August 29, 1996 the following step-up callable bond was offered to an investor:

- Westfälische Hypothekenbank public mortgage bond with a nominal principal of DM 10 million.
- Lifetime until Mar-15-2002.
- Issuer's option to terminate on Feb-15-1999 for Mar-15-1999 at 100%
- 5% coupon before the termination date
- 6.5% coupon after the termination date
- Accrued interest of DM 227,777.78 as of value date Aug-29-1996

Based on the investor's personal credit worthiness, the investor pays LIBOR plus 10 basis points for a 6-month loan. In order to determine the price of the step-up

callable, the investor transforms the bond into a synthetic floater paying the 6-month LIBOR rate plus ten basis points. Such a floater can be bought on the market and financed with a loan without requiring the investor to put up personal funds. This means that the cost incurred by the investor in transforming the bond is equal to its fair (arbitrage-free) market value. Determining the bond price takes the following steps:

- Identification of the individual components of the structured bond. The step-up callable bond consists of
  - A long bond with a 5.0% coupon maturing on Mar-15-1999
  - A long forward bond with a 6.5% coupon beginning at Mar-15-1999, and maturing at Mar-15-2002
  - A short call maturing on Feb-15-1999 on a bond with a coupon of 6.5% and a lifetime from Mar-15-1999 until Mar-15-2002.
- Transactions necessary for the transformation of the structured coupon into a floating cash flow.
  - To transform the step-up callable bond into a floater, a payer swap must be entered into in which the investor pays a coupon of 5% in return for a floating cash flow at the 6-month LIBOR plus 0.10% until Mar-15-1999. In pricing the swap, we must take into account that the swap partner receives the entire coupon in the amount of 5% (corresponding to DM 500,000 on a nominal principal of DM 10 million) from the investor on Mar-15-1997 (and no only the interest payment for the partial interest period starting on Aug-29-1996).
  - In order to neutralize the 6.5% bond and the call option, a payer swaption maturing on Feb-15-1999 must be bought, which gives the holder the right to enter into a swap paying the 6.5% fixed and in receiving 6-month LIBOR plus 0.10% for the period from Mar-15-1999 until Mar-15-2002.
- Valuation of the counter-transactions.
  - The current market price for a payer swap in which the investor pays 5% in return for the 6-month LIBOR plus ten basis points satisfying the above stated conditions is negative DM 355,407.06. Thus, the investor would *receive* this amount if he or she enters into this swap.
  - The costs for the purchase of the payer swaption which entitles the investor to pay 6.5% in exchange for LIBOR plus ten basis points at a volatility of 12.92% amount to DM 234,290.24. These prices can be computed from the current term structure valid for the investor using the procedure demonstrated in Sections 16.4 and 18.4.4.

The proceeds from the swap of DM 355,407.06 have to be set against the costs incurred in purchasing the payer swaption at a price of DM 234,290.24. There thus

remains a surplus of DM 121,116.82. Before the purchase price of the bond can be calculated, the accrued interest of DM 227.777,78 from the 5% coupon bond up to August 29, 1996 must first be accounted for. Consequently, we obtain the purchase price for the step-up bond as 98.93% (DM 10,000,000 plus DM 355,407.06 minus DM 234,290.24 minus DM 227,777.78). If the investor now buys the bond at 98.93%, he or she can immediately enter into a payer swap paying 5% fixed to the swap partner and receiving LIBOR plus ten basis points. In addition the investor can buy a payer swaption with the right to enter into a swap paying 6.5% in exchange for LIBOR plus ten basis points if the bond is not terminated. Then the investor would receive a net cash flow of LIBOR plus 0.10% on the nominal of DM 10,000,000 over exactly the lifetime of the step-up callable bond, no matter whether the bond is called or not.

Part IV

Risk

21

Fundamentals

The term *risk* signifies the uncertainty of the future developments in risk factors (for example, interest rate curves, stock prices, foreign exchange rates, volatilities, etc.) resulting in a negative deviation of the quantity of interest (e.g. the value of a portfolio) from a certain reference value. Or expressed in terms of a bank balance sheet, risk is the possibility that the value of assets decreases or that the value of liabilities increases. Among all the conceivable types of risk, for example, market risk, credit risk, operational risk, legal risk, etc., market risk and credit risk are those which are most commonly traded via various financial instruments.

Market risk can be defined as the possible negative deviation of one or more positions (portfolio) from a reference value, dependent on a change in *market risk factors* (interest rates, stock prices, foreign rates, commodity prices, etc.). Credit risk, on the other hand, are potential losses resulting from a business partner's default or a decline in credit worthiness.When solely interested in the risk of a partner's *default*, we speak of *default risk*. Credit risks are categorized into either *issuer risk* or *counterparty risk*.

The starting point for the analysis of market risk and credit risk of interest rate instruments are cash flows and their present value. Likewise, the current market price of stock and foreign currency exchange rates serve as starting points in analyzing the risk associated with investment in stocks and foreign currencies, respectively.

Candidates for the above-mentioned reference values are the current market value (present value) of a financial instrument, the initial investment capital, or the *book value* derived from balance sheet regulations.

The risk of a financial instrument or a portfolio can be determined on the basis of (fluctuations of) its underlyings. Hence, we refer to these as *risk factors*. A risk factor could be the 3-month LIBOR rate, the USD/EUR exchange rate or the Dow Jones index. A *scenario* for a portfolio is defined by specifying a value for each of its risk factors. In modern risk management, *risk* is quantified through the *value at risk*. This number is an upper bound for the loss incurred by a portfolio which – with a probability c – will not be exceeded during some (finite) time period δt. The probability c is referred to as the *confidence*. We will now formulate these concepts much more precisely.

21.1 Risk, Confidence, Percentile and Value at Risk

The value of a portfolio is a function of stochastic processes (the risk factors) and is thus itself a stochastic process. A *confidence interval* of an arbitrary stochastic variable X with density $p(X)$ is defined as the subset of the variable's range (the interval) attained with a previously specified probability c, called the *confidence,* or *level of confidence.* A level of confidence of 95% ($c = 0.95$) for example, implies that the probability that the value realized by the random variable lies outside the confidence interval is only 5%. Or equivalently, we can be 95% certain that a realization of the random variable will lie within the interval. Should the fluctuations about a mean value μ be of interest, the upper and lower bounds, a and $-a$, of a *symmetric confidence interval* can be determined as the values satisfying the condition

$$c \stackrel{!}{=} P(\mu - a < x < \mu + a) = \int_{\mu-a}^{\mu+a} p(x)dx$$

If, on the other hand, we want to know with a previously specified probability that the random variable will not fall below a certain value (the portfolio will not lose more than a certain amount of its value), we consider the *one-sided confidence interval* and determine the boundary a as that value satisfying the condition

$$c \stackrel{!}{=} P(x > a) = 1 - P(x \leq a) = 1 - \int_{-\infty}^{a} p(x)dx \tag{21.1}$$

The *percentile* or *quantile* associated with a previously specified probability c is defined as the value Q_c such that the probability is c, that a random value x is less than or equal to Q_c:

$$P(x \leq Q_c) = c \iff \int_{-\infty}^{Q_c} p(x)dx = c$$

This definition defines the percentile implicitly. The percentile can be determined by "inverting" the cumulative distribution function P:

$$Q_c = P^{-1}(c)$$

The boundary a of a one-sided confidence interval is thus precisely the $(1-c)$ percentile

$$\begin{aligned} c \stackrel{!}{=} P(x > a) &= 1 - P(x \leq a) \\ P(x \leq a) &= 1 - c \\ &\Longrightarrow \\ a = Q_{1-c} &= P^{-1}(1 - c) \end{aligned}$$

Using these definitions of confidence, confidence interval and percentile, the *value at risk* (VaR) of the value V of a financial instrument or a portfolio can be precisely

defined for a specified c. As mentioned above, the value at risk of a portfolio in a time span δt is intuitively the upper bound for the depreciation of the portfolio's value in this time period which will not be exceeded with a specified confidence, c. Here, the random variable is the change in the value δV (not the change in the underlying risk factors!) and VaR is the (value of the) *loss* which will not be exceeded with confidence c.

$$c \stackrel{!}{=} \mathrm{cpf}_{\delta V}(\delta V > -\mathrm{VaR}(c)) \tag{21.2}$$

For the sake of clarity in the notation, we have here used $\mathrm{cpf}_{\delta V}$ (instead of P) to denote the *cumulative probability function* of the random variable δV. This definition of the VaR can be expressed more explicitly as

$$c \stackrel{!}{=} 1 - \mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}(c)) = 1 - \int_{-\infty}^{-\mathrm{VaR}(c)} \mathrm{pdf}_{\delta V}(x)dx \tag{21.3}$$

where $\mathrm{pdf}_{\delta V}(x)$ denotes the *probability density function* of the random variable δV. We now see that the negative VaR is precisely the $(1-c)$ percentile of the distribution of δV :

$$-\mathrm{VaR}(c) = Q_{1-c}^{\mathrm{cpf}_{\delta V}}$$

or equivalently,

$$\mathrm{VaR}(c) = -Q_{1-c}^{\mathrm{cpf}_{\delta V}} = -\mathrm{cpf}_{\delta V}^{-1}(1-c) \tag{21.4}$$

In the discussion up to now, we have emphasized that the probability of the change in the portfolio's value δV (and *not* the risk factors and their probability distributions) defines the value at risk. However, changes in the portfolio's value are induced by the risk factors. In addition, it is not the usual practice to model the stochastic process δV or its distribution directly. Instead, the risk factors (as in pricing) are modeled as stochastic processes and the methods derived for pricing (see, for example Part II) are than used to calculate the value change δV and its distribution (and percentiles) induced by the changes in the risk factors.

Using the probability distribution of the *risk factor* instead of the probability distribution of the *portfolio* for the determination of the value at risk is only possible when V is a *monotonous* function of the risk factor process S. If, for example, V is a monotonously increasing function of S, we have

$$V(S) \leq V(a) \Longleftrightarrow S \leq a$$

and V lies below the confidence interval bound for the *portfolio* if and only if S lies below the confidence interval bound for the *risk factor*. The value at risk of the portfolio is then the difference between the current portfolio value and the value of the portfolio when the risk factor lies at the lower bound of the confidence interval for the *risk factor*. Likewise, for the case where the portfolio's value is a monotonously decreasing function, the value at risk of the portfolio is the difference between the current value of the portfolio and the portfolio's value when the *upper boundary* of the confidence interval of the *risk factor* is attained. The upper bound $\widetilde{a}$ of a one-sided confidence

interval is defined analogously to Equation 21.1 through

$$c \stackrel{!}{=} P(x < \widetilde{a}) = \int_{-\infty}^{\widetilde{a}} p(x)dx \tag{21.5}$$

Thus, if V is a monotonous function of the risk factor, the value at risk of the portfolio is

$$\begin{aligned} \mathrm{VaR}_V(c) &= V(S) - \min\{V(S=a), V(S=\widetilde{a})\} \\ &= \max\{V(S) - V(S=a), V(S) - V(S=\widetilde{a})\} \end{aligned} \tag{21.6}$$

The minimum function guarantees that precisely *that* boundary of the confidence interval is chosen which results in the greatest *loss* for the portfolio. Hence, the one-sided confidence interval for the risk factor which is bounded from *above* determines the value at risk for instruments (or portfolios) whose value *declines* with increasing underlying prices. Analogously, the one-sided confidence interval with a *lower* bound determines the value at risk for those instruments (or portfolios) whose prices *rise* with the rising price of the underlying. The values of many instruments like bonds, futures, and most options are monotonous functions of their risk factors. Instruments (or portfolios) which are *not* monotonous functions of their underlyings must be either stripped into their component elements which are monotonous functions of their underlyings or the distribution of V itself and its confidence interval must be determined directly. The confidence intervals of the *risk factors* are no longer of assistance in this case.

21.2 The Value at Risk of a Single Risk Factor

In order to illustrate the practical application of the concepts just introduced, we now consider a single risk factor S whose (infinitesimal) changes are governed by a *geometric Brownian motion* (*GBM*) as given in Equations 3.13 and 3.22. In other words, the risk factor satisfies the stochastic differential equation

$$d\ln S(t) = \mu dt + \sigma dW \quad \text{where} \quad dW \sim \mathrm{N}(0, dt) \tag{21.7}$$

The evolution of S in a *finite* time interval of length δt is the solution to the stochastic differential equation 21.7. This solution has already been given in Equation 3.25. It is a stochastic process which describes the distribution of the market parameter $S(t+\delta t)$ a finite time step δt later:

$$S(t+\delta t) = S(t)\exp(\mu\delta t + \sigma\delta W) \quad \text{where} \quad \delta W \sim \mathrm{N}(0, \delta t) \tag{21.8}$$

The volatility σ of the risk factor appearing here is *theoretically* the same volatility as appears in the differential equation, Equation 21.7. In practice, it is usually calculated as the standard deviation of the relative price changes over a finite time span δt (for example, through the analysis of historical time series),

$$\frac{S(t+\delta t)}{S(t)} \approx \delta\ln(S(t))$$

or obtained from the volatility values made available by commercial data providers or in internet sites such as for instance by RiskMetrics^{TM}.

As opposed to Equation 21.7 which only holds for infinitesimal time steps dt, the finite time span δt in Equation 21.8 can be taken to be arbitrarily long. Obviously, since the variance of the normally distributed stochastic component δW of the risk factor equals δt, the risk can be kept small by keeping the length of this time interval small. δt should, however, be chosen large enough for the position concerned can be liquidated within this time span. For this reason δt is sometimes referred to as the *liquidation period*. In the Capital Adequacy Directive *CAD II,* liquidation periods of $\delta t = 10$ days are required for internal models; in RiskMetrics^{TM} for instance, data for $\delta t = 1$ day and $\delta t = 25$ days are available .

We now consider the value at risk with respect to a specified confidence c of a portfolio consisting of a single position in N of one risk factor S. The value of this portfolio at time t is $V = NS(t)$. The change in the value $\delta V(t)$ caused by the change in the risk factor is

$$\delta V(t) = N\delta S(t) \quad \text{where} \quad \delta S(t) = S(t + \delta t) - S(t)$$

This case is by no means as special as it may seem. The change in S induces a change in V amplified by the *constant* factor N. The factor N is, so to speak, the *sensitivity* of V with respect to S and δV is a *linear* function of δS. The interpretation of N as the "number of instruments" in the portfolio is not essential in our deliberations. The same results hold for any portfolio whose change in value is a linear function of the change in the risk factor (or at least approximately so), as is the case for the delta-normal approximation in the variance-covariance method introduced below.

The value change δV can be explicitly derived from Equation 21.8 as

$$\begin{aligned} \delta V &= NS(t + \delta t) - NS(t) \\ &= NS(t) \left[\exp\left(\mu\delta t + \sigma\delta W\right) - 1\right] \end{aligned} \tag{21.9}$$

This allows us to express the probability required in Equation 21.3, i.e. the probability that $\delta V \leq -\text{VaR}(c)$, as

$$\text{cpf}_{\delta V}(\delta V \leq -\text{VaR}) = \text{cpf}_{\delta V}\left(NS(t)\left[e^{\mu\delta t + \sigma\delta W} - 1\right] \leq -\text{VaR}\right)$$

The strategy is now to write the unknown cumulative probability function $\text{cpf}_{\delta V}$ in terms of a probability whose distribution is known. The only stochastic variable involved is the Brownian motion whose distribution is given by

$$\delta W \sim \text{N}(0, \delta t) \Rightarrow \delta W \sim X\sqrt{\delta t} \quad \text{where} \quad X \sim \text{N}(0, 1)$$

We begin by rewriting the event that $\delta V \leq -\text{VaR}$ with the purpose of isolating the

stochastic component δW.

$$\begin{aligned}
&\mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}) \\
&\quad = \mathrm{cpf}_{\delta V}\left(e^{\mu\delta t+\sigma\delta W} - 1 \leq -\frac{\mathrm{VaR}}{NS(t)}\right) \\
&\quad = \mathrm{cpf}_{\delta V}\left(e^{\mu\delta t+\sigma\delta W} \leq 1 - \frac{\mathrm{VaR}}{NS(t)}\right) \\
&\quad = \mathrm{cpf}_{\delta V}\left(\mu\delta t + \sigma\delta W \leq \ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right)\right) \\
&\quad = \mathrm{cpf}_{\delta V}\left(\delta W \leq \frac{\ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) - \mu\delta t}{\sigma}\right)
\end{aligned}$$

The probability that δW is less than or equal to a certain value is dependent on the distribution of δW alone and not on that of δV. We can therefore simply replace $\mathrm{cpf}_{\delta V}$ with $\mathrm{cpf}_{\delta W}$:

$$\begin{aligned}
&\mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}) \\
&\quad = \mathrm{cpf}_{\delta W}\left(\delta W \leq \frac{1}{\sigma}\ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) - \frac{\mu}{\sigma}\delta t\right) \\
&\quad = \mathrm{cpf}_{\delta W}\left(\sqrt{\delta t}X \leq \frac{1}{\sigma}\ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) - \frac{\mu}{\sigma}\delta t\right) \\
&\quad = \mathrm{cpf}_{\delta W}\left(X \leq a\right)
\end{aligned}$$

having defined the parameter a as follows

$$a = \frac{1}{\sigma\sqrt{\delta t}}\ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) - \frac{\mu}{\sigma}\sqrt{\delta t} \tag{21.10}$$

The probability that X is smaller than a particular variable is dependent on the distribution of X alone and not on that of δW allowing $\mathrm{cpf}_{\delta W}$ to be simply replaced by cpf_X. The probability distribution of X is just the standard normal distribution $\mathrm{N}(0,1)$ whose density is known explicitly (see for example Section A.4.3). Thus the needed cumulative probability can be written as

$$\begin{aligned}
&\mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}) \\
&\quad = \mathrm{cpf}_X\left(X \leq a\right) \\
&\quad = \int_{-\infty}^{a} \mathrm{pdf}_X(x)dx \\
&\quad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp(-x^2/2)dx
\end{aligned}$$

The value at risk with respect to the confidence c is the value satisfying Equation 21.3. Inserting the above result into this requirement now yields

$$c \stackrel{!}{=} 1 - \mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp(-x^2/2)dx$$

The parameter a introduced in Equation 21.10 is thus the $(1-c)$ percentile of the standard normal distribution

$$a = Q_{1-c}^{\mathrm{N}(0,1)} \tag{21.11}$$

This value can be determined for any arbitrary confidence level. Examples of the boundary of a one-sided confidence interval of the standard normal distribution are

$$\begin{aligned} c = 95\% = 0.95 &\Rightarrow a = Q_{1-c}^{\mathrm{N}(0,1)} \approx -1.65 \\ c = 99\% = 0.99 &\Rightarrow a = Q_{1-c}^{\mathrm{N}(0,1)} \approx -2.326 \end{aligned} \tag{21.12}$$

Both of these confidences are frequently used in practice. *CAD II* requires $c = 99\%$ for internal models, whereas RiskMetrics$^{\mathrm{TM}}$ makes available data at the $c = 95\%$ confidence level.

Having computed the percentile a we are now in a position to apply Equation 21.10 to isolate the value at risk for a previously specified confidence level c:

$$\begin{aligned} a &= \left[\ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) - \mu\delta t\right] / (\sigma\sqrt{\delta t}) \\ a\sigma\sqrt{\delta t} + \mu\delta t &= \ln\left(1 - \frac{\mathrm{VaR}}{NS(t)}\right) \\ \exp\left(a\sigma\sqrt{\delta t} + \mu\delta t\right) &= 1 - \frac{\mathrm{VaR}}{NS(t)} \\ \frac{\mathrm{VaR}}{NS(t)} &= 1 - \exp\left(a\sigma\sqrt{\delta t} + \mu\delta t\right) \\ \mathrm{VaR} &= NS(t)\left[1 - \exp\left(a\sigma\sqrt{\delta t} + \mu\delta t\right)\right] \end{aligned}$$

Now using our main result Equation 21.11 we finally obtain for the value at risk with respect to the confidence level c of a position consisting of N of the same instruments depending solely on a single underlying risk factor S :

$$\mathrm{VaR}(c) = NS(t)\left[1 - \exp\left(\mu\delta t + Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right] \tag{21.13}$$

Thus, under the assumption that S behaves as the random walk in Equation 21.7, the probability is c that any loss due to this investment over the time span δt is smaller than this value at risk.

Now consider the risk of a *short* position consisting of N of the risk factor. The change in the portfolio's value is given by an equation analogous to Equation 21.9, namely

$$\delta V = -NS(t)\left[\exp\left(\mu\delta t + \sigma\delta W\right) - 1\right]$$

and the cumulative distribution function needed in Equation 21.3 is now

$$\begin{aligned} &\mathrm{cpf}_{\delta V}(\delta V \leq -\mathrm{VaR}) \\ &\quad = \mathrm{cpf}_{\delta V}\left(-NS(t)\left[e^{\mu\delta t + \sigma\delta W} - 1\right] \leq -\mathrm{VaR}\right) \\ &\quad = \mathrm{cpf}_{\delta V}\left(NS(t)\left[e^{\mu\delta t + \sigma\delta W} - 1\right] \geq \mathrm{VaR}\right) \end{aligned}$$

where in the last step, multiplication by (-1) has reversed the inequality to a "$\geq$" in the above condition. A "$\geq$" condition, however, is not well suited for the subsequent calculation in which the VaR condition is reduced to a percentile calculation since percentiles are always defined in terms of a "$\leq$" condition. It is simpler to start again from the beginning, this time from the original definition of the value at risk, Equation 21.2. This equation demands the probability $\mathrm{cpf}_{\delta V}(\delta V > -\mathrm{VaR})$. For our short position, this is given by

$$\begin{aligned}
&\mathrm{cpf}_{\delta V}(\delta V > -\mathrm{VaR}) \\
&\quad = \mathrm{cpf}_{\delta V}(-NS(t)\left[e^{\mu\delta t+\sigma\delta W}-1\right] > -\mathrm{VaR}) \\
&\quad = \mathrm{cpf}_{\delta V}(NS(t)\left[e^{\mu\delta t+\sigma\delta W}-1\right] < \mathrm{VaR}) \\
&\quad = \mathrm{cpf}_{\delta W}\left(\delta W < \frac{1}{\sigma}\ln\left(1+\frac{\mathrm{VaR}}{NS(t)}\right)-\frac{\mu}{\sigma}\delta t\right) \\
&\quad = \mathrm{cpf}_{X}\left(X < \frac{1}{\sigma\sqrt{\delta t}}\ln\left(1+\frac{\mathrm{VaR}}{NS(t)}\right)-\frac{\mu}{\sigma}\sqrt{\delta t}\right)
\end{aligned}$$

In the penultimate step, $\mathrm{cpf}_{\delta V}$ was once again replaced with $\mathrm{cpf}_{\delta W}$ since the probability that δW is smaller than a particular fixed value is dependent on the distribution of δW alone, and not on that of δV. The same reasoning applies in replacing $\mathrm{cpf}_{\delta W}$ with cpf_{X} in the last step.

Analogous to Equation 21.10, we introduce the following definition[1]:

$$\widetilde{a} = \frac{1}{\sigma\sqrt{\delta t}}\ln\left(1+\frac{\mathrm{VaR}}{NS(t)}\right)-\frac{\mu}{\sigma}\sqrt{\delta t} \tag{21.14}$$

It then follows that the definition given in 21.2 for the value at risk becomes

$$c \stackrel{!}{=} \mathrm{cpf}_{\delta V}(\delta V > -\mathrm{VaR}) = \mathrm{cpf}_{X}\left(X < \widetilde{a}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\widetilde{a}}\exp(-x^2/2)dx$$

The parameter $\widetilde{a}$ is thus the c percentile of the standard normal distribution

$$\widetilde{a} = Q_c^{\mathrm{N}(0,1)}$$

The symmetry A.53 of the standard normal distribution

$$\mathrm{N}(-x) = 1-\mathrm{N}(x)$$

implies that the percentile obeys the relation[2]

$$Q_c^{\mathrm{N}(0,1)} = -Q_{1-c}^{\mathrm{N}(0,1)}$$

[1] The difference to a in Equation. 21.10 lies in the sign of the VaR.

[2] This can be seen quite easily: Let $c = \mathrm{N}(x)$. By definition, x is the percentile associated with c. Or equivalently, the inverse of the cumulative distribution function gives the percentile

$$Q_c = x = \mathrm{N}^{-1}\left[\mathrm{N}(x)\right] = \mathrm{N}^{-1}(c)$$

It follows that we can use the same percentiles as given for instance in Equation 21.12 for two commonly used confidence levels. Inserting $\widetilde{a} = -Q_{1-c}^{\mathrm{N}(0,1)}$ and solving Equation 21.14 for the value at risk of a short position yields

$$\begin{aligned}\mathrm{VaR}(c) &= -NS(t)\left[1 - \exp\left(\widetilde{a}\sigma\sqrt{\delta t} + \mu\delta t\right)\right] \\ &= -NS(t)\left[1 - \exp\left(\mu\delta t - Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right]\end{aligned}$$

We can now write down the two formulas for the risks involved in both a long and a short position in one and the same risk factor:

$$\begin{aligned}\mathrm{VaR}_{\mathrm{long}}(c) &= NS(t)\left[1 - \exp\left(\mu\delta t + Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right] \\ \mathrm{VaR}_{\mathrm{short}}(c) &= -NS(t)\left[1 - \exp\left(\mu\delta t - Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right] \qquad (21.15)\end{aligned}$$

These two VaRs are *not* equal in magnitude! Although the probability distribution used in the derivation, namely the standard normal distribution $\mathrm{N}(0,1)$ has a density function symmetric about 0, the risk of a long position is different to that of a short position! This is firstly due to the drift and secondly to the fact that the normal distribution refers to the distribution of the *logarithmic* (i.e. *relative*) price changes of the underlying, as opposed to the VaR's in Equation 21.15 which contains the *log*normally distributed (*absolute*) price changes themselves. The lognormal distribution is not symmetric (see Figure 3.9). The approximations needed to make the risks of a long and short position equal will be discussed in greater detail in the next section.

21.3 Approximations in the Distribution of Risk Factors

For short liquidation periods (for example, $\delta t = 10$ days $= 0.0274$ years), the exponential function in Equation 21.8 is frequently approximated by its Taylor series[3] expanded up to and including its linear term. In addition, the contribution of the drift is often assumed to be negligible in comparison to the volatility in such short periods and is frequently excluded from consideration (for example by RiskMetrics$^{\mathrm{TM}}$). A third possibility, also frequently observed in practice is to combine the above two approximations, neglecting the drift and approximating the exponential function with

Applying N^{-1} to the symmetry equation $\mathrm{N}(-x) = 1 - \mathrm{N}(x)$ gives

$$\begin{aligned}&\mathrm{N}^{-1}\left[\mathrm{N}(-x)\right] = \mathrm{N}^{-1}\left[1 - \mathrm{N}(x)\right] \\ &\Leftrightarrow \\ &-x = Q_{1-\mathrm{N}(x)}\end{aligned}$$

Substituting c for $\mathrm{N}(x)$ and Q_c for x immediately yields

$$-Q_c = Q_{1-c}$$

[3] $\exp(x) \approx 1 + x + \cdots$

its linear Taylor series expansion as well. These three approximations correspond to a change in the risk factor, $\delta S(t) = S(t+\delta t) - S(t)$, in the following three ways:

$$\delta S(t) \approx \begin{cases} S(t)\left[e^{\sigma\delta W} - 1\right] & \mu \text{ neglected} \\ S(t)\left[\mu\delta t + \sigma\delta W\right] & \text{linear approximation for exp} \\ S(t)\sigma\delta W & \mu \text{ neglected and linear approximation} \end{cases} \tag{21.16}$$

We now show what effect these different approximations have on the value at risk of the portfolios treated above consisting of a long or a short position in a risk factor, respectively. The change in portfolio value δV induced by the risk factor changes are

$$\delta V_{\text{long}} = N\delta S(t)\ , \quad \delta V_{\text{short}} = -N\delta S(t)$$

for the long and short position, respectively. The approximations for $\delta S(t)$ thus induce an analogous approximation in δV. Consequently, the approximate value at risk of, for example, the long position can be derived as was done above yielding[4]

$$\text{VaR}_{\text{long}}(c) \approx \begin{cases} NS(t)\left[1 - \exp\left(Q_{1-c}^{\text{N}(0,1)}\sigma\sqrt{\delta t}\right)\right] & \mu \approx 0 \\ NS(t)\left[-\mu\delta t - Q_{1-c}^{\text{N}(0,1)}\sigma\sqrt{\delta t}\right] & \text{exp linear} \\ -NS(t)Q_{1-c}^{\text{N}(0,1)}\sigma\sqrt{\delta t} & \mu \approx 0, \text{exp linear} \end{cases} \tag{21.17}$$

This follows also – and more directly – by approximating the exponential function in Equation 21.15. The value at risk for the short position can be approximated in much the same way, the only difference being that the signs of N and $Q_{1-c}^{\text{N}(0,1)}$ are now negative. This implies that, even in linear approximation, the VaRs of a long and short position are not exactly equal:

$$\begin{aligned} \text{VaR}_{\text{long}}(c) &\approx NS(t)Q_{1-c}^{\text{N}(0,1)}\sigma\sqrt{\delta t} + NS(t)\mu\delta t \\ \text{VaR}_{\text{short}}(c) &\approx NS(t)Q_{1-c}^{\text{N}(0,1)}\sigma\sqrt{\delta t} - NS(t)\mu\delta t \end{aligned} \tag{21.18}$$

From these two linear approximations for the value at risk it is clear that the two VaRs can only then be equal when the drift is neglected in the linear approximation!

And only in this case is the well-known *square root of time law* valid. This law states that the value at risk with respect to one liquidation period δt can be calculated from the value at risk with respect to another liquidation period $\delta t'$ by multiplying by the square root of the ratio of the liquidation periods. Correspondingly, only in this approximation can the value at risk at a given confidence level a be converted into a VaR with respect to another confidence a' by simply multiplying by the ratio of the boundaries of the respective confidence intervals. Both conversions find their expression in the following equation:

$$\text{VaR}(c', \delta t') \approx \frac{Q_{1-c'}}{Q_{1-c}}\sqrt{\frac{\delta t'}{\delta t}}\ \text{VaR}(c, \delta t) \tag{21.19}$$

[4]Note that $Q_{1-c}^{\text{N}(0,1)} < 0$ for all reasonable confidence levels c, see for instance Equation 21.12. Since for all usually considered time spans δt and reasonable values of μ the drift term is smaller than the volatility term, the value at risk is a positive number.

On the basis of this equation, the value at risk computed for given data obtained from, for example, RiskMetrics^{TM}, with 95% confidence and a liquidation period of 1 day, can be converted to the value at risk corresponding to a different confidence level and/or liquidation period (for instance, 99% and 10 days for CAD II). Equation 21.19 can be easily derived from the third approximation of the value at risk listed in 21.17 (for this approximation, the risk for a long and short position are equal).

$$\begin{aligned}\mathrm{VaR}(c', \delta t') &\approx -NS(t)Q_{1-c'}\sigma\sqrt{\delta t'} \\ &= -NS(t)\frac{Q_{1-c'}}{Q_{1-c}}Q_{1-c}\sigma\sqrt{\frac{\delta t'}{\delta t}\delta t} \\ &= \underbrace{-NS(t)Q_{1-c}\sigma\sqrt{\delta t}}_{\mathrm{VaR}(c,\delta t)}\frac{Q_{1-c'}}{Q_{1-c}}\sqrt{\frac{\delta t'}{\delta t}}\end{aligned}$$

Equation 21.19 is valid *only* in linear approximation *and* when the drift is neglected. Furthermore, Equation 21.19 holds only for portfolios whose value changes are linear functions of the change in the underlying risk factor or when the linearity holds at least in good approximation. Examples are (as in the case currently under discussion) portfolios consisting of a single risk factor (the constant of proportionality is N) or the value at risk obtained by means of the delta-normal approximation of the variance-covariance method (see Section22) where the constant of proportionality is given by the portfolio delta.

21.4 The Covariance Matrix

In general, the value of the portfolio V depends on not only one but on several (often hundreds of) risk factors. We assume that these risk factors $S_i(t)$, $i = 1, \ldots, n$ can be modelled by random walks as in Equations 21.7, 3.13 or Equation 3.22 and obey the *coupled* stochastic differential equation

$$d \ln S_i(t) = \mu_i dt + dZ_i \tag{21.20}$$

where the dZ_i denote Brownian motions which are *correlated* with each other. If the random walks governing the risk factors were uncorrelated, their behavior could be completely determined through the specification of two parameters per risk factor, namely the drift and the volatility. The fluctuation (the variance) of a risk factor, which is of particular importance in risk management, could be described by the volatility alone. Since underlying prices, however, are in general correlated and the variance of the entire portfolio (a linear combination of the risk factors) is of interest, it is insufficient to specify a single variance for each risk factor. It is evident from Equation A.17, for example, that the *covariance* is needed to incorporate both the variance of each risk factor as well as the *correlation* between each pair of risk factors into a model which completely describes the value changes of the portfolio. These variances and correlations of the logarithm of the relative changes of each of the risk factors can be quite clearly represented by a matrix whose ij^{th} element is composed of the product

of the correlation between the i^{th} and j^{th} risk factor with the standard deviations of both these risk factors:

$$\delta\mathbf{\Sigma} = \begin{pmatrix} \delta\Sigma_{11} & \delta\Sigma_{12} & \cdots & \cdots & \delta\Sigma_{1n} \\ \delta\Sigma_{21} & \ddots & & \ddots & \delta\Sigma_{2n} \\ \vdots & & \delta\Sigma_{ij} & & \vdots \\ \vdots & \ddots & & \ddots & \vdots \\ \delta\Sigma_{n1} & \delta\Sigma_{n2} & \cdots & \cdots & \delta\Sigma_{nn} \end{pmatrix}$$

$$\text{where } \delta\Sigma_{ij} = \underbrace{\rho_{ij}}_{\substack{\text{Correlation} \\ \text{of } S_i \text{ with } S_j}} \underbrace{\sigma_i\sqrt{\delta t}}_{\substack{\text{Std. Dev.} \\ \text{of } S_i}} \underbrace{\sigma_j\sqrt{\delta t}}_{\substack{\text{Std. Dev.} \\ \text{of } S_j}} \text{ for } \; i,j = 1,2,\ldots,n \tag{21.21}$$

Here, δt is the reference time interval for the change in the risk factors, i.e. $\delta t = 1$ day for a daily change in the risk factors, $\delta t = 25$ days for a monthly change, etc. Matrices of this sort are called *covariance matrices*. With this matrix we can write

$$d\mathbf{Z} = \begin{pmatrix} dZ_1 \\ dZ_2 \\ \vdots \\ dZ_n \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, d\mathbf{\Sigma})$$

where the following general (vector) notation has been introduced

$$\begin{aligned} &\mathbf{X} \sim \mathrm{N}(\mathbf{R}, \mathbf{V}) \Longleftrightarrow \\ &X_i \text{ follow a normal distribution with} \\ &\mathrm{cov}[X_i, X_j] = V_{ij} \;,\;\; \mathrm{E}\,[X_i] = R_i \end{aligned}$$

The notation $d\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}dt)$ denotes that the vector $d\mathbf{Z}$ consists of the stochastic differentials dZ_i and that these are multivariate normally distributed with expectation 0 and covariance matrix $d\mathbf{\Sigma}$:

$$\begin{aligned} d\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, d\mathbf{\Sigma}) \implies \mathrm{cov}[dZ_i, dZ_j] = d\Sigma_{ij} \;,\;\; \mathrm{E}\,[dZ_i] = 0 \\ \text{where} \;\; d\Sigma_{ij} = \sigma_i\rho_{ij}\sigma_j dt \end{aligned} \tag{21.22}$$

As was the case for a single risk factor, the S_i are solutions to the associated stochastic differential equations 21.20. Analogously to Equation 3.25, theses solutions are given by the following stochastic processes

$$S_i(t+\delta t) = S_i(t)\exp\left(\mu_i\delta t + \delta Z_i\right) \;\text{ with } \delta\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \delta\mathbf{\Sigma}) \tag{21.23}$$

This equation specifies the distribution of the market parameter $S_i(t+\delta t)$ after a finite time interval δt has passed. The value $S_i(t)$ is known at time t as is the distribution of the random vector $\delta\mathbf{Z}$. The δZ_i are multivariate normally distributed with the covariance matrix

$$\delta\Sigma_{ij} = \sigma_i\rho_{i,j}\sigma_j\delta t$$

as given in Equation 21.21. As opposed to Equation 21.20, which holds for infinitesimal time steps dt, the finite time interval of length δt is arbitrary and can be taken to be the liquidation period used for the computation of the VaR, for example.

For short time intervals δt, the changes in the risk factors $\delta S_i(t) = S_i(t+\delta t) - S_i(t)$ can be approximated analogously as was done in Equation 21.16:

$$\delta S_i(t) \approx \begin{cases} S_i(t)\left[e^{\delta Z_i} - 1\right] & \text{Drift neglected} \\ S(t)\left[\mu_i \delta t + \delta Z_i\right] & \text{Linear approximation for exp} \\ S(t)\delta Z_i & \text{Drift neglected and linear approximation} \end{cases} \tag{21.24}$$

It is common practice to work with the latter and simplest of the three approximations above

$$\delta S_i(t) \approx S_i(t)\,\delta Z_i \quad \text{with} \quad \delta \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \delta\mathbf{\Sigma}) \tag{21.25}$$

21.4.1 Covariance Matrices of Data Providers

A price of the i^{th} risk factor is, of course, 100% correlated with itself. Observe that the correlation of two risk factors, say the i^{th} and j^{th} risk factors, is symmetric in the sense of Equation A.15. This symmetry makes almost half the information appearing in Equation 21.21 redundant. Only the "triangle" below the diagonal is needed. The matrix elements in the diagonal, i.e. those whose row index equals their column index, are composed of the volatilities since, as was mentioned above, the correlation of a risk factor with itself is always 1. Thus RiskMetrics$^{\text{TM}}$, for example, only makes the data in the "triangle" below the main diagonal of the correlation matrix (without the volatilities) available rather than the entire covariance matrix. Volatilities are not placed directly at the user's disposal. Instead, the standard deviations of the logarithmic change in the risk factors over $\delta t = 1$ day or $\delta t = 25$ days multiplied by 1.65 are given. Also, the current risk factor levels (called the *current levels*) are available. In summary, the following data can be obtained for both 1 day and 25 days:

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ \rho_{2,1} & \ddots & \ddots & & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & & 0 \\ \rho_{i,1} & & \rho_{i,i-1} & 1 & \ddots & \vdots \\ \vdots & \ddots & & \ddots & \ddots & 0 \\ \rho_{n,1} & \cdots & \rho_{n,i-1} & \cdots & \rho_{n,n-1} & 1 \end{pmatrix} \tag{21.26}$$

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \\ \vdots \\ Y_{n-1} \\ Y_n \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ \\ \vdots \\ S_{n-1} \\ S_n \end{pmatrix}$$

$\mathbf{Y}$ is a vector whose components are not the volatilities but the *standard deviations* of either daily ($\delta t = 1$ day) or monthly ($\delta t = 25$ days) logarithmic risk factor fluctuations multiplied by 1.65 given explicitly by

$$Y_i = -Q_{1-c}\sqrt{\text{var}\,[x_i]} = -Q_{1-c}\sqrt{\delta t}\,\sigma_i \qquad (21.27)$$
$$\text{where} \quad Q_{1-c} = -1.65 \Rightarrow c = 95\%$$

In order to obtain the actual volatilities, these Y_i need to be divided by 1.65 and $\sqrt{\delta t}$. The variables Y_i are referred to by RiskMetrics$^{\text{TM}}$ as the "value at risk" of the underlying (risk factor) S_i. In fact, neglecting the drift as in Equation 21.17, Y_i can be used to directly calculate the value at risk of a risk factor at a confidence level of $c = 95\%$ (see Equation 21.12):

$$\text{VaR}_i(95\%, \delta t) \approx S_i(t) - e^{-Y_i} S_i(t) \approx S_i(t) Y_i \qquad (21.28)$$

Thus interpreted, Y_i is a linear approximation (also neglecting the effect of the drift) of the *relative* value at risk of the i^{th} risk factor:

$$Y_i = \frac{\text{VaR}_i(95\%, \delta t)}{S_i(t)}$$

A linear approximation of the value at risk with respect to *another* confidence level and/or *another* liquidation period can be obtained (neglecting the drift) via 21.19.

Instead of relying on the volatilities and correlations of an external provider, these parameters can be determined by analyzing historical time series. The volatilities can also be implicitly calculated from prices of liquid options if such options are available for trade.

21.4.2 Cholesky Decomposition of the Covariance Matrix

As we will soon see, it is often quite convenient to work with a matrix which, when multiplied by itself, yields the covariance matrix 21.21. The need for such a matrix arises when applying several of the most important methods used in computing risk, such as Monte Carlo or the delta-gamma methods. This "square root" of the covariance matrix will be denoted by $\mathbf{A}$ and satisfies the condition

$$\mathbf{A}\mathbf{A}^T = \delta\mathbf{\Sigma} \qquad (21.29)$$

Here the notation $\mathbf{A}^T$ denotes the *transpose* of the matrix $\mathbf{A}$. We obtain the transpose by writing the row vectors of $\mathbf{A}$ as column vectors, i.e.

$$\left(\mathbf{A}^T\right)_{ij} = A_{ji}$$

Transformation of Uncorrelated Random Variables into Correlated Random Variables

A useful property of $\mathbf{A}$ (particularly for Monte Carlo simulations, see Section 23.1) is that it transforms uncorrelated random variables into correlated ones (with covariance $\delta\mathbf{\Sigma}$). Let $X_i, i = 1, \ldots, n$ be uncorrelated standard normally distributed random variables

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ with } \mathbf{X} \sim \mathrm{N}(\mathbf{0}, \mathbf{1}) \tag{21.30}$$

The symbol $\mathbf{1}$ is used to denote the *identity matrix*

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{21.31}$$

or equivalently[5], $(\mathbf{1})_{ij} = \delta_{ij}$. The notation in Equation 21.30 means explicitly:

$$\mathrm{cov}[X_i, X_j] = \delta_{ij}, \;\; \mathrm{E}[X_i] = 0 \;\; \forall i, j = 1, \ldots n$$

Applying the matrix $\mathbf{A}$ to $\mathbf{X}$ generates new random variables $\mathbf{Y}$ defined by

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \Longleftrightarrow Y_i = \sum_k A_{ik} X_k$$

The covariances of these new random variables are

$$\begin{aligned} \mathrm{cov}[Y_i, Y_j] &= \mathrm{cov}[\sum_k A_{ik} X_k, \sum_m A_{jm} X_m] \\ &= \sum_k A_{ik} \sum_m A_{jm} \underbrace{\mathrm{cov}[X_k, X_m]}_{\delta_{km}} \\ &= \sum_k A_{ik} A_{jk} \\ &= \left(\mathbf{A}\mathbf{A}^T\right)_{ij} \\ &= \delta\Sigma_{ij} \end{aligned}$$

where the defining property of the matrix $\mathbf{A}$, Equation 21.29, was used in the last step. The expectations of these random variables are

$$\mathrm{E}[Y_i] = \mathrm{E}[\sum_k A_{ik} X_k] = \sum_k A_{ik} \underbrace{\mathrm{E}[X_k]}_{0} = 0$$

[5]Here, we make use of the so called *Kronecker delta* often appearing in the science. It is defined as

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

This means that the random vector with components Y_i is multivariate normally distributed with covariance matrix $\delta\mathbf{\Sigma}$ and zero expectation:

$$\mathbf{AX} = \mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \delta\mathbf{\Sigma}) \tag{21.32}$$

Transformation of Correlated Random Variables into Uncorrelated Random Variables

Conversely, it is often convenient (for example, when using the delta-gamma method, see Section 22.3) to transform correlated random variables into uncorrelated ones. As we might suspect, this is accomplished through the mapping given by the inverse $\mathbf{A}^{-1}$ of the matrix $\mathbf{A}$. The proof is somewhat more involved than that given in the previous section: Let $Y_i, i = 1, \ldots, n$ be correlated multivariate normally distributed random variables with covariance $\delta\mathbf{\Sigma}$,

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ with } \mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \delta\mathbf{\Sigma})$$

or more explicitly

$$\mathrm{cov}[Y_i, Y_j] = \delta\Sigma_{ij} \ , \ \ \mathrm{E}[Y_i] = 0 \ \ \forall i, j = 1, \ldots n$$

The random variables X_i obtained by applying the inverse matrix $\mathbf{A}^{-1}$ to the random vector $\mathbf{Y}$ are given by

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y} \Longleftrightarrow X_i = \sum_k (A^{-1})_{ik} Y_k$$

The covariance of these X_i can be calculated as follows

$$\begin{aligned} \mathrm{cov}[X_i, X_j] &= \mathrm{cov}[\sum_k (A^{-1})_{ik} Y_k, \sum_m (A^{-1})_{jm} Y_m] \\ &= \sum_k (A^{-1})_{ik} \sum_m (A^{-1})_{jm} \underbrace{\mathrm{cov}[Y_k, Y_m]}_{\delta\Sigma_{km}} \\ &= \sum_k \sum_m (A^{-1})_{ik} \delta\Sigma_{km} (A^{-1})^T_{mj} \end{aligned}$$

This, however, is precisely the *matrix product*[6] of the matrices $\mathbf{A}^{-1}$, $\delta\mathbf{\Sigma}$ and $(\mathbf{A}^{-1})^T$. Decomposing the matrix $\delta\mathbf{\Sigma}$ as in Equation 21.29 yields

$$\begin{aligned} \mathrm{cov}[X_i, X_j] &= \left(\mathbf{A}^{-1}\delta\mathbf{\Sigma}(\mathbf{A}^{-1})^T\right)_{ij} \\ &= \left(\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^T(\mathbf{A}^{-1})^T\right)_{ij} \end{aligned}$$

[6] A short and simple overview of *matrix algebra* can be found in [65], for example.

The product $\mathbf{A}^{-1}\mathbf{A}$ of first two matrices appearing on the right hand side is by definition of *inverse matrices*, of course, the identity matrix given in Equation 21.31. Furthermore, the following property holds for every invertible matrix[7]

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \tag{21.33}$$

allowing the covariances to be calculated as

$$\begin{aligned}\mathrm{cov}[X_i, X_j] &= \left(\mathbf{A}^T(\mathbf{A}^{-1})^T\right)_{ij} \\ &= \left(\mathbf{A}^T(\mathbf{A}^T)^{-1}\right)_{ij} \\ &= (\mathbf{1})_{ij} \\ &= \delta_{ij}\end{aligned}$$

The expectations of the new random variables remain zero since, by the linearity of the expectation,

$$\mathrm{E}[X_i] = \mathrm{E}[\sum_k (A^{-1})_{ik} Y_k] = \sum_k (A^{-1})_{ik} \underbrace{\mathrm{E}[Y_k]}_{0} = 0$$

Since matrix multiplication is a *linear* transformation, the form of the distribution remains the same as well. In summary, the new random variables X_i are uncorrelated and all have the same standard normal distribution:

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \mathbf{1}) \tag{21.34}$$

However, uncorrelated random variables are in general not *independent*. For independence, the joint distribution of the variables has to be equal to the product of the distributions of each individual variable, see Equation A.62. Fortunately for uncorrelated standard normally distributed random variables this is indeed the case, see Equation A.61. Therefore, the X_i are *independent*. They are also all governed by the *same* distribution (in this case the standard normal distribution). Therefore the X_i are called *independent, identically distributed* random variables or *iid,* for short.

The Cholesky Decomposition

Having shown the usefulness of a matrix $\mathbf{A}$ satisfying the property in Equation 21.29, we proceed with the explicit construction of this matrix. In Equation 12.4 for $n = 2$, it has already been shown how the square root of a matrix can be calculated by means of the *Cholesky decomposition*. For arbitrarily large matrices ($n > 2$), the components A_{ij}, $i, j = 1, ..., n$ of the matrix $\mathbf{A}$ are obtained by iteratively solving the following

[7] Since $\mathbf{A}$ is invertible we have: $(\mathbf{A}\mathbf{A}^{-1})^T = \mathbf{1} \Rightarrow (\mathbf{A}^{-1})^T \mathbf{A}^T = \mathbf{1} \Rightarrow (\mathbf{A}^{-1})^T \underbrace{\mathbf{A}^T(\mathbf{A}^T)^{-1}}_{\mathbf{1}} = \mathbf{1}(\mathbf{A}^T)^{-1}$

system of equations:

$$A_{ji} = \begin{cases} 0 & \text{for } j < i \\ \sqrt{\delta\Sigma_{ii} - \sum_{k=1}^{i-1} A_{ik}^2} & \text{for } j = i \\ \frac{1}{A_{ii}} \left(\delta\Sigma_{ji} - \sum_{k=1}^{i-1} A_{ik}\, A_{jk} \right) & \text{for } j > i \end{cases} \tag{21.35}$$

with $\delta\Sigma_{ji}$ as given in Equation 21.21:

$$\delta\Sigma_{ji} = \begin{cases} \sigma_i^2 \delta t & \text{for } j = i \\ \rho_{ij}\, \sigma_i \sigma_j \delta t & \text{for } j \neq i \end{cases}$$

In solving the above system, we begin with $i = 1$, $j = 1$. Within a given Index i we first solve for all possible indices j before increasing i. This procedure is illustrated here for $n = 3$:

$$A_{11} = \sqrt{\delta\Sigma_{11}}\ , \quad A_{21} = \frac{\delta\Sigma_{21}}{A_{11}} = \frac{\delta\Sigma_{21}}{\sqrt{\delta\Sigma_{11}}}, \quad A_{31} = \frac{\delta\Sigma_{31}}{A_{11}} = \frac{\delta\Sigma_{31}}{\sqrt{\delta\Sigma_{11}}}$$

$$A_{22} = \sqrt{\delta\Sigma_{22} - A_{21}^2} = \sqrt{\delta\Sigma_{22} - \delta\Sigma_{21}^2/\delta\Sigma_{11}}$$

$$A_{32} = \frac{\delta\Sigma_{32} - A_{21}A_{31}}{A_{22}} = \frac{\delta\Sigma_{32} - \delta\Sigma_{21}\delta\Sigma_{31}/\delta\Sigma_{11}}{\sqrt{\delta\Sigma_{22} - \delta\Sigma_{21}^2/\delta\Sigma_{11}}}$$

$$A_{33} = \sqrt{\delta\Sigma_{33} - A_{31}^2 - A_{32}^2} = \sqrt{\delta\Sigma_{33} - \frac{\delta\Sigma_{31}^2}{\delta\Sigma_{11}} - \frac{(\delta\Sigma_{32} - \delta\Sigma_{21}\delta\Sigma_{31}/\delta\Sigma_{11})^2}{\delta\Sigma_{22} - \delta\Sigma_{21}^2/\delta\Sigma_{11}}}$$

Thus, all elements of the matrix **A** have been expressed in terms of the elements of the covariance matrix. Finally, substituting for $\delta\Sigma_{ij}$ as defined in Equation 21.21 yields

$$A_{11} = \sigma_1\sqrt{\delta t}$$

$$A_{21} = \sigma_2\sqrt{\delta t}\,\rho_{21}\ , \quad A_{22} = \sigma_2\sqrt{\delta t}\,\sqrt{1-\rho_{21}^2}$$

$$A_{31} = \sigma_3\sqrt{\delta t}\,\rho_{31}\ , \quad A_{32} = \sigma_3\sqrt{\delta t}\,\frac{\rho_{32} - \rho_{31}\rho_{21}}{\sqrt{1-\rho_{12}^2}}$$

$$A_{33} = \sigma_3\sqrt{\delta t}\,\sqrt{1-\rho_{31}^2 - \frac{(\rho_{32} - \rho_{31}\rho_{21})^2}{1-\rho_{12}^2}}$$

The first equation above holds when only a single risk factor is involved. The first *two* equations correspond to Equation 12.4 for *two* correlated random walks. The third and fourth rows complete the decomposition for an additional *third* correlated random walk.

22

The Variance-Covariance Method

The *variance-covariance method* makes use of covariances (volatilities and correlations) of the risk factors and the sensitivities of the portfolio values with respect to these risk factors with the goal of approximating the value at risk. This method leads directly to the final result, i.e. the portfolio's value at risk; no information regarding market scenarios arises. The variance-covariance method utilizes linear approximations of the risk factors themselves throughout the entire calculation, often neglecting the drift as well. In view of Equation 21.24, we have

$$\delta S_i(t) \approx S_i(t) \left[\mu_i \delta t + \delta Z_i\right] \approx S_i(t) \delta Z_i \tag{22.1}$$

The main idea characterizing this method, however, is that the portfolio value V is expanded in its Taylor series as a function of its risk factors $S_i, i = 1, \ldots n$, and approximated by breaking off after the first or second order term. Let

$$\mathbf{S}(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ S_n(t) \end{pmatrix}$$

denote the vector of risk factors. The Taylor expansion for the change in portfolio value $\delta V(\mathbf{S})$ up to second order is

$$\begin{aligned} \delta V(\mathbf{S}(t)) &= V(\mathbf{S}(t) + \delta \mathbf{S}(t)) - V(\mathbf{S}(t)) \\ &\approx \sum_i^n \frac{\partial V}{\partial S_i} \delta S_i(t) + \frac{1}{2} \sum_{i,j}^n \delta S_i(t) \frac{\partial^2 V}{\partial S_i \partial S_j} \delta S_j(t) \\ &= \sum_i^n \Delta_i \delta S_i(t) + \frac{1}{2} \sum_{i,j}^n \delta S_i(t) \Gamma_{ij} \delta S_j(t) \end{aligned}$$

$$\approx \sum_i^n \widetilde{\Delta}_i \, [\mu_i \delta t + \delta Z_i] + \frac{1}{2} \sum_{i,j}^n [\mu_i \delta t + \delta Z_i] \, \widetilde{\Gamma}_{ij} \, [\mu_j \delta t + \delta Z_j]$$

$$\approx \sum_i^n \widetilde{\Delta}_i \, \delta Z_i + \frac{1}{2} \sum_{i,j}^n \delta Z_i \widetilde{\Gamma}_{ij} \, \delta Z_j \tag{22.2}$$

The first "approximately equal" sign appears due to having broken off the Taylor series of the portfolio value, the second as a result of the linear approximation of the risk factors in accordance with Equation 22.1, and finally, in the last step, because the drift has been neglected. The last line in 22.2 is referred to as the *delta-gamma approximation.* An analogous approach leads to the *delta approximation*, for which the Taylor series in the above derivation is taken up to linear order only, resulting in an approximation solely consisting of the first of the two sums appearing in the last equation in 22.2.

The abbreviations Δ_i and Γ_{ij}, as usual, denote the *sensitivities* (at time t) of V with respect to the risk factors

$$\Delta_i := \frac{\partial V}{\partial S_i} \ , \ \ \Gamma_{ij} := \frac{\partial^2 V}{\partial S_i \partial S_j} \ , \ \ i, j = 1, \ldots n$$

Note here that the mixed partial derivatives arise in the expression for Γ_{ij}. In the literature, the matrix Γ_{ij} is sometimes called the *Hessian matrix*.

We will see below that the sensitivities usually appear in connection with the current levels $S_i(t)$ and $S_j(t)$. The notation $\widetilde{\Delta}_i$ and $\widetilde{\Gamma}_{ij}$ will be used to denote these sensitivities multiplied by the current levels:

$$\widetilde{\Delta}_i := S_i(t) \frac{\partial V}{\partial S_i} \ , \ \ \widetilde{\Gamma}_{ij} := S_i(t) S_j(t) \frac{\partial^2 V}{\partial S_i \partial S_j} \tag{22.3}$$

Using $\widetilde{\Delta}_i$ and $\widetilde{\Gamma}_{ij}$ will prove to substantially simplify the notation. Interpreting the $\widetilde{\Delta}_i$ as components of a vector $\widetilde{\mathbf{\Delta}}$, and $\widetilde{\Gamma}_{ij}$ as the elements of a matrix $\widetilde{\mathbf{\Gamma}}$, Equation 22.2 can be written in vector form as

$$\begin{aligned}
\delta V(\mathbf{S}(t)) &= \begin{pmatrix} \widetilde{\Delta}_1 & \cdots & \widetilde{\Delta}_n \end{pmatrix} \begin{pmatrix} \delta Z_1 \\ \vdots \\ \delta Z_n \end{pmatrix} \\
&\quad + \frac{1}{2} \begin{pmatrix} \delta Z_1 & \cdots & \delta Z_n \end{pmatrix} \begin{pmatrix} \widetilde{\Gamma}_{1,1} & \cdots & \widetilde{\Gamma}_{1,n} \\ \vdots & \ddots & \vdots \\ \widetilde{\Gamma}_{n,1} & \cdots & \widetilde{\Gamma}_{n,n} \end{pmatrix} \begin{pmatrix} \delta Z_1 \\ \vdots \\ \delta Z_n \end{pmatrix} \\
&= \widetilde{\mathbf{\Delta}}^T \delta \mathbf{Z} + \frac{1}{2} \delta \mathbf{Z}^T \widetilde{\mathbf{\Gamma}} \delta \mathbf{Z}
\end{aligned}$$

The approximation of $V(\mathbf{S})$ through its Taylor series expansion up to second order is presented in Figure 22.1 for a *straddle* (a portfolio made up of a call and a put option) on a risk factor S. The figure has been extracted from the accompanying Excel

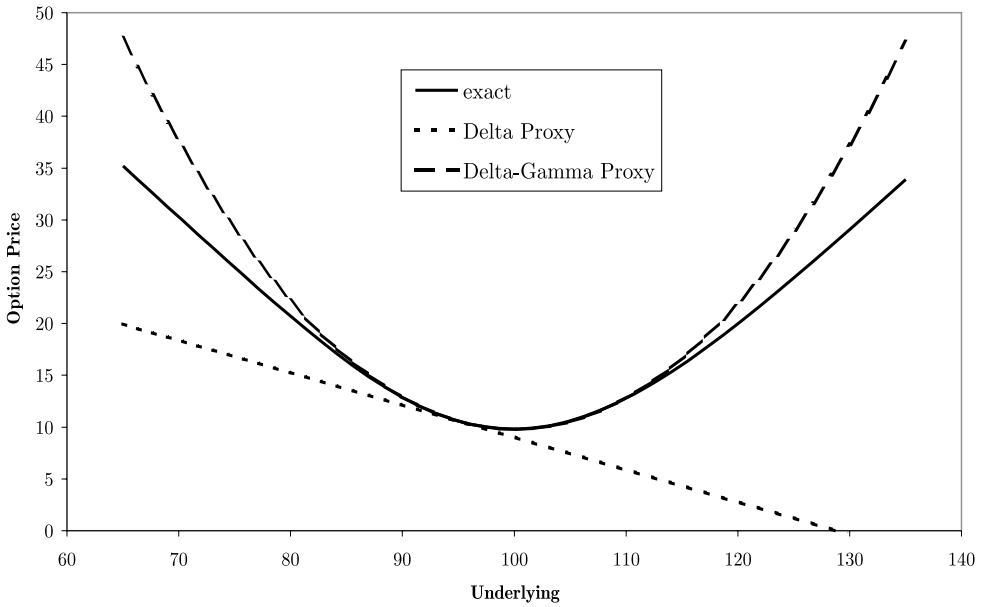


Figure 22.1: Black-Scholes price of a straddle (strike = 100, time to maturity = 1 year) on an underlying S (dividend yield 6%, repo rate 3%). The dashed line is the delta-gamma proxy, the dotted line is the simple delta proxy. The Taylor expansion was done about $S = 95$.

workbook STRADDLE.XLS. We can recognize that the delta-gamma approximation for a simple payoff profile is quite a good approximation. For somewhat more complicated portfolios, however, the delta-gamma approximation fails to be a reasonable representation of the payoff profile. In such cases, we recommend simulation with the *Monte Carlo method* as described in Section 23.1. However, the disadvantage of the simulation approach is of course the large number of numerical computations necessary.

22.1 Portfolios vs. Financial Instruments

Although we continually refer to *portfolio* sensitivities, the same results hold for individual *financial instruments* as well. In fact, sensitivity of a portfolio composed of financial instruments on the same underlying as described in Section 13.5.1 can be obtained by simply adding together the sensitivities of the individual instruments. This is the approach most commonly taken when calculating portfolio sensitivities. For the sake of clarity, we once again present this method explicitly here.

Consider a portfolio with a value V consisting of M different financial instruments with values V_k , $k = 1, \ldots M$. N_k denotes the number of each instrument with value V_k held in the portfolio. The total value of the portfolio is naturally the sum of the

values of each individual position

$$V(t) = \sum_{k=1}^{M} N_k V_k(\mathbf{S}(t)),$$

The change in value $\delta V(t)$ of this portfolio is (approximated up to second order)

$$\begin{aligned}
\delta V(\mathbf{S}(t)) &= \sum_{k=1}^{M} N_k \delta V_k(\mathbf{S}(t)) \\
&\approx \sum_{k=1}^{M} N_k \left[\sum_{i}^{n} \frac{\partial V_k}{\partial S_i} \delta S_i(t) + \frac{1}{2} \sum_{i,j}^{n} \delta S_i(t) \frac{\partial^2 V_k}{\partial S_i \partial S_j} \delta S_j(t) \right] \\
&= \sum_{k=1}^{M} N_k \sum_{i}^{n} \Delta_i^k \delta S_i(t) + \frac{1}{2} \sum_{k=1}^{M} N_k \sum_{i,j}^{n} \delta S_i(t) \Gamma_{ij}^k \delta S_j(t)
\end{aligned}$$

Where the sensitivities of the *financial instruments* have been introduced in the last step. For example, Δ_i^k is the linear sensitivity of the k^{th} financial instrument in the portfolio with respect to the i^{th} risk factor, etc.:

$$\Delta_i^k := \frac{\partial V_k}{\partial S_i} \ , \ \ \Gamma_{ij}^k := \frac{\partial^2 V_k}{\partial S_i \partial S_j} \ , \ \ i,j = 1, \dots n \ ; \ \ k = 1, \dots M$$

Simply rearranging the terms makes it clear that the summing over the index k (which denotes the different financial instruments) yields the portfolio sensitivities:

$$\begin{aligned}
\delta V(\mathbf{S}(t)) &= \sum_{i}^{n} \delta S_i(t) \sum_{k=1}^{M} N_k \Delta_i^k + \frac{1}{2} \sum_{i,j}^{n} \delta S_i(t) \delta S_j(t) \sum_{k=1}^{M} N_k \Gamma_{ij}^k \\
&= \sum_{i}^{n} \delta S_i(t) \Delta_i + \frac{1}{2} \sum_{i,j}^{n} \delta S_i(t) \delta S_j(t) \Gamma_{ij}
\end{aligned}$$

Thus, a portfolio sensitivity like Δ_i, for example, contains the sensitivities of all instruments in the portfolio (including all the position sizes N_k) with respect to the considered risk factor. This then yields (as is intuitively clear) the sensitivities of the entire portfolio as sums over the sensitivities of all instruments in the portfolio:

$$\Delta_i = \sum_{k=1}^{M} N_k \Delta_i^k \ , \ \ \Gamma_{ij} = \sum_{k=1}^{M} N_k \Gamma_{ij}^k$$

In practice, this procedure is usually referred to as *position mapping*. Using the approximation in Equation 22.1 for the change in risk factor, we finally obtain an expression

for the portfolio's change in value as

$$\begin{aligned}\delta V(\mathbf{S}(t)) &\approx \sum_i^n \widetilde{\Delta}_i \, [\mu_i \delta t + \delta Z_i] + \frac{1}{2} \sum_{i,j}^n [\mu_i \delta t + \delta Z_i] \, \widetilde{\Gamma}_{ij} \, [\mu_j \delta t + \delta Z_j] \\ &\approx \sum_i^n \widetilde{\Delta}_i \, \delta Z_i + \frac{1}{2} \sum_{i,j}^n \delta Z_i \widetilde{\Gamma}_{ij} \, \delta Z_j\end{aligned}$$

using the modified portfolio sensitivities defined as in Equation 22.3:

$$\begin{aligned}\widetilde{\Delta}_i &:= S_i(t)\Delta_i = S_i(t) \sum_{k=1}^M N_k \Delta_i^k \\ \widetilde{\Gamma}_{ij} &:= S_i(t)\Gamma_{ij} = S_i(t)S_j(t) \sum_{k=1}^M N_k \Gamma_{ij}^k\end{aligned}$$

Adding the sensitivities of the financial instruments and multiplying by the current levels of the risk factors to obtain these modified portfolio sensitivities is sometimes referred to as *VaR mapping*.

22.2 The Delta-Normal Method

In the *delta-normal method,* the Taylor series 22.2 of the portfolio value is broken off after the linear term.

$$\begin{aligned}\delta V(\mathbf{S}(t)) &\approx \sum_i^n \frac{\partial V}{\partial S_i} \delta S_i(t) \\ &\approx \sum_i^n \widetilde{\Delta}_i \, [\mu_i \delta t + \delta Z_i] \\ &\approx \sum_i^n \widetilde{\Delta}_i \, \delta Z_i = \widetilde{\boldsymbol{\Delta}}^T \delta \mathbf{Z}\end{aligned} \tag{22.4}$$

22.2.1 The Value at Risk with Respect to a Single Risk Factor

For a single risk factor this means

$$\delta V(S(t)) \approx \Delta \delta S(t) \tag{22.5}$$

with the sensitivity $\Delta := \partial V / \partial S$. The change in the portfolio's value is thus approximated to be a linear function of the change in the underlying risk factor. This corresponds to the situation described in Section 21.2. There, the constant of proportionality (the sensitivity) was not Δ but N or $-N$ for a long or short position, respectively. The linear approximation implies intuitively that a portfolio with a linear sensitivity Δ with respect to a risk factor can be interpreted as a portfolio consisting

of Δ risk factors. The only subtlety in this argumentation is that in Section 21.2, we distinguished between a long and a short position, treating the two cases differently on the basis of whether the proportionality constant N was positive or negative (which leads to the two different VaRs in Equation 21.15). However, we can not know a priori whether Δ is greater than or less than 0. We do know, however, that V is linear and in consequence, a monotone function of S. Therefore, in the sense of Equation 21.6, the results following from Equation 21.15 hold with the correspondence $\Delta \widehat{=} N$ for $\Delta > 0$, and $\Delta \widehat{=} -N$ for $\Delta < 0$. Making use of this fact allows us to write

$$\begin{aligned} \mathrm{VaR}_V(c) \approx \max\{&\widetilde{\Delta}\left[1 - \exp\left(\mu\delta t + Q_{1-c}\sigma\sqrt{\delta t}\right)\right], \\ &\widetilde{\Delta}\left[1 - \exp\left(\mu\delta t - Q_{1-c}\sigma\sqrt{\delta t}\right)\right]\} \end{aligned} \tag{22.6}$$

using the notation $\widetilde{\Delta} := S(t)\Delta = S(t)\partial V/\partial S$. As usual, Q_{1-c} is the $(1-c)$ percentile of the standard normal distribution[1]. The maximum function in Equation 21.6 effects the correct choice for the VaR. If $\Delta > 0$, the *lower bound* of the confidence interval of the risk factor[2] is relevant and consequently the VaR function as defined above takes on the value corresponding to a *long* position in Equation 21.15. Likewise for $\Delta < 0$, the *upper bound* of the confidence interval of the risk factor[3] is relevant and the above defined maximum function takes on the value corresponding to the VaR of the short position in Equation 21.15.

In all our deliberations up to this point, only the portfolio value has been approximated with Equation 22.5. The change in the *risk factor* in Equation 22.6 is still exact. Approximating this risk factor change with Equation 22.1, the VaR becomes

$$\begin{aligned} \mathrm{VaR}_V(c) \approx \max\{&\widetilde{\Delta}\left[-\mu\delta t - Q_{1-c}\sigma\sqrt{\delta t}\right], \\ &\widetilde{\Delta}\left[-\mu\delta t + Q_{1-c}\sigma\sqrt{\delta t}\right]\} \end{aligned}$$

This corresponds exactly to Equation 21.18, since $\widetilde{\Delta} \widehat{=} N$ for $\widetilde{\Delta} > 0$ and $\widetilde{\Delta} \widehat{=} -N$ for $\widetilde{\Delta} < 0$.

The common summand $-\widetilde{\Delta}\mu\delta t$ can now be taken out of the maximum function

$$\begin{aligned} \mathrm{VaR}_V(c) &\approx \max\left\{-\widetilde{\Delta}Q_{1-c}\sigma\sqrt{\delta t}, +\widetilde{\Delta}Q_{1-c}\sigma\sqrt{\delta t}\right\} - \widetilde{\Delta}\mu\delta t \\ &= \left|\widetilde{\Delta}Q_{1-c}\sigma\sqrt{\delta t}\right| - \widetilde{\Delta}\mu\delta t \end{aligned} \tag{22.7}$$

In this approximation, the maximum function produces precisely the absolute value of the risk which is caused by the volatility of the risk factor. A positive drift μ of the risk factor reduces the portfolio risk when $\widetilde{\Delta} > 0$ (intuitively, the portfolio then represents a long position). If, on the other hand, the portfolio sensitivity is negative, i.e. $\widetilde{\Delta} < 0$, a positive drift *increases* the portfolio risk (intuitively, the

[1] For all relevant confidence levels, this percentile is a negative number, see Equation 21.12.
[2] This corresponds to the percentile Q_{1-c} of the standard normal distribution.
[3] This corresponds to the percentile $-Q_{1-c} = Q_c$ of the standard normal distribution.

portfolio represents a short position). The drift's influence is of course lost if the drift is neglected in the approximation. The value at risk then reduces to

$$\mathrm{VaR}_V(c) \approx \left|\widetilde{\Delta} Q_{1-c}\sigma\sqrt{\delta t}\right| \tag{22.8}$$

where the absolute value makes it immediately clear that the sign of the portfolio sensitivity no long plays a role.

22.2.2 The Value at Risk with Respect to Several Risk Factors

In the previous section, linear approximations enabled us to reduce the VaR with respect to a single risk factor to that of a position consisting of Δ instruments representing the risk factor. We were then able, as was done in Section 21.2, to deduce information about the unknown distribution of the *portfolio's* value V from the known distribution of the *risk factor* S. The extension of these results to the case of several risk factors is not trivial even for the delta-normal approximation. Only by using the roughest approximation in Equation 22.1 for the change in the risk factors, namely $\delta S_i(t) \approx S_i(t)\delta Z_i$, can we manage to avoid involving the distribution of V in the discussion.

The approximation $\delta S(t) \approx S(t)\delta Z$ led to the value at risk equation 22.8 with respect to a single risk factor. Squaring both sides of this equation yields

$$\begin{aligned}
\mathrm{VaR}_V^2(c) &\approx \widetilde{\Delta}^2 \left(Q_{1-c'}\right)^2 \sigma^2 \delta t \\
&= \Delta^2 \left(Q_{1-c'}\right)^2 S(t)^2 \sigma^2 \delta t \\
&= \Delta^2 \left(Q_{1-c'}\right)^2 S(t)^2 \mathrm{var}\left[\delta Z\right] \\
&= \Delta^2 \left(Q_{1-c'}\right)^2 \mathrm{var}\left[\delta S(t)\right]
\end{aligned}$$

since the approximation in Equation 21.25 allows the approximation of the variance of $\delta S(t)$ with $S(t)^2\,\mathrm{var}[\delta Z]$. On the other hand, the variance of V can be calculated from Equation 22.5 simply as

$$\mathrm{var}\left[\delta V\right] \approx \mathrm{var}\left[\Delta \delta S(t)\right] = \Delta^2 \mathrm{var}\left[\delta S(t)\right]$$

This means that this approximation can be used to express the square of the value at risk in terms of a multiple of the *variance of the portfolio*:

$$\begin{aligned}
\mathrm{VaR}_V^2(c) &\approx \left(Q_{1-c'}\right)^2 \Delta^2 \mathrm{var}\left[\delta S(t)\right] \\
&= \left(Q_{1-c'}\right)^2 \mathrm{var}\left[\delta V\right]
\end{aligned}$$

Now, only the *variance* of the portfolio's value needs to be determined for the computation of the VaR and not its *distribution* or its percentiles.

If several risk factors are involved, Equation 22.4 can be used to write the portfolio's change in value, δV, in the approximation given by Equation 21.25, as a linear combination of normally distributed random variables δZ_i (with deterministic coefficients

$\widetilde{\Delta}_i$)

$$\delta V \approx \sum_{i=1}^{n} \widetilde{\Delta}_i \delta Z_i$$

The variance of a sum of random variables is equal to the sum of the covariances of these random variable as can be seen from Equation A.17. The variance of the portfolio value is thus

$$\begin{aligned} \text{var}\left[\delta V\right] &\approx \sum_{i,j=1}^{n} \widetilde{\Delta}_i \widetilde{\Delta}_j \text{cov}\left[\delta Z_i, \delta Z_j\right] \\ &= \sum_{i,j=1}^{n} \widetilde{\Delta}_i \delta \Sigma_{ij} \widetilde{\Delta}_j \\ &= \delta t \sum_{i,j=1}^{n} \widetilde{\Delta}_i \sigma_i \rho_{ij} \sigma_j \widetilde{\Delta_j} \end{aligned} \tag{22.9}$$

where the definition of the covariance matrix in Equation 21.21 was used in the last step. This means that the value at risk can be approximated as

$$\begin{aligned} \text{VaR}_V(c) &\approx |Q_{1-c}| \sqrt{\text{var}\left[\delta V\right]} \\ &= |Q_{1-c}| \sqrt{\widetilde{\boldsymbol{\Delta}} \delta \boldsymbol{\Sigma} \widetilde{\boldsymbol{\Delta}}} \\ &= |Q_{1-c}| \sqrt{\delta t} \sqrt{\sum_{i,j=1}^{n} \widetilde{\Delta}_i \sigma_i \rho_{ij} \sigma_j \widetilde{\Delta}_j} \end{aligned} \tag{22.10}$$

This is the central equation for the delta-normal method. It summarizes all assumptions, approximations, and computation methods of the delta-normal method.

If all portfolio sensitivities are non-negative[4], then this equation can be rewritten in an alternative and quite intuitive form. Let $\text{VaR}_i(c)$ denote the value at risk of the portfolio with respect to a particular risk factor S_i. Using Equation 22.8, we can approximate this by

$$\text{VaR}_i(c) \approx \left| \widetilde{\Delta}_i Q_{1-c} \sigma_i \sqrt{\delta t} \right|$$

For the special case that *none* of the portfolio deltas is negative[5], the portfolio VaR in the delta-normal method can also be expressed in terms of the VaRs of the individual risk factors:

$$\text{VaR}_V(c) \approx \sqrt{\sum_{i,j=1}^{n} \text{VaR}_i(c) \rho_{ij} \text{VaR}_j(c)} \quad \text{if} \;\; \widetilde{\Delta}_i \geq 0 \forall i \in \{1, \ldots, n\} \tag{22.11}$$

In this case the VaR with respect to all risk factors can thus be obtained by computing the square root of the weighted sum of the products of the VaRs with respect to the

[4]This is for instance the case for a portfolio containing only long positions in stocks.

[5]This is often the case for instance for a typical private investor's portfolio containing only long positions in stocks.

individual risk factors; the weights under consideration are the respective correlations between the risk factors.

In the linear approximation, the effect of the drifts can be subsequently introduced into the approximation. The *expected* change in portfolio values are calculated using the deltas and the drifts of the risk factors. Analogously to Equation 22.7 for a single risk factor, this expected change is then subtracted from the value at risk of the portfolio given in Equation 22.11:

$$\mathrm{VaR}_V(c) \approx |Q_{1-c}| \sqrt{\delta t} \sqrt{\sum_{i,j=1}^{n} \widetilde{\Delta}_i \sigma_i \rho_{ij} \sigma_j \widetilde{\Delta}_j} - \delta t \sum_i \widetilde{\Delta}_i \, \mu_i \tag{22.12}$$

The delta-normal approach to the calculation of the value at risk can be summarized as follows:

- Calculate the sensitivities of the portfolio with respect to all risk factors.
- Multiply the covariance matrix with the sensitivities of the portfolio and the current values of the risk factors as in Equation 22.9 to obtain the variance of the portfolio. The covariance matrix's elements consist of the product of volatilities and correlations of the risk factors as defined in Equation 21.21.
- Multiply the portfolio variance as in Equation 22.10 by the liquidation period and the square of the percentile corresponding to the desired confidence interval (for example, -2.326 for 99% confidence).
- The square root of the thus obtained number is the value at risk of the entire portfolio, neglecting the effect of the drifts of the risk factors.
- The effect of the drifts can be taken into account using Equation 22.12.

22.3 The Delta-Gamma Method

The delta-gamma method for calculating the portfolio's VaR makes use of the Taylor series expansion of the value of the portfolio up to and including the second order terms along with the approximation in Equation 22.1 for the risk factors. The starting point for the delta-gamma method is thus the last line in Equation 22.2, which when written in vector notation is given by

$$\delta V(\mathbf{S}(t)) = \widetilde{\boldsymbol{\Delta}}^T \delta \mathbf{Z} + \frac{1}{2} \delta \mathbf{Z}^T \widetilde{\boldsymbol{\Gamma}} \delta \mathbf{Z} \tag{22.13}$$

where

$$\delta \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \delta \boldsymbol{\Sigma}) \implies \mathrm{cov}[\delta Z_i, \delta Z_j] = \sigma_i \rho_{ij} \sigma_j \delta t \ , \ \mathrm{E}\left[\delta Z_i\right] = 0$$

The right-hand side of Equation 22.13 can *not* be written as the sum of the contributions of each risk factor as was the case for the delta-normal method:

$$\sum_i^n \widetilde{\Delta}_i \, \delta Z_i + \frac{1}{2} \sum_{i,j}^n \delta Z_i \widetilde{\Gamma}_{ij} \, \delta Z_j \neq \sum_i^n \left(\text{Contribution of the } i^{\text{th}} \text{ Risk Factor} \right)$$

The contributions of the individual risk factors can not be considered separately since they are *coupled* in the above equation by the matrix $\widetilde{\mathbf{\Gamma}}$. Furthermore the random variables δZ_j are not independent. They are *correlated* through the covariance matrix $\delta\mathbf{\Sigma}$. Two essential elements of the method presented here are[6]:

- Using the Cholesky decomposition of the covariance matrix $\delta\mathbf{\Sigma}$ to transform the δZ_j into independent random variables.
- Diagonalizing the gamma matrix $\widetilde{\mathbf{\Gamma}}$ thereby decoupling the contributions of the individual risk factors.

Furthermore, it is no longer possible in this situation, as in the delta-normal case, to reduce the VaR with respect to a risk factor to the VaR of a position consisting of Δ of these risk factors. Thus the (unknown) distribution of the *portfolio value* can no longer be substituted by the (known) distribution of the individual *risk factors*, as could still be done in Section 21.2. Therefore, the distribution of δV must be determined *directly* in order to calculate the value at risk defined in Equations 21.2 or 21.3. A third essential step of the delta-gamma method presented here involves the determination of the distribution of δV.

22.3.1 Decoupling of the Risk Factors

Motivated by Equation 21.34, we first introduce a matrix $\mathbf{A}$ satisfying the property 21.29. This matrix can be constructed through the Cholesky decomposition of the covariance matrix as described in detail in Section 21.4.2. This matrix transforms the correlated δZ_i into uncorrelated random variables. With this goal in mind, we rewrite Equation 22.13, first introducing identity matrices into the equation and then replacing them with $\mathbf{A}\mathbf{A}^{-1}$ or $(\mathbf{A}^T)^{-1}\mathbf{A}^T$ as follows:

$$\begin{aligned}
\delta V(\mathbf{S}(t)) &= \widetilde{\mathbf{\Delta}}^T\,\mathbf{1}\,\delta\mathbf{Z}+\frac{1}{2}\delta\mathbf{Z}^T\,\mathbf{1}\,\widetilde{\mathbf{\Gamma}}\,\mathbf{1}\,\delta\mathbf{Z}\\
&= \widetilde{\mathbf{\Delta}}^T\,\mathbf{A}\mathbf{A}^{-1}\,\delta\mathbf{Z}+\frac{1}{2}\delta\mathbf{Z}^T\,(\mathbf{A}^T)^{-1}\mathbf{A}^T\,\widetilde{\mathbf{\Gamma}}\,\mathbf{A}\mathbf{A}^{-\mathbf{1}}\,\delta\mathbf{Z}\\
&= \widetilde{\mathbf{\Delta}}^T\,\mathbf{A}\mathbf{A}^{-1}\,\delta\mathbf{Z}+\frac{1}{2}\delta\mathbf{Z}^T\,(\mathbf{A}^{-1})^T\mathbf{A}^T\,\widetilde{\mathbf{\Gamma}}\,\mathbf{A}\mathbf{A}^{-\mathbf{1}}\,\delta\mathbf{Z}\\
&= \widetilde{\mathbf{\Delta}}^T\mathbf{A}\,(\mathbf{A}^{-1}\,\delta\mathbf{Z})+\frac{1}{2}(\mathbf{A}^{-1}\delta\mathbf{Z})^T\mathbf{A}^T\widetilde{\mathbf{\Gamma}}\mathbf{A}\;(\mathbf{A}^{-1}\delta\mathbf{Z})
\end{aligned}$$

In the penultimate step, the property 21.33 has been used; recall that this holds for every invertible matrix. In the last step, the parentheses are intended to emphasize the fact that $\delta\mathbf{Z}$ only appears in combination with $\mathbf{A}^{-1}$. We have shown before that the components of the vector $\mathbf{A}^{-1}\delta\mathbf{Z}$ are *iid* random variables, see Equation 21.34. We

[6]This goes back to a paper by Rouvinez, see [142].

can therefore write

$$\begin{aligned} \delta V(\mathbf{S}(t)) &= \widetilde{\boldsymbol{\Delta}}^T \mathbf{A}\, \delta\mathbf{Y} + \frac{1}{2}\delta\mathbf{Y}^T \mathbf{M}\, \delta\mathbf{Y} \\ \text{where} \quad \delta\mathbf{Y} &:= \mathbf{A}^{-1}\delta\mathbf{Z} \sim \mathrm{N}(\mathbf{0},\mathbf{1})\ , \quad iid \\ \text{and} \quad \mathbf{M} &:= \mathbf{A}^T \widetilde{\boldsymbol{\Gamma}} \mathbf{A} \end{aligned} \tag{22.14}$$

Thus, the first goal has been accomplished. The δZ_i have been transformed into *iid* random variables δY_i.

Because $\widetilde{\boldsymbol{\Gamma}}$ is by definition a symmetric matrix, i.e. $\widetilde{\Gamma}_{ij} = \widetilde{\Gamma}_{ji}$, we can show that the newly defined matrix $\mathbf{M}$ is symmetric as well:

$$\begin{aligned} M_{ij} &= \left(\mathbf{A}^T \widetilde{\boldsymbol{\Gamma}} \mathbf{A}\right)_{ij} = \sum_k \sum_m (A^T)_{ik} \widetilde{\Gamma}_{km} A_{mj} \\ &= \sum_k \sum_m (A^T)_{ik} \widetilde{\Gamma}_{mk} A_{mj} = \sum_k \sum_m A_{mj} \widetilde{\Gamma}_{mk} (A^T)_{ik} \\ &= \sum_k \sum_m (A^T)_{jm} \widetilde{\Gamma}_{mk} A_{ki} = \left(\mathbf{A}^T \widetilde{\boldsymbol{\Gamma}} \mathbf{A}\right)_{ji} = M_{ji} \end{aligned}$$

22.3.2 Diagonalization of the Gamma Matrix

The next step is to decouple the contributions to δV of the individual random variables in Equation 22.14. This is accomplished by diagonalizing the gamma matrix, or more precisely, the transformed gamma matrix $\mathbf{M}$ introduced in Equation 22.14. Diagonalizing a matrix is a standard procedure in *linear algebra*. We refer the reader to the relevant literature[7]. We nevertheless take the opportunity to demonstrate the fundamental operations for *diagonalizing a matrix* here since they contain essential elements of the practical value at risk computations to be performed in the delta-gamma method.

The *eigenvectors* $\mathbf{e}^i$ of a matrix $\mathbf{M}$ are the vectors which are mapped by $\mathbf{M}$ to the same vector multiplied by a number (called a *scalar* in algebra):

$$\begin{aligned} \mathbf{M}\mathbf{e}^i &= \lambda_i \mathbf{e}^i \Leftrightarrow \\ (\mathbf{M} - \lambda_i \mathbf{1})\, \mathbf{e}^i &= \mathbf{0} \end{aligned} \tag{22.15}$$

These scalars λ_i are called *eigenvalues* of the matrix. As known from linear algebra, an equation of this kind has a non-trivial solution ($\mathbf{e}^i \neq \mathbf{0}$) if and only if the matrix $(\mathbf{M} - \lambda_i \mathbf{1})$ is *singular*. For this to be the case, the *determinant* of this matrix must be zero:

$$\det\left(\mathbf{M} - \lambda_i \mathbf{1}\right) = 0 \tag{22.16}$$

The solutions of Equation 22.16 are the eigen*values* λ_i. Having determined these values, they can be substituted into Equation 22.15 to determine the eigen*vectors* $\mathbf{e}^i$ of the matrix. The eigenvectors have as yet only been defined up to a multiplicative scalar

[7] The most important results required for the analysis here receive a clear and concise treatment in [65], for example.

since if $\mathbf{e}^i$ solves Equation 22.15 then $c\mathbf{e}^i$ does as well, for any arbitrary scalar c. The uniqueness of the eigenvalues can be guaranteed by demanding that the eigenvectors have norm 1:

$$(\mathbf{e}^i)^T\mathbf{e}^i = 1 \tag{22.17}$$

As is known from linear algebra, a symmetric $n \times n$ matrix has n linearly independent eigenvectors which are *orthogonal*. This means that the *inner product* of each pair of different eigenvectors equals zero (graphically: the angle formed by the two vectors is 90 degrees). Together with the normalization the eigenvectors they thus have the following property

$$(\mathbf{e}^i)^T\mathbf{e}^j = \sum_k e^i_k e^j_k = \delta_{ij} \tag{22.18}$$

where δ_{ij} denotes the well-known *Kronecker delta*. A collection of vectors satisfying this property is called *orthonormal*. Since we have shown that the $n \times n$ matrix $\mathbf{M}$ in Equation 22.14 is symmetric, we can be sure that it indeed has n orthonormal eigenvectors satisfying Equation 22.18.

To clarify the notation for these eigenvectors: The subscript k and the superscript i of e^i_k identify this value as the k^{th} component of the i^{th} eigenvector $\mathbf{e}^i$.

$$\mathbf{e}^j = \begin{pmatrix} e^j_1 \\ e^j_2 \\ \vdots \\ e^j_n \end{pmatrix}, \quad (\mathbf{e}^j)^T = \begin{pmatrix} e^j_1 & e^j_2 & \cdots & e^j_n \end{pmatrix}, \quad j = 1, \ldots, n$$

A matrix $\mathbf{O}$ can now be constructed whose column vectors are composed of the eigenvectors of $\mathbf{M}$:

$$\mathbf{O} = \begin{pmatrix} \mathbf{e}^1 & \mathbf{e}^2 & \ldots & \mathbf{e}^n \end{pmatrix} = \begin{pmatrix} e^1_1 & e^2_1 & \cdots & e^n_1 \\ e^1_2 & e^2_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ e^1_n & \cdots & \cdots & e^n_n \end{pmatrix} \Rightarrow O_{ij} = e^j_i \tag{22.19}$$

The j^{th} eigenvector $\mathbf{e}^j$ is in the j^{th} column of the matrix. In the i^{th} row, we find the i^{th} components of all eigenvectors. As we will soon see, this matrix is an indispensable tool for the purpose of the diagonalization. As can be immediately verified[8]

$$\mathbf{O}^T\mathbf{O} = \mathbf{1} \tag{22.20}$$

from which it follows that

$$\mathbf{O}^T = \mathbf{O}^{-1} \Rightarrow \mathbf{O}\mathbf{O}^T = \mathbf{1} \tag{22.21}$$

Equation 22.20 characterizes a group of matrices known as *orthonormal transformations.* Applying such a matrix to a vector effects a rotation of the vector.

[8] $(\mathbf{O}^T\mathbf{O})_{ij} = \sum_k (O^T)_{ik} O_{kj} = \sum_k e^i_k e^j_k = \delta_{ij}$

The eigen*values* of the matrix $\mathbf{M}$ can be used to construct a matrix as well, namely the diagonal matrix

$$\lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \tag{22.22}$$

From Equation 22.15, it follows immediately that the relation

$$\mathbf{MO} = \mathbf{O}\lambda \tag{22.23}$$

holds for the matrices $\mathbf{M}$, $\mathbf{O}$ and λ. Such matrix equations can often be verified quite easily by comparing the matrix elements individually:

$$\begin{aligned} (\mathbf{MO})_{ij} &= \sum_k M_{ik} O_{kj} = \sum_k M_{ik} e_k^j = \left(\mathbf{M}\mathbf{e}^j\right)_i \\ &= \left(\lambda_j \mathbf{e}^j\right)_i = \lambda_j e_i^j = O_{ij}\lambda_j = (\mathbf{O}\lambda)_{ij} \end{aligned}$$

The decisive step in the above proof is the first equality in the second line where the eigenvector equation 22.15 was used. Multiplying both sides of this equation *from the left* by the matrix $\mathbf{O}^T$ and using Equation 22.20 directly yields the desired diagonalization of $\mathbf{M}$, since λ is a diagonal matrix:

$$\mathbf{O}^T\mathbf{MO} = \mathbf{O}^T\mathbf{O}\lambda = \lambda \tag{22.24}$$

Multiplying both sides of Equation 22.23 *from the right* by the matrix $\mathbf{O}^T$ and using Equation 22.21), also yields a very useful representation of $\mathbf{M}$, namely the *spectral representation* (also referred to as *eigenvector decomposition*)

$$\mathbf{M} = \mathbf{O}\lambda\mathbf{O^T} = \sum_k \lambda_k \left(\mathbf{e}^k (\mathbf{e}^k)^T\right)$$

We are now in a position to introduce the diagonalized matrix $\mathbf{O}^T\mathbf{MO}$ into Equation 22.14 by inserting identity matrices in Equation 22.14 and subsequently replacing them by $\mathbf{OO^T}$. Equation 22.21 ensures that equality is maintained**.**

$$\begin{aligned} \delta V(\mathbf{S}(t)) &= \widetilde{\mathbf{\Delta}}^T \mathbf{A}\,\mathbf{1}\,\delta\mathbf{Y} + \frac{1}{2}\delta\mathbf{Y}^T\,\mathbf{1}\,\mathbf{M}\,\mathbf{1}\,\delta\mathbf{Y} \\ &= \widetilde{\mathbf{\Delta}}^T \mathbf{A}\mathbf{O}\mathbf{O^T}\delta\mathbf{Y} + \frac{1}{2}\delta\mathbf{Y}^T\mathbf{O}\underbrace{\mathbf{O}^T\mathbf{M}\mathbf{O}}_{\lambda}\mathbf{O^T}\delta\mathbf{Y} \\ &= \widetilde{\mathbf{\Delta}}^T \mathbf{A}\mathbf{O}\left(\mathbf{O^T}\delta\mathbf{Y}\right) + \frac{1}{2}\left(\mathbf{O}^T\delta\mathbf{Y}\right)^T \lambda \left(\mathbf{O^T}\,\delta\mathbf{Y}\right) \end{aligned}$$

In the last equality, the parentheses are meant to emphasize that $\delta\mathbf{Y}$ appears only in

combination with $\mathbf{O^T}$. In consequence, we can write

$$\begin{aligned} \delta V(\mathbf{S}(t)) &= \widetilde{\mathbf{\Delta}}^T \mathbf{A}\mathbf{O}\, \delta\mathbf{X} + \frac{1}{2}\delta\mathbf{X}^T \lambda \delta\mathbf{X} \\ \text{with} \quad \delta\mathbf{X} &:= \mathbf{O^T}\delta\mathbf{Y} = \mathbf{O^T}\mathbf{A}^{-1}\delta\mathbf{Z} \\ \text{and} \quad \lambda &:= \mathbf{O}^T\mathbf{M}\mathbf{O} = \mathbf{O}^T\mathbf{A}^T\widetilde{\mathbf{\Gamma}}\mathbf{A}\mathbf{O} \end{aligned} \tag{22.25}$$

The δY_i were *iid*, standard normally distributed random variables. This was accomplished in the previous section by the mapping $\mathbf{A}^{-1}$. Now to achieve the diagonalization of the gamma matrix, the random variables must undergo a further transformation under the mapping $\mathbf{O^T}$. The question remains as to whether the accomplishments of the previous section was undone by this new transformation, in other words, whether the transformed random variables have remained independent. We therefore consider the covariance of the new random variables

$$\begin{aligned} \mathrm{cov}[\delta X_i, \delta X_j] &= \mathrm{cov}[\sum_k O^T_{ik}\delta Y_k, \sum_m O^T_{jm} Y_m] \\ &= \sum_k \sum_m O^T_{ik} O^T_{jm} \underbrace{\mathrm{cov}[\delta Y_k, \delta Y_m]}_{\delta_{km}} \\ &= \sum_k O^T_{ik} O^T_{jk} \\ &= \sum_k O^T_{ik} O_{kj} = \left(\mathbf{O}^T\mathbf{O}\right)_{ij} = \mathbf{1}_{ij} = \delta_{ij} \end{aligned}$$

The covariances have remained invariant under the transformation $\mathbf{O}^T$. Thus, the new random variables are also uncorrelated and all have variance 1. Also, the zero expectation does not change under the transformation $\mathbf{O}^T$:

$$\mathrm{E}[\delta X_i] = \mathrm{E}[\sum_k O^T_{ik}\delta Y_k] = \sum_k O^T_{ik} \underbrace{\mathrm{E}[\delta Y_k]}_{0}$$

Since matrix multiplication is a linear transformation (we are operating in the realm of *linear* algebra), the form of the distribution remains the same as well. In summary, the new random variables are uncorrelated and all have the same standard-normal distribution. We have argued after Equation 21.34 that such variables are indeed *iid* random variables. Summarizing the above deliberations, we can write for the δX_i:

$$\delta\mathbf{X} \sim \mathrm{N}(\mathbf{0}, \mathbf{1}) \ , \quad iid$$

We have only used property 22.20, i.e. the *iid* property of random variables remains invariant under *every* orthonormal transformation.

If we define a "transformed sensitivity vector" as

$$\mathbf{L} := \mathbf{O}^T\mathbf{A}^T\widetilde{\mathbf{\Delta}}$$

(which implies for its transposed $\mathbf{L}^T := \widetilde{\mathbf{\Delta}}^T\mathbf{A}\mathbf{O}$), the portfolio-change Equation 22.25 can be brought into the following simple form

$$\delta V(\mathbf{S}(t)) = \mathbf{L}^T\, \delta\mathbf{X} + \frac{1}{2}\delta\mathbf{X}^T\lambda\delta\mathbf{X} \quad \text{with} \quad \delta\mathbf{X} \sim \mathrm{N}(\mathbf{0}, \mathbf{1}) \tag{22.26}$$

or, expressed component-wise

$$\delta V(\mathbf{S}(t)) = \sum_i \left[L_i\, \delta X_i + \frac{1}{2}\lambda_i \delta X_i^2 \right] = \sum_i \delta V_i$$

$$\text{where} \quad \delta V_i = L_i\, \delta X_i + \frac{1}{2}\lambda_i \delta X_i^2 \ , \quad i = 1, \ldots, n \tag{22.27}$$

The change in the portfolio's value is now the decoupled sum of the individual contributions of *iid* random variables as was our original intention.

At this stage, we collect all transformations involved in mapping the original random variables in Equation 22.13 into the *iid* random variables in the above expression Equation 22.26 :

$$\delta \mathbf{X} := \mathbf{O}^T \mathbf{A}^{-1} \delta \mathbf{Z} \ , \quad \lambda := \mathbf{O}^T \mathbf{A}^T \widetilde{\mathbf{\Gamma}}\, \mathbf{A}\mathbf{O} \ , \quad \mathbf{L} := \mathbf{O}^T \mathbf{A}^T \widetilde{\mathbf{\Delta}}$$

From Equation 22.21 we know that $\mathbf{O}^T\mathbf{A}^{-1} = \mathbf{O}^{-1}\mathbf{A}^{-1} = (\mathbf{A}\mathbf{O})^{-1}$. We now recognize that all of these transformations can be represented with a single matrix defined as

$$\mathbf{D} := \mathbf{A}\mathbf{O} \tag{22.28}$$

With this matrix, the transformations become simply

$$\delta \mathbf{X} := \mathbf{D}^{-1} \delta \mathbf{Z} \ , \quad \lambda := \mathbf{D}^T \widetilde{\mathbf{\Gamma}}\, \mathbf{D} \ , \quad \mathbf{L} := \mathbf{D}^T \widetilde{\mathbf{\Delta}} \tag{22.29}$$

The matrix $\mathbf{D}$ *directly* diagonalizes the gamma matrix $\widetilde{\mathbf{\Gamma}}$ (as opposed to $\mathbf{O}$, which diagonalizes the matrix $\mathbf{M}$). *In addition,* $\mathbf{D}$ is by definition, an orthonormal transformation (a "rotation") of $\mathbf{A}$, the Cholesky decomposition of the covariance matrix. $\mathbf{D}$ is likewise a "square root" of the covariance matrix, since the square of a matrix remains invariant under orthonormal transformations of the matrix. Explicitly:

$$\mathbf{D}\mathbf{D}^T = \mathbf{A}\mathbf{O}\,(\mathbf{A}\mathbf{O})^T = \mathbf{A}\mathbf{O}\mathbf{O}^T\mathbf{A}^T = \mathbf{A}\mathbf{1}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \delta\mathbf{\Sigma} \tag{22.30}$$

Therefore, the matrix $\mathbf{D}$ satisfies *both* tasks, namely the decoupling of the gamma matrix *and* the transformation of the correlated random variables into uncorrelated ones.

As a little consistence check, using the matrix $\mathbf{D}$ we immediately recognize the equivalence of Equations 22.26 and 22.13:

$$\begin{aligned}
\delta V(\mathbf{S}(t)) &= \left(\mathbf{D}^T \widetilde{\mathbf{\Delta}}\right)^T \mathbf{D}^{-1}\delta\mathbf{Z} + \frac{1}{2}\left(\mathbf{D}^{-1}\delta\mathbf{Z}\right)^T \mathbf{D}^T \widetilde{\mathbf{\Gamma}}\, \mathbf{D}\mathbf{D}^{-1}\delta\mathbf{Z} \\
&= \widetilde{\mathbf{\Delta}}^T \underbrace{\mathbf{D}\,\mathbf{D}^{-1}}_{\mathbf{1}}\delta\mathbf{Z} + \frac{1}{2}\delta\mathbf{Z}^T \underbrace{\left(\mathbf{D}^T\right)^{-1}\mathbf{D}^T}_{\mathbf{1}} \widetilde{\mathbf{\Gamma}}\, \underbrace{\mathbf{D}\mathbf{D}^{-1}}_{\mathbf{1}}\delta\mathbf{Z}
\end{aligned}$$

where Equation 21.33 was again used in verifyng this equivalence.

22.3.3 The Distribution of the Portfolio Value Changes

Having decoupled the individual contributions to δV in Equation 22.27 into standard normally distributed *iid* random variables, we can now determine the distribution of the sum. δV is nevertheless not simply the sum of normally distributed random variables alone, since the expression also includes the *square* of normally distributed random variables. These additional random variables represent the difference in complexity compared to the delta-normal method. According to Section A.4.6, the square of a standard normally distributed random variable is χ^2-distributed with one degree of freedom. Wie can thus write Equation 22.27 as

$$\delta V(\mathbf{S}(t)) = \sum_{i=1}^{n} L_i\,\delta X_i + \frac{1}{2}\sum_{i=1}^{n} \lambda_i \widetilde{X}_i \quad \text{mit } \delta X_i \sim \mathrm{N}(0,1)\ ,\ \ \widetilde{X}_i \sim \chi^2(1)$$

However, the $\widetilde{X}_i$ are not independent random variables since we obviously have

$$\widetilde{X}_i = (\delta X_i)^2 \quad \forall i$$

We need to re-write δV in such a way that every term appearing is statistically independent of every other term. First note that a δX_i is independent of every other term in δV if and only if the corresponding eigenvalue λ_i is zero, since in this case the corresponding $\widetilde{X}_i$ does not appear in the sum. We will emphasize this by introducing the index set J which contains only the indices of non-zero eigenvalues:

$$J = \{1, \ldots, n \,|\lambda_j \neq 0\} \tag{22.31}$$

With this index set we can write[9]

$$\begin{aligned}\delta V(\mathbf{S}(t)) &= \sum_{i\notin J} L_i\,\delta X_i + \sum_{j\in J} L_j\,\delta X_j + \frac{1}{2}\sum_{j\in J}\lambda_j \delta X_j^2 \\ &= \sum_{i\notin J} L_i\,\delta X_i + \sum_{j\in J}\left[L_j\,\delta X_j + \frac{1}{2}\lambda_j \delta X_j^2\right]\end{aligned} \tag{22.32}$$

The first sum in Equation 22.32 is actually a sum of normally distributed random variables and as such is again a normally distributed random variable which we denote by u_0. The expectation of this random variable can be calculated as

$$\mathrm{E}\,[u_0] = \mathrm{E}\left[\sum_{i\notin J} L_i\,\delta X_i\right] = \sum_{i=1}^{n} L_i \underbrace{\mathrm{E}\,[\,\delta X_i]}_{0} = 0$$

and the variance is

$$\begin{aligned}\mathrm{var}\,[u_0] &= \mathrm{var}\left[\sum_{i\notin J} L_i\,\delta X_i\right] = \sum_{i,j\notin J} \mathrm{cov}\,[L_i\,\delta X_i, L_j\,\delta X_j] \\ &= \sum_{i,j\notin J} L_i L_j \underbrace{\mathrm{cov}\,[\,\delta X_i,\,\delta X_j]}_{\delta_{ij}} = \sum_{i\notin J} L_i^2\end{aligned}$$

[9]The notation $i \notin J$ denotes all indices i with eigenvalue $\lambda_j = 0$, i.e., the set $\{1, \ldots, n \,|\lambda_j = 0\}$.

with the components L_i of the transformed sensitivity vector $\mathbf{L}$ defined in Equation 22.29. Thus

$$u_0 := \sum_{i \notin J} L_i \, \delta X \sim \mathrm{N}(0, \sum_{i \notin J} L_i^2)$$

Consider now the sums over $j \in J$ in Equation 22.32. To combine the *dependent* random numbers δX_j and δX_j^2 in the square brackets of Equation 22.32 into one single random number, we complete the square for each $j \in J$:

$$\frac{1}{2}\lambda_j \delta X_j^2 + L_j \, \delta X_j = \frac{1}{2}\lambda_j \left(\delta X_j^2 + 2\frac{L_j}{\lambda_j} \, \delta X_j \right) = \frac{1}{2}\lambda_j \left(\delta X_j + \frac{L_j}{\lambda_j} \right)^2 - \frac{L_j^2}{2\lambda_j}$$

Since δX_j is a standard normal random variable wie have

$$\delta X_j \sim \mathrm{N}(0,1) \quad \Longrightarrow \quad \delta X_j + \frac{L_j}{\lambda_j} \sim \mathrm{N}\left(\frac{L_j}{\lambda_j}, 1 \right)$$

Therefore, according to Equation A.88 in Section A.4.6 $u_j := (\delta X_j + L_j/\lambda_j)^2$ has a *non-central* χ^2-distribution with one degree of freedom and non-central parameter L_j^2/λ_j^2:

$$\left(\delta X_j + \frac{L_j}{\lambda_j} \right)^2 =: u_j \sim \chi^2 \left(1, \, \frac{L_j^2}{\lambda_j^2} \right) \quad \forall j \in J$$

In summary, δV has now become a sum of non-central χ^2-distributed random variables u_j plus a normally distributed random variable u_0 (plus a constant), where all the random variables appearing are independent of each other:

$$\delta V(\mathbf{S}(t)) = u_0 + \frac{1}{2}\sum_{j \in J} \lambda_j u_j - \underbrace{\frac{1}{2}\sum_{j \in J} L_j^2/\lambda_j}_{\text{constant}} \tag{22.33}$$

$$\text{with} \;\; u_0 \sim \mathrm{N}(0, \sum_{i \notin J} L_i^2) \;\; , \;\; u_j \sim \chi^2 \left(1, (L_j/\lambda_j)^2 \right) \;\; , \;\; j \in J$$

The problem now consists in determining the distribution of the sum of independent but *differently* distributed random variables (or at least its percentiles). According to Equation 21.4, the value at risk at a specified confidence level is then computed with precisely these percentiles or, equivalently, by inverting the cumulative distribution function of δV.

22.3.4 Moments of the Portfolio Value Distribution

We begin by calculating the moments of the random variable δV. The first moments, the expectation, the variance, etc. (see Equations A.22 and A.23) have intuitive interpretations and their explicit forms provide an intuitive conception of the distribution. In addition, approximations for the distribution (the Johnson approximation) and for the percentiles (the Cornish-Fisher approximation) will later be presented which can be computed with the moments.

The *Moment Generating Function* is a very useful tool for calculating the moments. The moment generating function (abbreviated as *MGF*) G_x of a random variable x with density function $\mathrm{pdf}(x)$ is defined in Section A.3.1 by

$$G_x(s) \equiv \mathrm{E}[e^{sx}] = \int_{-\infty}^{\infty} e^{sx}\,\mathrm{pdf}(x)dx = \sum_{n=0}^{\infty} \frac{s^n}{n!}\,\mathrm{E}[x^n] \tag{22.34}$$

where in the last step the exponential function e^{sx} has been expanded in its Taylor series. The name *moment generating function* stems from the fact, stated mathematically in Equation A.27, that the derivatives of the function $G_x(s)$ with respect to s evaluated at $s = 0$ generate the moments of the random variable x

$$\mathrm{E}[x^n] = \left.\frac{\partial^n G_x(s)}{\partial s^n}\right|_{s=0} \tag{22.35}$$

As demonstrated in Section A.3.1, the MGF can be explicitly computed for many distributions from their integral representation, Equation A.25. In particular, according to Equation A.56, the MGF of u_0, i.e. the MGF of a $\mathrm{N}(0, \sum_{i\notin J} L_i^2)$ distributed random variable is given by

$$G_{u_0}(s) = \exp\left(\frac{1}{2}\,s^2 \sum_{i\notin J} L_i^2\right) \tag{22.36}$$

while Equation A.89 gives the MGF of u_j, i.e. of a non-central χ^2-distribution with one degree of freedom $\chi^2(1, (L_j/\lambda_j)^2)$

$$G_{u_j}(s) = \frac{1}{\sqrt{1-2s}} \exp\left\{\frac{s}{1-2s}\frac{L_j^2}{\lambda_j^2}\right\}\ ,\ \ j \in J \tag{22.37}$$

This function is well-defined for $s < 1/2$, which is sufficient for our needs since as is clear from Equation 22.35 that we are particularly interested in values of s in a neighborhood of zero.

The usefulness of the MGF in the calculation of the distribution of δV in Equation 22.33 stems from the fact that according to Equation A.30 the MGF of the sum of *independent* random variables x, y is simply the product of the MGFs of each of the random variables:

$$G_{x+y}(s) = G_x(s)\,G_y(s) \tag{22.38}$$

and that furthermore, from Equation A.31

$$G_{ax+b}(s) = e^{bs} G_x(as) \tag{22.39}$$

for all non-stochastic values a, b and random variables x. The MGF of δV can thus be written as the product of the each of the MGFs appearing in the sum:

$$G_{\delta V}(s) = \exp\left\{-s\sum_{j\in J}\frac{L_j^2}{2\lambda_j}\right\} G_{u_0}(s) \prod_{j\in J} G_{u_j}(\frac{1}{2}\lambda_j s)$$

The MGFs of each of the individual random variables are given explicitly in Equations 22.36 and 22.37. Performing the substitution $s \to \lambda_j s/2$ in Equation 22.37 yields the required MGF for the argument $\lambda_j s/2$:

$$G_{u_j}(\frac{1}{2}\lambda_j s) = \frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{L_j^2}{2\lambda_j}\frac{s}{1-\lambda_j s}\right\}$$

We thus obtain an *explicit* expression for the moment generating function of the distribution of the portfolio's value changes:

$$\begin{aligned} G_{\delta V}(s) &= \exp\left\{-s\sum_{j\in J}\frac{L_j^2}{2\lambda_j}\right\} \exp\left(\frac{1}{2}\,s^2\sum_{i\notin J}L_i^2\right) \prod_{j\in J}\frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{L_j^2}{2\lambda_j}\frac{s}{1-\lambda_j s}\right\} \\ &= \exp\left(\frac{1}{2}\,s^2\sum_{i\notin J}L_i^2\right) \prod_{j\in J}\left[\frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{L_j^2}{2\lambda_j}\frac{s}{1-\lambda_j s}\right\} \exp\left\{-s\frac{L_j^2}{2\lambda_j}\right\}\right] \\ &= \exp\left(\frac{1}{2}\,s^2\sum_{i\notin J}L_i^2\right) \prod_{j\in J}\left[\frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{L_j^2}{2\lambda_j}s\left(\frac{1}{1-\lambda_j s}-1\right)\right\}\right] \end{aligned}$$

Using the same denominator in the second exp-function finally yields

$$G_{\delta V}(s) = \exp\left(\frac{1}{2}\,s^2\sum_{i\notin J}L_i^2\right) \prod_{j\in J}\frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{1}{2}L_j^2\frac{s^2}{1-\lambda_j s}\right\} \tag{22.40}$$

This can be simplified even further by the following trick: Since $\lambda_i = 0$ for all $i \notin J$, we can re-write the exp-function in the following way:

$$\exp\left(\frac{1}{2}\,s^2\sum_{i\notin J}L_i^2\right) = \prod_{i\notin J}\exp\left(\frac{1}{2}\,s^2L_i^2\right) = \prod_{i\notin J}\frac{1}{\sqrt{1-\lambda_i s}} \exp\left\{\frac{1}{2}L_i^2\frac{s^2}{1-\lambda_i s}\right\}$$

Using this form in Equation 22.40 allows us to write δV very compactly as a product over *all* indexes $j = 1, \ldots, n$

$$G_{\delta V}(s) = \prod_{j=1}^{n}\frac{1}{\sqrt{1-\lambda_j s}} \exp\left\{\frac{1}{2}L_j^2\frac{s^2}{1-\lambda_j s}\right\} \tag{22.41}$$

This function is well-defined for all $s < \min_{j\in J}\left(\frac{1}{2|\lambda_i|}\right)$, which is sufficient for our needs since because of Equation 22.35, we are only interested in values of s which are close to zero.

Now, using Equation 22.35, arbitrary moments of δV can be computed. We start the calculation of the first moment by introducing the abbreviation

$$a_j := \frac{1}{2}L_j^2\frac{s^2}{1-\lambda_j s}$$

Application of the well-known *product rule* yields

$$\begin{aligned}
\mathrm{E}[\delta V] &= \left.\frac{\partial G_{\delta V}(s)}{\partial s}\right|_{s=0} \\
&= \left.\frac{\partial}{\partial s}\prod_{j=1}^{n}\frac{e^{a_j}}{\sqrt{1-\lambda_j s}}\right|_{s=0} \\
&= \left.\sum_{j=1}^{n}\left(\frac{\partial}{\partial s}\frac{e^{a_j}}{\sqrt{1-\lambda_j s}}\right)\prod_{k=1,k\neq j}^{n}\frac{e^{a_k}}{\sqrt{1-\lambda_k s}}\right|_{s=0}
\end{aligned}$$

The derivative we need to calculate is

$$\begin{aligned}
\frac{\partial}{\partial s}\frac{e^{a_j}}{\sqrt{1-\lambda_j s}} &= e^{a_j}\frac{\partial}{\partial s}\frac{1}{\sqrt{1-\lambda_j s}} + \frac{1}{\sqrt{1-\lambda_j s}}\frac{\partial}{\partial s}e^{a_j} \\
&= \frac{\frac{1}{2}\lambda_j e^{a_j}}{(1-\lambda_j s)^{3/2}} + \frac{\frac{1}{2}L_j^2 e^{a_j}}{\sqrt{1-\lambda_j s}}\left(\frac{2s}{1-\lambda_j s} + \frac{\lambda_j\, s^2}{(1-\lambda_j s)^2}\right) \\
&= \frac{1}{2}\frac{e^{a_j}}{(1-\lambda_j s)^{3/2}}\left(\lambda_j + 2L_j^2 s + \lambda_j L_j^2\frac{s^2}{1-\lambda_j s}\right)
\end{aligned}$$

For $s = 0$ almost all terms vanish and we are left with $\lambda_j/2$. Thus $\mathrm{E}[\delta V]$ becomes simply

$$\mathrm{E}[\delta V] = \sum_{j=1}^{n}\frac{1}{2}\lambda_j \left.\prod_{k\in J,k\neq j}\frac{e^{a_k}}{\sqrt{1-\lambda_k s}}\right|_{s=0} = \frac{1}{2}\sum_{j=1}^{n}\lambda_j$$

The expectation of the portfolio's value changes in the delta-gamma approximation Equation 22.2 is thus just half the sum of the eigenvalues of the transformed gamma matrix $\mathbf{M}$. This is by definition half the *trace* of the eigenvalue matrix λ. With Equations 22.29 and 22.30 we arrive at the conclusion that the expectation of δV equals half the trace of the product of the gamma matrix and the covariance matrix[10]:

$$\mathrm{E}[\delta V] = \frac{1}{2}\,\mathrm{tr}\,(\lambda) = \frac{1}{2}\,\mathrm{tr}\left(\mathbf{D}^T\,\widetilde{\mathbf{\Gamma}}\,\mathbf{D}\right) = \frac{1}{2}\,\mathrm{tr}\left(\widetilde{\mathbf{\Gamma}}\,\mathbf{D}\mathbf{D}^T\right) = \frac{1}{2}\,\mathrm{tr}\left(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\right) \tag{22.42}$$

Note that the drifts of all risk factors have been neglected (see Equations 22.13 and 22.1). The *risk factors* are thus all approximated to be *drift-free*. But then, for a portfolio depending only *linearly* on the risk factors (or in the linear approximation of the delta-normal method) the expectation (the drift) of the portfolio value changes also equals zero. In Equation 22.42, the expectation (the drift) of the portfolio changes is *not* zero because *non-linear* effects were taken into consideration. It is readily seen that the gamma matrix gives rise to the drift of δV in contrast to the linear sensitivities $\widetilde{\mathbf{\Delta}}$ which do *not* appear in 22.42.

[10]Here the well-known cyclic property of the trace has been used: $\mathrm{tr}\,(\mathbf{ABC}) = \mathrm{tr}\,(\mathbf{BCA})$ for arbitrary matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

To find out more about the distribution of δV, we proceed by computing its variance. According to Equation A.5 the variance is the second central moment which can be calculated via Equation A.29:

$$\begin{aligned}
\operatorname{var}[\delta V] &= \mathrm{E}[(\delta V - \mathrm{E}[\delta V])^2] \\
&= \frac{\partial^2}{\partial s^2} \exp\left(-s\mathrm{E}[\delta V]\right) G_{\delta V}(s)\Big|_{s=0} \\
&= \frac{\partial^2}{\partial s^2} \exp\left(-s\frac{1}{2}\sum_{i=1}^{n}\lambda_i\right) \prod_{j=1}^{n} \frac{\exp\left(\frac{1}{2}L_j^2 \frac{s^2}{1-\lambda_j s}\right)}{\sqrt{1-\lambda_j s}}\Bigg|_{s=0} \\
&= \frac{\partial^2}{\partial s^2} \prod_{j=1}^{n} \frac{1}{\sqrt{1-\lambda_j s}} \exp\left(\frac{1}{2}L_j^2 \frac{s^2}{1-\lambda_j s} - \frac{1}{2}\lambda_j s\right)\Bigg|_{s=0} \\
&= \frac{\partial^2}{\partial s^2} \prod_{j=1}^{n} a_j \Bigg|_{s=0}
\end{aligned} \tag{22.43}$$

with the abbreviation

$$a_j := \frac{1}{\sqrt{1-\lambda_j s}} \exp\left(\frac{1}{2}L_j^2 \frac{s^2}{1-\lambda_j s} - \frac{1}{2}\lambda_j s\right)$$

The second derivative of this product is quite involved. We nonetheless present it explicitly here to demonstrate how moments of δV can be determined in practice. Such moments are needed quite often, for instance for the Cornish-Fisher expansion. We start by repeatedly applying the product rule to arrive at

$$\begin{aligned}
\frac{\partial^2}{\partial s^2} \prod_{j=1}^{n} a_j &= \frac{\partial}{\partial s}\left(\sum_{j=1}^{n} \frac{\partial a_j}{\partial s} \prod_{k=1, k\neq j}^{n} a_k\right) \\
&= \sum_{j=1}^{n}\left(\frac{\partial}{\partial s}\frac{\partial a_j}{\partial s}\right) \prod_{k=1,k\neq j}^{n} a_k + \sum_{j=1}^{n} \frac{\partial a_j}{\partial s}\frac{\partial}{\partial s} \prod_{k=1,k\neq j}^{n} a_k \\
&= \sum_{j=1}^{n} \frac{\partial^2 a_j}{\partial s^2} \prod_{k=1,k\neq j}^{n} a_k + \sum_{j=1}^{n} \frac{\partial a_j}{\partial s} \sum_{k=1,k\neq j}^{n} \frac{\partial a_k}{\partial s} \prod_{\substack{m=1, \\ m\neq k,\, m\neq j}}^{n} a_m
\end{aligned} \tag{22.44}$$

Thus, we mainly have to differentiate a_j. For ease of notation, we introduce yet another abbreviation, namely

$$b_j := \frac{1}{2}L_j^2 \frac{s^2}{1-\lambda_j s} - \frac{1}{2}\lambda_j s \quad \Longrightarrow \quad a_j = \frac{e^{b_j}}{\sqrt{1-\lambda_j s}}$$

The first derivative with respect to a_j is now calculated as

$$\begin{aligned}\frac{\partial a_j}{\partial s} &= \frac{1}{\sqrt{1-\lambda_j s}}\frac{\partial e^{b_j}}{\partial s} + e^{b_j}\frac{\partial}{\partial s}\frac{1}{\sqrt{1-\lambda_j s}} \\ &= \frac{e^{b_j}}{\sqrt{1-\lambda_j s}}\left(\frac{s\,L_j^2}{1-\lambda_j s} + \frac{\frac{1}{2}s^2 L_j^2\lambda_j}{(1-\lambda_j s)^2} - \frac{\lambda_j}{2}\right) + \frac{\frac{1}{2}e^{b_j}\lambda_j}{(1-\lambda_j s)^{3/2}} \\ &= \frac{sL_j^2 e^{b_j}}{(1-\lambda_j s)^{3/2}} + \frac{\frac{1}{2}s^2 L_j^2 e^{b_j}\lambda_j}{(1-\lambda_j s)^{5/2}} - \frac{\frac{1}{2}e^{b_j}\lambda_j}{(1-\lambda_j s)^{1/2}} + \frac{\frac{1}{2}e^{b_j}\lambda_j}{(1-\lambda_j s)^{3/2}}\end{aligned} \tag{22.45}$$

For $s = 0$ the first two terms in the last line vanish and the last two terms just compensate each other so that $\partial a_j/\partial s$ vanishes completely at $s = 0$:

$$\left.\frac{\partial a_j}{\partial s}\right|_{s=0} = -\frac{1}{2}e^{b_j}\lambda_j + \frac{1}{2}e^{b_j}\lambda_j = 0 \tag{22.46}$$

Therefore only the term involving the second derivative in Equation 22.44 contributes to Equation 22.43. Using the result 22.45, this second derivative is explicitly:

$$\begin{aligned}\frac{\partial^2 a_j}{\partial s^2} &= \frac{\partial}{\partial s}\left[\frac{sL_j^2 e^{b_j}}{(1-\lambda_j s)^{3/2}} + \frac{\frac{1}{2}s^2 L_j^2 e^{b_j}\lambda_j}{(1-\lambda_j s)^{5/2}} - \frac{\frac{1}{2}e^{b_j}\lambda_j}{(1-\lambda_j s)^{1/2}} + \frac{\frac{1}{2}e^{b_j}\lambda_j}{(1-\lambda_j s)^{3/2}}\right] \\ &= L_j^2\frac{\partial e^{b_j}}{\partial s}\frac{s}{(1-\lambda_j s)^{3/2}} + L_j^2 e^{b_j}\frac{3}{2}\frac{\lambda_j s}{(1-\lambda_j s)^{5/2}} + L_j^2 e^{b_j}\frac{1}{(1-\lambda_j s)^{3/2}} \\ &+ \frac{1}{2}L_j^2\frac{\partial e^{b_j}}{\partial s}\frac{\lambda_j s^2}{(1-\lambda_j s)^{5/2}} + \frac{5}{4}L_j^2 e^{b_j}\frac{\lambda_j^2 s^2}{(1-\lambda_j s)^{7/2}} + L_j^2 e^{b_j}\frac{\lambda_j s}{(1-\lambda_j s)^{5/2}} \\ &- \frac{1}{2}\frac{\partial e^{b_j}}{\partial s}\frac{\lambda_j}{(1-\lambda_j s)^{1/2}} - \frac{1}{4}e^{b_j}\frac{\lambda_j^2}{(1-\lambda_j s)^{3/2}} \\ &+ \frac{1}{2}\frac{\partial e^{b_j}}{\partial s}\frac{\lambda_j}{(1-\lambda_j s)^{3/2}} + \frac{3}{4}e^{b_j}\frac{\lambda_j^2}{(1-\lambda_j s)^{3/2}}\end{aligned}$$

Most terms above have s as a factor. They all vanish for $s = 0$. The only terms remaining are:

$$\begin{aligned}\left.\frac{\partial^2 a_j}{\partial s^2}\right|_{s=0} &= L_j^2 e^{b_j} - \frac{1}{2}\frac{\partial e^{b_j}}{\partial s}\lambda_j - \frac{1}{4}e^{b_j}\lambda_j^2 + \frac{1}{2}\frac{\partial e^{b_j}}{\partial s}\lambda_j + \frac{3}{4}e^{b_j}\lambda_j^2 \\ &= \underbrace{e^{b_j}}_{1 \text{ for } s=0}\left(L_j^2 + \frac{1}{2}\lambda_j^2\right)\end{aligned} \tag{22.47}$$

Inserting all these results into Equation 22.43 finally yields

$$\begin{aligned}\mathrm{var}[\delta V] &= \sum_{j=1}^{n} \frac{\partial^2 a_j}{\partial s^2} + \sum_{j=1}^{n} \frac{\partial a_j}{\partial s} \sum_{k=1,k\neq j}^{n} \frac{\partial a_k}{\partial s}\bigg|_{s=0} \\ &= \sum_{j=1}^{n} \frac{\partial^2 a_j}{\partial s^2}\bigg|_{s=0} \\ &= \sum_{j=1}^{n} \left(L_j^2 + \frac{1}{2}\lambda_j^2\right)\end{aligned}$$

In the first step we used Equation 22.44 and $a_k|_{s=0} = 1$. In the second step the result 22.46 was inserted and in the third step the result 22.47. The sum $\sum L_j^2$ is just the square of the transformed sensitivity *vector* and $\sum \lambda_j^2$ is the trace of the square of the the *matrix* of eigenvalues, i.e.

$$\mathrm{var}[\delta V] = \mathbf{L}^T\mathbf{L} + \frac{1}{2}\,\mathrm{tr}\left(\lambda^2\right)$$

Finally, making use of the transformations in Equation 22.29 and applying Equation 22.30, the variance of the portfolio's value change in the framework of the delta-gamma method becomes

$$\begin{aligned}\mathrm{var}[\delta V] &= \widetilde{\mathbf{\Delta}}^T\mathbf{D}\mathbf{D}^T\widetilde{\mathbf{\Delta}} + \frac{1}{2}\,\mathrm{tr}\left(\mathbf{D}^T\,\widetilde{\mathbf{\Gamma}}\,\mathbf{D}\mathbf{D}^T\,\widetilde{\mathbf{\Gamma}}\,\mathbf{D}\right) \\ &= \widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\widetilde{\mathbf{\Delta}} + \frac{1}{2}\,\mathrm{tr}\,\left(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\right)\end{aligned} \tag{22.48}$$

Note that the first term resulting from the linear portfolio sensitivities $\widetilde{\mathbf{\Delta}}$ is identical to the portfolio variance in the delta-normal method (see Equation 22.9). The non-linear sensitivities $\widetilde{\mathbf{\Gamma}}$ effect a correction of the linear portfolio variance which has a form similar to the drift correction from the non-linear term in Equation 22.42. While in Equation 22.42, the trace of the product of the gamma matrix and the covariance matrix was relevant, now the trace of the *square* of this product is required for computing the variance.

The variance is the second *central* moment of the random variable. The *central moments* μ_i of a random variable are defined in general terms in Equation A.23 as the "expectation of powers of the deviation from the expectation":

$$\mu_i := \mathrm{E}[(\delta V - \mathrm{E}[\delta V])^i]\ ,\ \ i > 1$$

Analogously to the approach for the first two moments demonstrated above, we can continue to calculate the further moments of δV. The first (central) moments are

compiled here:

$$\begin{aligned}
\mu &= \mathrm{E}[\delta V] = \frac{1}{2}\,\mathrm{tr}\,\left(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\right)\\
\mu_2 &= \mathrm{E}[(\delta V - \mathrm{E}[\delta V])^2] = \widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\widetilde{\mathbf{\Delta}} + \frac{1}{2}\,\mathrm{tr}\,\left((\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^2\right)\\
\mu_3 &= \mathrm{E}[(\delta V - \mathrm{E}[\delta V])^3] = 3\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\,\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\,\widetilde{\mathbf{\Delta}} + \mathrm{tr}\,\left((\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^3\right)\\
\mu_4 &= \mathrm{E}[(\delta V - \mathrm{E}[\delta V])^4] = 12\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\,(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^2\,\widetilde{\mathbf{\Delta}} + 3\,\mathrm{tr}\,\left((\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^4\right) + 3\mu_2^2
\end{aligned} \tag{22.49}$$

In this way, a great deal of additional information about the distribution of δV can be generated. For instance *skewness* and *curtosis* of the distribution of δV are[11]

$$\begin{aligned}
\text{Skewness} &\equiv \frac{\mu_3}{\mu_2^{3/2}} = \frac{3\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\,\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma}\,\widetilde{\mathbf{\Delta}} + \mathrm{tr}\,(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^3}{\left(\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\widetilde{\mathbf{\Delta}} + \frac{1}{2}\,\mathrm{tr}\,(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^2\right)^{3/2}}\\
\text{Curtosis} &\equiv \frac{\mu_4}{\mu_2^2} = \frac{12\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\,(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^2\,\widetilde{\mathbf{\Delta}} + 3\,\mathrm{tr}\,\left((\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^4\right) + 3\mu_2^2}{\left(\widetilde{\mathbf{\Delta}}^T\delta\mathbf{\Sigma}\widetilde{\mathbf{\Delta}} + \frac{1}{2}\,\mathrm{tr}\,(\widetilde{\mathbf{\Gamma}}\,\delta\mathbf{\Sigma})^2\right)^2}
\end{aligned}$$

A *percentile*, however, is needed for the computation of the value at risk as given in Equation 21.4.

Johnson Transformation

Computation of a percentile necessitates knowledge of the distribution of the random variable *directly* and not of its moments. In order to be able to proceed, we could assume a particular functional form of the distribution and then establish a relation between the parameters of this functional form and the moments of the random variable via *moment matching*. Since the moments, as shown above, can be explicitly computed, the parameters of the assumed distribution can thus be determined. For example, if we assume that a random variable is normally or lognormally distributed, we would take Equations 22.42 and 22.48 as parameter values. Additional functional forms for approximating the distribution of δV were suggested by Johnson [99]. These *Johnson transformations* have four parameters which can be determined from the first four moments in Equation 22.49. They represent a substantially better approximation than, for example a lognormal distribution.

Cornish-Fisher Expansion

One possibility of approximating the *percentiles* of a distribution from its *central moments* and the percentiles $Q^{\mathrm{N}(0,1)}$ of the standard normal distribution is the *Cornish-Fisher expansion*. Since this expansion makes use of the *standard* normal distribution,

[11] Recall that a normal distribution has skewness 0 and kurtosis 3, see Equation A.58.

we must first transform δV into a centered and normalized random variable $\widetilde{\delta V}$ with expectation 0 and variance 1. This is accomplished by defining

$$\widetilde{\delta V} := \frac{\delta V - \mathrm{E}[\delta V]}{\sqrt{\mathrm{var}[\delta V]}} = \frac{\delta V - \mu}{\sqrt{\mu_2}}$$

The percentile of the distribution of $\widetilde{\delta V}$ can now be approximated with the Cornish-Fisher expansion [33], [168] as follows

$$\begin{aligned} Q^{\mathrm{cpf}_{\widetilde{\delta V}}} \approx\; & Q^{\mathrm{N}(0,1)} + \frac{1}{6}\left[(Q^{\mathrm{N}(0,1)})^2 - 1\right]\frac{\mu_3}{\mu_2^{3/2}} \\ & + \frac{1}{24}\left[(Q^{\mathrm{N}(0,1)})^3 - 3Q^{\mathrm{N}(0,1)}\right]\left(\frac{\mu_4}{\mu_2^2} - 3\right) \\ & + \frac{1}{36}\left[2\,(Q^{\mathrm{N}(0,1)})^3 - 5Q^{\mathrm{N}(0,1)}\right]\left(\frac{\mu_3}{\mu_2^{3/2}}\right)^2 \end{aligned} \tag{22.50}$$

where the expansion is taken up to the order, which uses only the first four moments from Equation 22.49. The probability that $\widetilde{\delta V}$ is less than a number a is, naturally, the same as the probability that δV is less than $\mu + \sqrt{\mu_2}\,a$. Thus

$$Q^{\mathrm{cpf}_{\delta V}} = \mu + \sqrt{\mu_2}\,Q^{\mathrm{cpf}_{\widetilde{\delta V}}}$$

holds for the percentiles. From Equation 21.4, the value at risk is thus

$$\mathrm{VaR}(c) = -Q_{1-c}^{\mathrm{cpf}_{\delta V}} = -\mu - \sqrt{\mu_2}\,Q_{1-c}^{\mathrm{cpf}_{\widetilde{\delta V}}}$$

where we now can use the approximation 22.50 for $Q_{1-c}^{\widetilde{\delta V}}$ since the percentiles of the standard normal distribution for the confidence level $(1-c)$ are well known (see for instance Equation 21.12).

22.3.5 Fourier-Transformation of the Portfolio Value Distribution

While a great deal of information about the distribution function can be gleaned from the computation of the moments with the moment generating function, we are still not able to calculate the distribution itself directly. *Characteristic functions* (CFs), however, generate the distribution directly (this can at least be accomplished numerically). As defined in Section A.3.2, the *characteristic function* Φ_x of a random variable x with density function $\mathrm{pdf}(x)$ is[12]

$$\Phi_x(s) \equiv \mathrm{E}[e^{isx}] = \int_{-\infty}^{\infty} e^{isx}\,\mathrm{pdf}(x)dx \tag{22.51}$$

This is precisely the definition of the *Fourier transformation* of the density function. As demonstrated in Section A.3.2, the CFs of many random variables with density

[12]Here, i denotes the *imaginary number* satisfying the property $i^2 = -1$, thus intuitively $i = \sqrt{-1}$.

functions can be computed explicitly. In particular, the CF of u_0, i.e. the CF of a normally distributed random variable $\mathrm{N}(0, \sum_{i \notin J} L_i^2)$ is given by

$$\Phi_{u_0}(s) = \exp\left(-\frac{1}{2}s^2 \sum_{i \notin J} L_i^2\right) \tag{22.52}$$

And according to Equation A.91, the CF of the u_j , i.e. of a non-central χ^2-distributed random variable with one degree of freedom $\chi^2(1, (L_j/\lambda_j)^2)$ is given by

$$\Phi_{u_j}(s) = \frac{1}{\sqrt{1-2is}} \exp\left\{\frac{is}{1-2is}\frac{L_j^2}{\lambda_j^2}\right\} \ , \ j \in J \ , \ i \equiv \sqrt{-1} \tag{22.53}$$

Similarly to the moment generating function, the usefulness of the CF in computing the distribution of δV in Equation 22.33 stems from the property A.34. The CF of a sum of independent random variables x, y is simply the product of the CFs of each of these variables:

$$\Phi_{x+y}(s) = \Phi_x(s)\, \Phi_y(s) \tag{22.54}$$

Furthermore, according to Equation A.35,

$$\Phi_{ax+b}(s) = e^{ibs}\Phi_x(as) \tag{22.55}$$

holds for all non-stochastic values a, b and random variables x. Thus the CF of δV can be expressed as the product of the CFs of the random variables appearing in the definition of δV:

$$\Phi_{\delta V}(s) = \exp\left\{-is \sum_{j \in J} \frac{L_j^2}{2\lambda_j}\right\} \Phi_{u_0}(s) \prod_{j \in J} \Phi_{u_j}(\frac{1}{2}\lambda_j s) \tag{22.56}$$

The CFs of the individual random variables are given explicitly in Equations 22.52 and 22.53. We now have all information at our disposal to calculate an *explicit* expression for the characteristic function of the distribution of δV. The result is of course the same as Equation 22.40 for the MGF with the obvious substitution $s \to is$:

$$\Phi_{\delta V}(s) = \exp\left(-\frac{1}{2}\, s^2 \sum_{i \notin J} L_i^2\right) \prod_{j \in J} \frac{1}{\sqrt{1-i\lambda_j s}} \exp\left\{-\frac{1}{2} L_j^2 \frac{s^2}{1-i\lambda_j s}\right\} \tag{22.57}$$

Since $\lambda_i = 0$ for $i \notin J$,we can – as we did with the MGF – write the first exp-function as

$$\exp\left(-\frac{1}{2}\, s^2 \sum_{i \notin J} L_i^2\right) = \prod_{i \notin J} \frac{1}{\sqrt{1-i\lambda_i s}} \exp\left\{-\frac{1}{2} L_i^2 \frac{s^2}{1-i\lambda_i s}\right\}$$

Thus, the characteristic function can be written as a product over *all* indexes j:

$$\Phi_{\delta V}(s) = \prod_{j=1}^{n} \frac{1}{\sqrt{1-i\lambda_j s}} \exp\left\{-\frac{1}{2} L_j^2 \frac{s^2}{1-i\lambda_j s}\right\} \text{ with } i \equiv \sqrt{-1} \tag{22.58}$$

In contrast to the moment generating function, there exists an inverse transformation for the characteristic function, namely the *inverse Fourier Transformation* (see Section A.3.2). From Equation 22.58 the density function $\mathrm{pdf}(\delta V)$ can thus be computed (at least numerically).

$$\begin{aligned}\mathrm{pdf}_{\delta V}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\Phi_{\delta V}(s)ds \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\prod_{j=1}^{n}\frac{1}{\sqrt{1-i\lambda_j s}}\exp\left\{-\frac{1}{2}L_j^2\frac{s^2}{1-i\lambda_j s}-isx\right\}ds \end{aligned} \tag{22.59}$$

The *cumulative* probability function of δV can now be obtained through the (numerical) integration of this probability density

$$\begin{aligned}\mathrm{cpf}_{\delta V}(c) &\equiv \int_{-\infty}^{c}\mathrm{pdf}_{\delta V}(x)dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{c}\int_{-\infty}^{\infty}\prod_{j=1}^{n}\frac{\exp\left\{-\frac{1}{2}L_j^2\frac{s^2}{1-i\lambda_j s}-isx\right\}}{\sqrt{1-i\lambda_j s}}ds\,dx\end{aligned}$$

A method which can likewise be applied in practice does not use the Fourier transformation of the density, but the Fourier transformation of the *cumulative* distribution directly:

$$F_x(s) \equiv \int_{-\infty}^{\infty} e^{isx}\,\mathrm{cpf}(x)dx$$

This Fourier transformation has the analogous properties to those indicated in Equations 22.55 and 22.54. Hence, the Fourier transformation of the cumulative distribution function of the portfolio's change is (up to a constant)

$$F_{\delta V}(s) \sim F_{u_0}(s)\prod_{j\in J}F_{u_j}(\frac{1}{2}\lambda_j s) \tag{22.60}$$

analogous to Equation 22.56. Here, the individual Fourier transformations of the *cumulative* distribution function of a normally distributed random variable (for u_0) and the *cumulative* distribution of the non-central χ^2 random variables with one degree of freedom (for the u_j) appear. These can only be computed numerically. Then, the results as shown in Equation 22.60 are multiplied in Fourier space. Finally, the function $F_{\delta V}(s)$ thus obtained must be transformed back with the inverse Fourier transformation to obtain the cumulative probability function of δV as

$$\mathrm{cpf}_{\delta V}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}F_{\delta V}(s)ds \tag{22.61}$$

The Wiener-Chintchine theorem states that the approach just described is equivalent to taking the *convolution* of the cumulative distribution functions. The twice computed Fourier transformation is numerically preferable to computing the convolution. The recommended method for numerically performing Fourier transformations (or inverse

Fourier transformations like Equation 22.61) is the *fast Fourier transformation* (FFT). This method requires significantly fewer computations as compared to other common procedures for numerical integration[13].

22.3.6 Monte Carlo Simulation of the Portfolio Value Distribution

Calculating the cumulative distribution of δV with characteristic functions involves complicated numerical procedures. Using moment-generating functions to calculate the moments we need additional assumptions and approximations to establish a relation between those moments and the distribution or the percentiles[14]. All methods introduced here therefore offer sufficient scope for error to creep into the calculations. Additionally, significant difficulties are often involved in calculating the gamma and covariance matrices. Hence, it is by all means legitimate to apply a simple Monte Carlo simulation, instead of the often complicated methods described above, to generate the distribution of δV. The statistical error in doing so is often no larger than the errors of the above methods assuming of course that a sufficient number of simulation runs have been carried out.

For a Monte Carlo Simulation we proceed directly from Equation 22.27. n normally distributed random numbers are generated with which the simulated change in the portfolio's value can be immediately computed with the expression

$$\delta V = \sum_{i=1}^{n} \left[L_i \, \delta X_i + \frac{1}{2} \lambda_i \delta X_i^2 \right]$$

This procedure is repeated N times (as a rule, several thousand times) thus obtaining N simulated changes in the portfolio's value from which the distribution of δV can be approximated. The percentiles of this distribution can be approximated by simply sorting the simulated values of δV in increasing order as described in Section 23.1 (a detailed discussion of value at risk computations using the Monte Carlo Method can be found in this section). The advantage of the method described above over the more general methods described in Section 23.1 is that a complicated revaluation of the portfolio for the generated scenarios is not necessary; the change in the portfolio's value is generated directly with Equation 22.27. However, before the simulation can be performed, the eigenvalues of the transformed gamma matrix must be calculated by solving Equation 22.16, and the transformed sensitivities L_i need to be determined as well. Because of Equations 22.29 and 22.28, both the Cholesky decomposition of the covariance matrix as well as the eigenvectors of the gamma matrix must be computed. The eigenvectors are determined by solving Equation 22.15.

[13] In contrast to other common numerical procedures, the FFT reduces the number of necessary multiplications from order $O(N^2)$ to $O(N \ln(N))$. See, for example [16] oder [133].

[14] It should not be forgotten that the Delta-Gamma method itself is only an approximation of the portfolio's value obtained from the second-order Taylor series approximation.

23

Simulation Methods

23.1 Monte Carlo Simulation

In the calculation of the value at risk by means of Monte Carlo simulations, all of the risk factors influencing a portfolio are simulated over the liquidation period δt as stochastic processes satisfying, for example, Equation 3.13 or even more general processes of the form 3.16. The value at risk of the risk factors themselves are taken into complete consideration using Equation 21.15 sometimes neglecting the drift in the simulation if the liquidation period is short:

$$\mathrm{VaR}_{\mathrm{long}}(c) \approx NS(t)\left[1-\exp\left(+Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right]$$
$$\mathrm{VaR}_{\mathrm{short}}(c) \approx -NS(t)\left[1-\exp\left(t-Q_{1-c}^{\mathrm{N}(0,1)}\sigma\sqrt{\delta t}\right)\right]$$

As explained in Section 21.1, the value at risk of a long position in an underlying is only then equal to that of a short position if the drift is neglected *and* the linear approximation has been used. Since the linear approximation is usually not assumed in the Monte Carlo method, the VaRs of a long position will not equal that of a short position on the same underlying.

In carrying out the simulation, it will be taken into consideration that the risk factors are not independent of one another, but are *correlated*. This has already been demonstrated in Section 12.3.2 for the case of *two* correlated underlying prices. Processes of the form 3.13 involve Wiener processes whose stochastic components are coupled as given by the covariance matrix 21.21. In other words, the logarithmic changes in the risk factors are multivariate normally distributed with the covariance matrix 21.21. In this way, market scenarios (combinations of all risk factors) possibly occurring up to the end of the liquidation period (up to time T) are simulated. The *portfolio* values at the conclusion of the liquidation period are then computed on the basis of all these simulated market scenarios. With this information, the distribution of the potential *portfolio* values at the conclusion of the liquidation period can be approximated. The value at risk can then be obtained through the statistical evaluation of this portfolio value distribution.

The advantage of the Monte Carlo simulation compared to other methods (such as the variance-covariance method) is that for the portfolio valuation for each market scenario we can in principle use the same valuation methods as for determining the portfolio's *current* value. No additional approximations for the *valuation* of the financial instruments in the portfolio need to be made (called *full valuation*). From the aspect of the valuation, the value at risk is, in principle, just as precise as the current pricing. Nevertheless, it is in practice often not possible to use the same (computationally intensive) routines for both the valuation of a portfolio with respect to, for example, 10,000 scenarios as for (one single) determination of the portfolio's current value, the *mark-to-market*. It is therefore often necessary to use simpler and less precise methods for the revaluation of financial instruments with respect to the Monte Carlo scenarios. The statistical error arising in connection with such simulation methods is also unavoidable since only a finite number of scenarios can be simulated and thus only mean values rather than expectations can be computed (see Section 29.3 for more on this subject).

23.1.1 The Risk Factors as Correlated Random Walks

A random number generator usually produces single, uncorrelated random numbers. However, with the methods described in Section 21.4.2, via the Cholesky decomposition **A** of the covariance matrix, uncorrelated random numbers can be transformed into correlated ones. This can be exploited when carrying out Monte Carlo simulations:

With the help of the Cholesky decomposition **A** of the covariance matrix standard normally distributed random variables X_j are transformed into components of a multivariate normally distributed random vector Y_j, for $j = 1, 2, ..., n$ whose covariances are given in Equation 21.21:

$$Y_i = \sum_{j=1}^{n} A_{ij} X_j$$

The random walks of the risk factors expressed in terms of the random variables Y_j are

$$d \ln S_j(t + \delta t) = \mu_j \, \delta t \; + Y_j$$

Similar to Equation 12.2, the values of the risk factors for a scenario simulated to occur at the end of the time interval δt are

$$\begin{aligned} \ln S_j(t + \delta t) &= \ln S_j(t) + \mu_j \, \delta t \; + Y_j \\ S_j(t + \delta t) &= e^{Y_j} e^{\mu_j \, \delta t} S_j(t) \quad , \quad j = 1, \; \ldots \; , n \end{aligned} \tag{23.1}$$

where here, as has received mention on numerous occasions, the drifts μ_j are often neglected in the analysis.

To generate a complete market scenario for the time $t + \delta t$ a random number Y_j for each risk factor is required. The portfolio is then revaluated at the value date $t + \delta t$ on the basis of this scenario. We thus obtain a simulated portfolio value at the end of the liquidation period. The approach in a Monte Carlo simulation in risk management is summarized below. This type of simulation is sometimes called *structured Monte Carlo*.

23.1.2 Structured Monte Carlo

Simulation

- Generate n standard normally distributed, uncorrelated random numbers X_j, one for each risk factor.
- Generate correlated random numbers Y_j, $j = 1, 2, ..., n$ using the equation

$$Y_i = \sum_{j=1}^{n} A_{ij} X_j$$

  The elements of the matrix $\mathbf{A}$ are given through the Cholesky decomposition of the covariance matrix in accordance with Equation 21.35.
- Using these Y_j, the risk factors for the simulated scenario at the end of a time interval δt are calculated via Equation 23.1. If the portfolio contains path-dependent derivatives, it is not possibly to simply jump to the end of the liquidation period in a single step if the liquidation period is longer than one day. Smaller steps are necessary to simulate the paths of the risk factors up to the conclusion of the liquidation period similar to Equation 12.1, instead of simulating directly with Equation 12.2.
- Perform a new valuation of the portfolio with respect to these simulated risk factors. If computationally possible a *full valuation* is preferable[1].

Thus, one single market scenario is simulated and the portfolio is re-valued with respect to this single scenario. This simulation is now repeated (for example, 10,000 times) in order to generate numerous scenarios and a portfolio value for each of these scenarios. Finally, the statistical evaluation is performed.

Evaluation

The change in the value of the portfolio observed in the i^{th} simulated scenario will be denoted by δV_i, the vector containing all risk factor values in the i^{th} simulated scenario by $\mathbf{S}_i$. We let m denote the number of simulations and n the number of relevant risk factors. For every simulated scenario, the induced simulated value change of the portfolio is the difference between the portfolio's value with respect to the simulated scenario and its current value:

$$\delta V_i = V\left(\mathbf{S}_i(t + \delta t)\right) - V\left(\mathbf{S}(t)\right) \quad \text{for} \;\; i = 1, \; \ldots \; , m \tag{23.2}$$

We thus obtain m simulated value changes. The value at risk of the Monte Carlo simulation is the minimum of these δV_i, where a certain number of the least favorable value changes are ignored dependent on the desired confidence level. For 95% confidence,

[1] If the portfolio valuation requires for example Monte Carlo *pricing* methods for some (exotic) financial instruments, these additional Monte Carlo simulations (for pricing) have to run *inside* the simulation loop for the VaR calculation. Clearly this may lead to unacceptably large computation times.

for example, these are 5% of the least favorable value changes. For 10,000 simulated scenarios, for example, the 500 worst scenarios are ignored. We denote by δV_{1-c} the most favorable of the value changes which are ignored at a level of confidence c. The value at risk of the portfolio is now the least favorable result among the set of results remaining after those $(1-c)\%$ least favorable simulations have been removed from consideration or equivalently, the least favorable value greater than V_{1-c}:

$$\mathrm{VaR}_V(c) = -\min_i \{\delta V_i \mid \delta V_i > \delta V_{1-c}\} \quad \text{for} \quad i = 1, \ldots, m \tag{23.3}$$

With a confidence c, the portfolio will depreciate in value by no more than this value at risk by the end of the liquidation period.

23.2 Historical Simulation

Historical simulations are performed by investigating historical time series with the objective of identifying market changes which have actually occurred in the past and using these changes to compute the value at risk. The covariance matrix in Equation 21.21 is not necessary for a historical simulation nor is it necessary to assume that the risk factors behave as random walks with constant yields and volatilities or even that they behave as random walks at all! This freedom from model assumptions is the primary advantage of this method.

The independence from model assumptions is at the expense of very involved data management. While only three values provide sufficient statistical information about the past behavior of two risk factors (both volatilities and the correlation between the two) for variance-covariance and Monte Carlo methods, entire time series of prices for all risk factors relevant to the portfolio must be kept for a historical simulation to be performed. For example, the closing prices of every underlying for the previous 250 days. For two underlyings, this amounts to 500 values in comparison to just the 3 required for the methods mentioned above.

From the historical time series, the value changes $\delta S_j(\delta t_i)$ of all risk factors S_j over time intervals δt_i with the same length as the liquidation period are determined over the entire available history of the risk factors[2]:

$$\delta S_j(\delta t_i) = S_j(t - i\,\delta t + \delta t) - S_j(t - i\,\delta t) \quad \text{with} \quad i = 1, \ldots, m \;\; ; \;\; j = 1, \ldots, n \tag{23.4}$$

For example, the time series over 250 days yields 249 daily changes or 240 changes for a liquidation period of 10 days[3].

The historical risk factor changes applied to the price of the risk factors at time t (today) provide m different scenarios. For reasons of consistency, the *relative* changes should be applied to today's price rather than the absolute changes[4]. For each scenario

[2] The number of available liquidation periods obtained from the historical time span for which data is available is denoted by m, the number of relevant risk factors again by n.

[3] In the second case, the liquidation periods overlap resulting in auto-correlations.

[4] A historical change for example, a 12 point change in the DAX index which stood at 1200 at the outset of a liquidation period is quite different from a 12 point change when the DAX is at 7000. A relative change of 1% (i.e., 70 points) is therefore more suitable.

i the thus induced value change δV_i of the portfolio V is computed. This can be accomplished with a full valuation of the portfolio. However, in practice it is often the case that a simple linear (delta valuation) or quadratic (delta-gamma valuation) approximation as in Equation 22.2 is performed.

In this way, m "historical" value changes $\delta V\,(t_i)$ are generated from the past time series data. The value at risk of a historical simulation is now the minimum of of all δV_i, where - similar to the Monte Carlo simulation - unfavorable changes in the portfolio's value falling outside a previously specified confidence interval are ignored. For a confidence level of 95%, for example, and a history consisting of 250 days, the 12 worst out of the 249 portfolio value changes are not considered when finding the minimum over daily changes.

If δV_{1-c} denotes the most favorable change among the ignored value changes at a confidence level of c, then the value at risk of the portfolio is the least favorable portfolio change greater than V_{1-c}:

$$\mathrm{VaR}_V(c) = -\min_i \{\delta V_i \mid \delta V_i > \delta V_{1-c}\} \quad \text{with} \quad i = 1, \ldots, m$$

With a confidence c, the portfolio at the end of the liquidation period depreciates by an amount no larger than this value at risk.

At this point we can clearly see the greatest disadvantage of this method: the weak statistical information on the basis of which the probabilistic conclusions such as confidence levels are drawn. Despite the immense effort in data management of all relevant historical time series, usually only a dozen (in the above example) or so values remain for the final analysis, namely those falling below the lower boundary of the confidence interval. The probabilistic conclusion is drawn on the basis of these few values. In contrast, the statistical basis deriving from 10,000 Monte Carlo simulation runs is approximately 50 times larger (of course, this statistical advantage of the Monte Carlo method is at the expense of assuming that the risk factors are random walks).

A further disadvantage of historical simulations is the following effect: For each change of position in a portfolio (after each transaction), the new portfolio (and its value changes) must be recalculated for all 250 days. In doing so, it may happen that *another* historical risk factor change affects the make-up of the set falling outside the confidence interval for the new portfolio so that suddenly the value at risk based on *another scenario* is relevant. Thus, a trader, after entering into a transaction intended to optimize the VaR according to the *original* scenario, is then informed of a value at risk computed on the basis of *another* scenario. This greatly increases the difficulty of evaluating the success of the transaction.

For the sake of consistency, the following is also to be considered when conducting historical simulations: if we wish to avoid model assumptions by carrying out historical simulations (at the cost of the difficulties described above in determining the scenarios), then we should be consistent in doing so throughout the *valuation*, too. In particular, pricing models which are based on the assumption that the underlyings are random walks should then in principle not be used. This of course significantly increases the level of difficulty involved in pricing, in particular, of options. For example, neither the Black-Scholes models nor most of the numerical methods discussed in this book could be utilized in this case. This level of consistency between risk calculation and pricing

is therefore almost never feasible in practice.

23.3 Crash and Stress Testing: Worst Case Scenarios

Each value at risk concept introduced up to this point yields the *potential loss* in the course of a liquidation period and the *probability* with which this loss occurs. The confidence levels most commonly used are 95% or 99%. This means that losses amounting to the value at risk or higher actually occur between 2 and 12 times per year, as they actually should. Otherwise the model on the basis of which these probabilities are derived is incorrect. The value at risk can thus not be considered a worst case scenario, but rather as part of daily business: losses of this magnitude must occur on average once a month at a confidence level of 95%! Accordingly, the value of these losses must be kept below an acceptably small limit.

In order to obtain a measure of a portfolio's risk should a catastrophe occur, a *worst case* or *crash scenario* is constructed by hand through the explicit specification of all risk factors influencing the portfolio. The portfolio is then revaluated on the basis of this market scenario. Such a scenario could, for example, be the crash of October 1987. The difference between the calculated value and the current portfolio value is the "value at risk" of the portfolio with respect to the crash-scenario. Obviously, this value expresses the potential loss as a result of the crash but no information is available about the *probability* of such an event.

A further method used to get a feeling for the risk of a portfolio in the case of rare and very unfavorable market developments is to use 6 or 8 standard deviations from the expectation as the boundary of the confidence interval rather than the usual 1.65 or 2.36. This approach is sometimes referred to as a *stress test.* In this way, a potential loss is obtained as well as a *theoretical* probability that a loss of this magnitude is incurred. For example, for a standard normal distribution, the probability of a loss of more than six standard deviations in a one-sided confidence interval is approximately one to one billion:

$$1 - \frac{1}{\sqrt{2\pi}} \int_{-6}^{\infty} e^{-x^2/2} dx \approx 9,86610^{-10} \approx 10^{-9}$$

No great importance should be attached to such probability statements since the random walk assumptions, constant volatilities and correlations, for instance, are in all probability no longer satisfied when such events occur. As a rule, market scenarios of this type change the correlations drastically and the volatilities explode. Thus, de facto, stress tests, like crash tests, provide information on the potential loss involved without specifying the probability of such an event.

	Variance-Covariance	Monte Carlo	Historical Simulation	Stress Simulation
Models Risk Factors as Random Walks	−	−		
Assumes Constant Vol & Correlation	−	(−)		
Requires Historical Time Series			−	
Requires Vol & Correlations	(−)	(−)		
Neglects the Mean Yield	−			
Linear Proxy for the Risk Factors	−		(−)	
Delta(-Gamma) Proxy for Prices		−	−	−
Full Valuation		+	+	+
Specified Scenarios	+	+	+	
Specified Probabilities	+	+		
Based on Large Data Sets		+		
Valid for Long Liquidation Periods		+		+
Takes Vega Risk into Account		(+)	+	+
Takes Theta Risk into Account		+	+	+

Table 23.1: The pros and cons of the most common value at risk methods.

Advantages and Disadvantages of the Commonly Used Value at Risk Methods

In Table 23.1 the advantages and disadvantages of the VaR methods introduced above are summarized. A "+" in the method column indicates that this particular method has the advantage of the property associated with the corresponding row. A "-" means that it has the disadvantage of the corresponding row. No entry indicates that the method does not have the property of the row under consideration. A symbol in parentheses means that the indicated property is usually assumed in the application of the method but that the advantage or disadvantage is not, in *principle*, characteristic of the method.

24

Interest Rate Risk and Cash Flows

24.1 Cash Flow Structures of Financial Instruments

A consistent relationship between *interest rate* risk factors and the changes in the value of a financial instrument induced by them is established through the decomposition of the financial instrument into *cash flows* insofar as such a decomposition is possible. The philosophy behind this approach is that the value of any financial instrument is equal to the sum of all discounted, expected future cash flows. The decomposition into expected future cash flows can, in general, be accomplished to a point where the *size* of the cash flow no longer depends explicitly on the interest rate (in the case of interest rate options, this only holds when the option price is approximated linearly as a function of the interest rate). The *interest rate risk* of an instrument then consists solely in the fact that the discount factors used in computing the instrument's present value at a specified value date change with changes in the interest rates. However, the size of the cash flows can by all means be affected by other risk factors; these could be foreign exchange risk or other price risks influencing the value of the instrument.

In order to more clearly illustrate the decomposition described above, we consider the example of a plain vanilla call on a stock price (or foreign currency exchange rate) $S(t)$. The interpretation of the Black-Scholes equation given in Section 11.4.1 enables us to replicate a call by buying $\Delta = B_q(t,T)\mathrm{N}(x)$ underlyings at the spot price $S(t)$ on the current value date t, while partially financing the costs incurred with a loan in the amount of $B_r(t,T)K\mathrm{N}(x-\sigma\sqrt{T-t})$. The remaining costs for the purchase of the underlying are paid for "out of pocket". These costs are precisely the option premium. The loan must be repaid, with interest, upon maturity of the option at time T. Since the loan must be compounded with $B_r(t,T)^{-1}$ over the lifetime of the option, the total amount to be repaid at T is $K\mathrm{N}(x-\sigma\sqrt{T-t})$. A plain vanilla call can therefore be decomposed into two cash flows:

- an *incoming* cash flow as of the present value date t in the amount of the value of the stocks purchased, $+B_q(t,T)\mathrm{N}(x)S(t)$

- an *outgoing* cash flow at option's maturity T in the amount of the strike multiplied by the (risk-neutral) probability of exercise, $-K\mathrm{N}(x - \sigma\sqrt{T-t})$

The first cash flow occurs at the *current* value date and must not be discounted. The amount of the cash flow is its current value. It follows that no interest rate risk is associated with this payment. In our example, this amount depends solely on the current price of the underlying (and its dividend). The risk associated with the underlying price is expressed by this value; it is precisely the *delta risk* and, as such, equals the coefficient $+B_q(t,T)\mathrm{N}(x)$ of $S(t)$, i.e. the delta of the call.

The second cash flow occurs at a future time and is thus exposed to interest rate risk since it must be discounted back to the present date with the discount factor $B_r(t,T)$. This is precisely the *rho risk;* the amount of the cash flow (multiplied by $dB_r(t,T)/dr$) yields the rho of the call.

The procedure described above is somewhat more exact (for example, not only the risk but the price of the option is correctly replicated by the described decomposition) than the frequently quoted interpretation of an option as a "delta-weighted underlying" as prescribed in the *standard methods* of the Capital Adequacy Directive (*CAD*). In our example, this interpretation would have generated only the first of the two cash flows. This aspect of a call option, namely that it is not simply a delta-weighted underlying but includes a partial financing *as well* (for a put, the opposite position is assumed: the sale of delta-weighted underlyings plus a partial investment of the proceeds) is not accounted for by the standard methods of the Capital Adequacy Directive. This means that for stock, foreign exchange and commodity options, the rho risk is neglected. For interest rate options, this has the consequence that a part of the underlying risk is ignored. This imprecision is neither necessary nor justified since the computation of the cash flows generated from the financing or investment in the case of a call or put, respectively, is no more difficult to calculate than the delta-weighting of the underlying. Furthermore, the size of these cash flows can by all means be of the same order as the delta-weighted underlying (for example, for at the money options). Below, we will systematically present the cash flows for the plain vanilla interest rate instruments (specifically for a long position in these instruments) treated in Part III. Here, we will generally use the same notation as in the corresponding sections of Part III. Most of the work in constructing the cash flow structures has already been done in Part III. In particular, the cash flows for interest rate instruments in the spot market were presented explicitly in most cases. The cash flows can also often be inferred directly from the pricing equations: The factors appearing as coefficients of the spot rate discount factors identify the cash flows occurring at the end of the corresponding lifetime of the discount factor, as long as the size of the cash flow itself is not explicitly dependent on an interest rate (in the form of forward rates, for instance).

24.1.1 Spot Transactions

Although the cash flow structures of spot instruments have largely been presented explicitly in Chapter 16, we again provide an overview of these cash flows here for the sake of completeness.

The *zero bond* in Section 16.1 has the following cash flow

- at maturity T: an *in*coming cash flow in the amount of the principal

$$+N$$

From its pricing formula, the *floater* in Section 16.2 can also be represented as a single cash flow

- at the next fixing date t_{m+1}: an *in*coming cash flow in the amount of the principal compounded with the previously fixed interest rate

$$+N\ B_R^{\text{fix}}(t_m, t_{m+1})^{-1}$$

The *coupon bond* in Section 16.3 has the following cash flows

- at the future coupon payment dates t_i, $i = m+1, \ldots, n-1$: *in*coming cash flows in the amount of the coupon yield on the principal earned over the previous coupon period

$$+N[B_K(t_{i-1}, t_i)^{-1} - 1]$$

- at the maturity date t_n: an *in*coming cash flow in the amount of the last coupon plus the face value

$$+NB_K(t_{n-1}, t_n)^{-1}$$

The *payer swap* in Section 16.4 has the following cash flows

- at the next fixing date t_{m+1}: an *in*coming cash flow in the amount of the face value compounded at the variable interest rate fixed at the start t_m of the current swap period and an *out*going cash flow in the amount of the interest earned on the principal at the fixed rate, both taken over the previous coupon period. In total

$$+N[\ B_R^{\text{fix}}(t_m, t_{m+1})^{-1} - B_K^{\text{fix}}(t_m, t_{m+1})^{-1} + 1]$$

- at each future fixing date t_i, $i = m+2, \ldots, n-1$: an *out*going cash flow in the amount of the fixed leg interest earned on the principal over the respective swap period

$$-N[B_K(t_{i-1}, t_i)^{-1} - 1]$$

- upon maturity t_n: an *out*going cash flow in the amount of the last interest payment plus the face value

$$-NB_K(t_{i-1}, t_i)^{-1}$$

Analogously, the *annuity loan* described in Section 16.5 has the coefficients of the spot rate discount factors in Equation 16.23 as cash flows, each occurring at the end of the respective lifetime of the discount factor, where the residual debt is to be substituted as indicated by Equation 16.21.

24.1.2 Forward Transactions

The *FRA* in Section 17.1 has the following cash flows

- on the fixing date T, i.e. at the beginning of the forward period: an *in*coming cash flow, in the amount of the principal

$$N$$

- at the end T' of the forward period: an *out*going cash flow in the amount of the principal compounded at the fixed forward rate

$$-NB_K(T,T')^{-1}$$

The *future* on a zero bond in Section 17.2.1 has the following cash flows

- at maturity T of the future: an *out*going cash flow in the amount of the fixed (clean) delivery price K (in %) multiplied by the principal

$$-NK$$

- at maturity T' of the underlying (the zero bond): an *in*coming cash flow in the amount of the principal

$$+N$$

The *future on a coupon bond* in Section 17.2.2 has the following cash flows

- at maturity T of the forward transaction: an *out*going cash flow in the amount of the agreed upon (clean) delivery price K times the principal multiplied by the conversion factor P of the cheapest-to-deliver bond plus the accrued interest Z of the CTD

$$-(NKP+Z)$$

- at each coupon payment date t_i of the CTD occurring after the maturity T of the future, i.e. on the dates t_i, with $i = m+1,\ldots,n-1$ and $t_{m+1} > T$: an *in*coming cash flow in the amount of the CTD's coupon yield on the principal over the respective coupon period

$$+N[B_K(t_{i-1},t_i)^{-1}-1]$$

- at maturity $t_n = T'$ of the underlying, i.e. at the redemption date of the CTD: an *in*coming cash flow in the amount of the last CTD coupon plus the nominal

$$+NB_K(t_{n-1},t_n)^{-1}$$

The *forward payer swap* in Section 17.3 has the following cash flows

- at the last fixing date T_0 before (or at) the maturity T of the forward transaction: an *in*coming cash flow in the amount of the principal (this cash flow represents the floating leg of the swap)
$$+N$$

- at each fixing date after the maturity T of the forward transaction, i.e. at times T_i, with $i = 1, \ldots, n-1$ and $T_1 > T$: an *out*going cash flow in the amount of the fixed interest earned on the principal over the respective swap period
$$-N[B_K(t_{i-1}, t_i)^{-1} - 1]$$

- at the maturity T' of the underlying, i.e. at the last payment date T_n of the swap: an *out*going cash flow in the amount of the last fixed interest earned on the principal plus the principal
$$-NB_K(T_{n-1}, T_n)^{-1}$$

Analogously, the *forward bond* in Section 17.4 has the coefficients of the spot rate discount factors in Equation 17.14 as cash flows, each occurring at the end of the respective lifetime of the discount factor.

24.1.3 Options

The determination of the expected cash flows of interest rate options nicely demonstrates the application of the hedging deliberations in Chapter 13. As mentioned at the outset, for this purpose the option is replicated by a portfolio consisting of an underlying and cash (a zero bond) and thus, as indicated in Equation 11.5, can be represented by the hedging portfolio:

$$\text{Option} = \text{Delta} \times \text{Underlying} + \text{Zero Bond}$$

For a call, the replicating portfolio is long in the underlying (delta is positive) and short in the zero bond (a loan which partially finances the purchase of the underlyings). For a put, the replicating portfolio is short in the underlying (delta is negative) and long in the zero bond (a partial investment of the proceeds from the sale of the underlyings).

The computation of the required deltas have already been carried out in Section 13.4 for all possible combinations of options and underlyings. The following cash flow determination is thus only a demonstration of the application of the extremely powerful Equation 13.16 together with Table 13.1

Options on Bonds

The Call A *call* with maturity date T on a bond with a coupon yield R (and a face value of 1) is priced in Section 18.4.1 using the Black-76 equation 9.10, where the dirty forward price $S(t, T)$ of the underlying bond and the dirty strike K appear in the equation. For the dirty forward price, we now formally introduce a "dividend yield" q

of the bond. This represents the yield earned from the coupon payments during the lifetime of the option resulting from the coupon R (not the accrued interest, but the real coupon payments, should they occur). This means that compounding with q up to maturity T must yield the same amount as coupon payments occurring during the lifetime of the option, compounded from the respective coupon payment date up to maturity; formally

$$B_q(t,T)^{-1} - 1 = \sum_i \left[B_R(t_{i-1},t_i)^{-1} - 1\right] B_r(t_i,T)^{-1} \quad \text{with} \quad t < t_i \leq T$$

This "dividend yield" allows us to write the dirty forward price in 7.1 as

$$S(t,T) = \frac{B_q(t,T)}{B_r(t,T)} S(t)$$

This forward price substituted into the Black-76 model gives, of course, formally the Black-Scholes equation of a call, for example:

$$c(t,T,K) = B_q(t,T)\ S(t)\mathrm{N}(x) - B_r(t,T)K\mathrm{N}(x - \sigma\sqrt{T-t})$$

We now set this option price equal to the value of the replicating portfolio, composed of an amount $g(t)$ of cash and $\triangle(t)$ bonds with the current price $S(t)$:

$$c(t,T,K) = \triangle(t)S(t) + g(t)$$

The required delta is obtained by taking the derivative of the expression in 13.16 with respect to the spot price of the underlying bond. The needed derivative can be found in Table 13.1 and is simply "call on the spot with respect to the spot" (the notation is the same as in Section 13.4) :

$$\triangle(t) = B_q(t,T)N(x)$$

The cash sum $g(t)$ must then be equal to the difference between the option price and the bond position weighted by delta:

$$\begin{aligned} g(t) &= c(t,T,K) - \triangle(t)S(t) \\ &= B_q(t,T)\ S(t)\mathrm{N}(x) - B_r(t,T)K\mathrm{N}(x - \sigma\sqrt{T-t}) \\ &- B_q(t,T)\mathrm{N}(x)S(t) \\ &= -B_r(t,T)K\mathrm{N}(x - \sigma\sqrt{T-t}) \end{aligned}$$

This cash sum is negative. In the framework of a replicating portfolio, it can be interpreted as a loan with repayment due upon the maturity of the option entered into with the purpose of partially financing the purchase of the underlying. Through compounding, $g(t)$ accumulates interest until T at which time it is equal to $K\mathrm{N}(x - \sigma\sqrt{T-t})$. This is the amount due for repayment at T, the maturity of the option. As such, it is an *out*going cash flow. Together with the "delta bonds" making up the bond position, the cash flows of the replicating portfolio for a call on a bond are

- at maturity T of the option: an *out*going cash flow in the amount of the compounded loan $g(t)$

$$-KN(x - \sigma\sqrt{T-t})$$

where K is the dirty strike (=clean strike plus the accrued interest up to T)

- at each coupon payment date t_i of the bond after the option's maturity, i.e. at each t_i with $i = m+1, \ldots, n-1$ and $t_m + 1 > T$: an *in*coming cash flow in the amount of the coupon yield earned on Δ (the sum of the face values in the bond position) during the previous respective coupon period

$$\begin{aligned} +\Delta(t)[B_R(t_{i-1}, t_i)^{-1} - 1] &= +B_q(t,T)\mathrm{N}(x)[B_R(t_{i-1}, t_i)^{-1} - 1] \\ &= \frac{\mathrm{N}(x)[B_R(t_{i-1}, t_i)^{-1} - 1]}{\sum_j [B_R(t_{j-1}, t_j)^{-1} - 1]\, B_r(t_j, T)^{-1} + 1} \\ \text{where} \quad t < t_j \leq T \end{aligned}$$

- at the maturity of the underlying, i.e. upon maturity t_n of the bond: an *in*coming cash flow in the amount of the last coupon plus the principal

$$\begin{aligned} +\Delta(t)B_R(t_{n-1}, t_n)^{-1} &= \frac{\mathrm{N}(x)B_R(t_{n-1}, t_n)^{-1}}{\sum_j [B_R(t_{j-1}, t_j)^{-1} - 1]\, B_r(t_j, T)^{-1} + 1} \\ \text{where} \quad t < t_j \leq T \end{aligned}$$

As yet, the decomposition of the option is exact. Up to this point, no approximation has entered into the discussion. In particular, discounting these cash flows with the current discount factors reproduces the option price on the value date t. This is immediately apparent for the construction of the cash flows but can, of course, be explicitly verified as well.

The linear approximation comes into play when saying that the *size* of the cash flows determined up to now is *independent* of the interest rates. Strictly speaking, Δ still depends on the interest rate: the argument x of the standard normal distribution depends on the spot rates, and spot rate discount factors are also hidden in B_q. The linear approximation now means that delta is *assumed* to be constant, in particular, independent of the interest rate, and consequently the size of the above cash flows can be assumed to be constant *in linear approximation*. The interest rate risk of this cash flow structure now only appears in the form of a change in the *present value* (and not in the *size*) of the expected future cash flows, resulting from discounting with stochastic discount factors.

The Put The approach in dealing with the corresponding *put* is completely analogous: The option price p is set equal to the value of the replicating portfolio consisting of a cash amount $g(t)$ and delta bonds at the current price $S(t)$:

$$p(t, T, K) = \Delta(t)S(t) + g(t)$$

The necessary delta is obtained from Equation 13.16 by taking the derivative with respect to the spot price $S(t)$. The required derivative in Table 13.1 is simply "put on the spot with respect to the spot" (the notation is the same as in Section 13.4):

$$\Delta(t) = -B_q(t,T)\mathrm{N}(-x)$$

The cash amount $g(t)$ equals the difference of the option price and the price of the delta bonds:

$$\begin{aligned} g(t) &= p(t,T,K) - \triangle(t)S(t) \\ &= -B_q(t,T)\ S(t)\mathrm{N}(-x) + B_r(t,T)K\mathrm{N}(-x+\sigma\sqrt{T-t}) \\ &+ B_q(t,T)\mathrm{N}(-x)S(t) \\ &= +B_r(t,T)K\mathrm{N}(-x+\sigma\sqrt{T-t}) \end{aligned}$$

The amount is positive. In the framework of the replicating portfolio, it is invested until the maturity of the option. The compounding up to time T increases the value of the investment $g(t)$ to $K\mathrm{N}(-x+\sigma\sqrt{T-t})$. This is the amount we get back at the maturity of the option and as such is an incoming cash flow. Together with the "delta bonds" in the bond position, the cash flows of the replicating portfolio for the put on a bond are

- at maturity T of the option: an *in*coming cash flow in the amount of the compounded investment $g(t)$

  $$KN(-x+\sigma\sqrt{T-t})$$

  where K is the dirty strike (=clean strike plus the interest accrued up to time T)

- at each coupon payment date t_i after the maturity of the option, i.e. at each t_i with $i = m+1,\ldots,n-1$ and $t_m + 1 > T$: an *out*going cash flow in the amount of the bond's coupon yield earned on delta multiplied by the face value of the bond over the previous respective coupon period

  $$\begin{aligned} \Delta(t)[B_R(t_{i-1},t_i)^{-1}-1] &= -B_q(t,T)\mathrm{N}(-x)[B_R(t_{i-1},t_i)^{-1}-1] \\ &= -\frac{\mathrm{N}(-x)[B_R(t_{i-1},t_i)^{-1}-1]}{\sum_j [B_R(t_{j-1},t_j)^{-1}-1]\, B_r(t_j,T)^{-1}+1} \end{aligned}$$

  where $t < t_j \le T$

- at maturity of the underlying, i.e. on the repayment date t_n of the bond: an *out*going cash flow in the amount of the last coupon plus the bond's face value

  $$\Delta(t)B_R(t_{n-1},t_n)^{-1} = -\frac{\mathrm{N}(-x)B_R(t_{n-1},t_n)^{-1}}{\sum_j [B_R(t_{j-1},t_j)^{-1}-1]\, B_r(t_j,T)^{-1}+1}$$

  where $t < t_j \le T$

Options on Bond Futures

The Call A call with maturity T on a bond future with maturity T' (and face value = 1 monetary unit) is priced with the Black-Scholes equation for futures options, Equation 18.4.2 as described in Section 18.4.2. Here, the *clean* forward price of the underlying bond and the clean strike K enter into the equation since the difference of the two equals the payoff should the option be exercised. We set this option price c_F equal to the value of the replicating portfolio consisting of a cash sum $g(t)$ and delta futures with delivery price K and a current value of F_S:

$$c_F(t,T,K) = \Delta_F(t) F_S(t,T',K) + g(t)$$

The required delta is obtained from Equation 13.16 by taking the derivative with respect to the spot price of the underlying bond. The needed derivatives (call on a future with respect to the spot and the future with respect to the spot) can be taken directly from Table 13.1 (the notation is the same as in Section 13.4):

$$\begin{aligned}\Delta_F(t) &= B_q(t,T')B_r(T,T')^{-1}\mathrm{N}(x')\left[B_q(t,T')B_r(t,T')^{-1}\right]^{-1} \\ &= \frac{B_r(t,T')}{B_r(T,T')}\mathrm{N}(x') = B_r(t,T)\mathrm{N}(x')\end{aligned}$$

where Equation 3.4 was used in the last step. The cash amount $g(t)$ can now be expressed as the difference between the option price and the delta futures position. With Equation 18.4.2 for the option price and Equation 7.5 giving the value of the futures position, we establish the following equation for $g(t)$:

$$\begin{aligned}g(t) &= c_F(t,T,K) - \Delta_F(t)F_S(t,T',K) \\ &= c_F(t,T,K) - B_r(t,T)\mathrm{N}(x')\left[S(t,T') - K\right] \\ &= B_r(t,T)\left[S(t,T')\mathrm{N}(x') - K\mathrm{N}(x' - \sigma\sqrt{T-t})\right] \\ &\quad - B_r(t,T)\mathrm{N}(x')\left[S(t,T') - K\right] \\ &= B_r(t,T)K\left[\mathrm{N}(x') - \mathrm{N}(x' - \sigma\sqrt{T-t})\right]\end{aligned}$$

This cash sum, in contrast to the call on a spot, is *positive*. We do not need to borrow money in order to purchase the "delta" underlying futures since a future is always worth less than its corresponding option. The difference $g(t)$ between the option price (which is paid out of pocket) and the (lower) price for the corresponding underlying is, in the framework of a replicating portfolio, invested up to maturity of the option. The amount $g(t)$ is compounded and thus increases to a value of $K\left[\mathrm{N}(x') - \mathrm{N}(x' - \sigma\sqrt{T-t})\right]$ at option's maturity. This is the amount received at maturity and, as such, is an *incoming* cash flow (this cash flow would be ignored were the option treated simply as a delta-weighted underlying.)

The cash flows generated through the other component of the replicating portfolio, namely the future, have already been determined in the Section 24.1.2. We adopt those results here to complete the cash flow structure of the option. It should be noted that only "delta" futures are contained in the replicating portfolio.

In total, the call on a bond future can be represented by the following cash flows generated by the replicating portfolio:

- at maturity T of the option: an *in*coming cash flow in the amount of the cash investment $g(t)$ compounded up to this time

$$K\left[\mathrm{N}(x') - \mathrm{N}(x' - \sigma\sqrt{T-t})\right]$$

- upon maturity T' of the future: an *out*going cash flow in the amount of the agreed upon (clean) delivery price K times the principal multiplied by the conversion factor P of the Cheapest-to-Deliver bond (with a face value of 1) in addition to the CTD's accrued interest Z, for each of the "delta" futures

$$-\Delta_F(t)(KP + Z) = -B_r(t,T)\mathrm{N}(x')(KP + Z)$$

- at each coupon payment date t_i of the CTD after maturity $T'$of the future, i.e. on each date t_i with $i = m+1, \ldots, n-1$ and $t_m + 1 > T'$: an *in*coming cash flow in the amount of the CTD coupon yield on the face value for each of the Δ_F futures over the respective previous coupon period

$$+\Delta_F(t)\left[B_R(t_{i-1}, t_i)^{-1} - 1\right] = +B_r(t,T)\mathrm{N}(x')\left[B_R(t_{i-1}, t_i)^{-1} - 1\right]$$

- at maturity of the underlying CTD, i.e. on the redemption date t_n of the CTD: an *in*coming cash flow in the amount of the last CTD coupon plus the CTD's face value for each of the Δ_F futures

$$+\Delta_F(t)B_R(t_{n-1}, t_n)^{-1} = +B_r(t,T)\mathrm{N}(x')B_R(t_{n-1}, t_n)^{-1}$$

This decomposition of the option is again exact up to this point. In particular, discounting these cash flows with current discounting factors back to the current value date t reproduces the option price. The linear approximation comes into play when we *assume* that the *size* of the as yet determined cash flows is *independent* of the interest rates. Strictly speaking, Δ_F still depends on the interest rate: the forward rate appears in the expression for x' and $B_r(t,T)$ and the coefficient of the $\mathrm{N}(x')$ in the explicit expression for Δ_F is the spot rate discount factor over the lifetime of the option. Assuming delta to be constant in the linear approximation means, in particular, that it is independent of interest rate fluctuations. The size of the above cash flows can then likewise be *assumed* to be constant. The interest rate risk of the cash flow structure then arises solely in consequence of the changes in the *present value* (and not in the *size*) of the expected future cash flows, resulting from discounting with stochastic discount factors.

The Put

The approach for the corresponding bond future *put* is completely analogous: The option price p_F is set equal to the value of the replicating portfolio, composed of a cash amount $g(t)$ and delta futures with a delivery price K and current value F_S:

$$p_F(t,t,K) = \Delta_F(t)F_S(t,T',K) + g(t)$$

The required delta is obtained from Equation 13.16 by taking the derivative with respect to the spot price $S(t)$ of the underlying bond. The derivatives needed from Table 13.1 are now the "put on a future with respect to the spot" and "future with respect to the spot" (the notation is the same as in Section 13.4):

$$\begin{aligned}\Delta_F(t) &= -B_q(t,T')B_r(T,T')^{-1}\mathrm{N}(-x')\left[B_q(t,T')B_r(t,T')^{-1}\right]^{-1} \\ &= -\frac{B_r(t,T')}{B_r(T,T')}\mathrm{N}(-x') = -B_r(t,T)\mathrm{N}(-x')\end{aligned}$$

where we have again made use of Equation 3.4 in the final step. The cash sum $g(t)$ can now be expressed as the difference between the option price and the price of the delta future positions:

$$\begin{aligned}g(t) &= p_F(t,T,K) - \Delta_F(t)F_S(t,T',K) \\ &= -B_r(t,T)\left[S(t,T')\mathrm{N}(-x') - K\mathrm{N}(-x'+\sigma\sqrt{T-t})\right] \\ &+ B_r(t,T)\mathrm{N}(-x')\left[S(t,T') - K\right] \\ &= B_r(t,T)K\left[\mathrm{N}(-x'+\sigma\sqrt{T-t}) - \mathrm{N}(-x')\right]\end{aligned}$$

This cash sum is again *positive*. In the framework of a replicating portfolio, this amount is invested up to the maturity of the option. Compounding this sum increases the value of the investment to $K\left[\mathrm{N}(-x'+\sigma\sqrt{T-t}) - \mathrm{N}(-x')\right]$ at time T. This is the amount received at option's maturity and, as such, is an *incoming* cash flow. Together with the delta short futures positions, the cash flows of the replicating portfolio for the put on a bond future can be given as follows:

- at maturity T of the option: an *in*coming cash flow in the amount of the return on the investment of the cash amount $g(t)$

$$K\left[\mathrm{N}(-x'+\sigma\sqrt{T-t}) - \mathrm{N}(-x')\right]$$

- at maturity T' of the short future: an *in*coming cash flow in the amount of the (clean) delivery price K times the face value (here equal to 1) multiplied by the conversion factor P of the cheapest-to-deliver bond in addition to the accrued interest Z from the CTD; all of this for each of the delta futures

$$-\Delta_F(KP+Z) = +B_r(t,T)\mathrm{N}(-x')(KP+Z)$$

- for each coupon payment date t_i of the CTD after the maturity T' of the future, i.e. on each date t_i with $i = m+1,\ldots,n-1$ and $t_{m+1} > T'$: an *out*going cash flow in the amount of the CTD coupon yield on the face value for each of the Δ_F futures over the respective previous coupon period

$$\Delta_F[B_R(t_{i-1},t_i)^{-1} - 1] = -B_r(t,T)\mathrm{N}(-x')[B_R(t_{i-1},t_i)^{-1} - 1]$$

- at maturity of the underlying, i.e. on the redemption date t_n of the CTD: an *out*going cash flow in the amount of the last CTD coupon plus the CTD's face value for the Δ_F futures

$$-B_r(t,T)\mathrm{N}(-x')[B_R(t_{n-1},t_n)^{-1}$$

Caps and Floors

In Section 18.4.3 caps and floors were interpreted as a series of caplets and floorlets. Each caplet corresponds to a call and each floorlet to a put on the reference interest rate (a zero bond yield) for the forward period from fixing time T to payment time T'. These options are priced in the market with the Black-76 model as indicated in Equation 18.8 using the forward rate as the underlying forward price. Thus, a forward rate agreement for the corresponding forward period on the same principal N as the option and with respect to the same reference rate as the option is the ideal instrument for replicating a caplet. In Section 17.1 the price of such an FRA was shown to be

$$\begin{aligned} F_r(T,T',K|t) &= NB_r(t,T')\left[B_r(T,T')^{-1} - B_K(T,T')^{-1}\right] \\ &= NB_r(t,T')\left[(B_r(T,T')^{-1}-1) - (B_K(T,T')^{-1}-1)\right] \\ &= NB_r(t,T')\left[r(T,T'|t) - K\right](T'-T) \end{aligned}$$

where in the last step, we have replaced the general notation for the compounding factor over the forward period with simple compounding commonly used in these markets. The replicating portfolio composed of FRAs and cash amounts is now set equal to the caplet price

$$c_r^{\text{cap}}(T,T',K|t) = \Delta(t)F_r(T,T',K|t) + g(t)$$

The required delta is obtained from Equation 13.16 by taking the derivative with respect to the spot price $S(t)$ of the underlying (here the spot zero bond yield r). The derivative needed ("call on a forward with respect to spot" and "forward with respect to spot") can be taken directly from Table 13.1 (the notation is the same as that used in Section 13.4):

$$\Delta(t) = \mathrm{N}(x')$$

Now, $g(t)$ can be expressed as the difference between the option price and the price of delta FRA positions:

$$\begin{aligned} g(t) &= c_r^{\text{cap}}(T,T',K|t) - \Delta(t)F_r(T,T',K|t) \\ &= N(T'-T)B_r(t,T')\left[r(T,T'|t)\mathrm{N}(x') - K\mathrm{N}(x'-\sigma\sqrt{T-t})\right] \\ &- N(T'-T)B_r(t,T')\mathrm{N}(x')\left[r(T,T'|t) - K\right] \\ &= B_r(t,T')N(T'-T)K\left[\mathrm{N}(x') - \mathrm{N}(x'-\sigma\sqrt{T-t})\right] \end{aligned}$$

This sum is positive and is invested until T'. Through compounding until T', the investment $g(t)$ increases in value to $N(T'-T)K\left[\mathrm{N}(x') - \mathrm{N}(x'-\sigma\sqrt{T-t})\right]$. This is the amount received on the repayment date T' and as such is an *incoming* cash flow.

The cash flows generated by the FRA have been determined already in Section 24.1.2. We use them here in order to complete the cash flow structure of the option. Note that only "delta" FRAs are in the replicating portfolio. In total, a caplet is represented by the following cash flows generated by its replicating portfolio:

- at the beginning of the caplet period, i.e. on the fixing date T of the FRA: an *in*coming cash flow in the amount of Δ multiplied by the principal

$$\Delta N = \mathrm{N}(x')N$$

- at the conclusion of the caplet period, i.e. at the settlement date T: an *out*going cash flow from the FRA in the amount of the face values compounded over the caplet period at the caplet's strike rate (multiplied by Δ)

$$-\Delta N B_K(T,T')^{-1} = -\mathrm{N}(x')NB_K(T,T')^{-1}$$

  and an *in*coming cash flow in the amount of the compounded investment $g(t)$

$$NK\left[\mathrm{N}(x') - \mathrm{N}(x' - \sigma\sqrt{T-t})\right]$$

  Adopting the convention of simple compounding for the caplet period, the compounding factor at the strike rate can be written as $B_K(T,T')^{-1} = 1+K(T'-T)$. The cash flows due at T' can then be collected together into a net outflow in the amount of

$$-N\left[\mathrm{N}(x') + (T'-T)K\mathrm{N}(x' - \sigma\sqrt{T-t})\right]$$

We obtain the same result if the forward rate (simple compounding) in the option price Equation 18.8 is expressed using spot rate discount factors and the coefficients of those discount factors are identified as the cash flows: adopting the convention of simple compounding allows us to write $B_r(T,T')^{-1} - 1 = r(T,T'|t)(T'-T)$. Therefore the caplet price can be written in the form

$$\begin{aligned} c_r^{\mathrm{cap}}(T,T',K|t) &= NB_R(t,T')\left[B_r(T,T'|t)^{-1} - 1\right]\mathrm{N}(x') \\ &\quad - NB_R(t,T')K(T'-T)\mathrm{N}(x' - \sigma\sqrt{T-t}) \\ &= B_R(t,T)N\mathrm{N}(x') \\ &\quad - B_R(t,T')N\left[\mathrm{N}(x') + (T'-T)K\mathrm{N}(x' - \sigma\sqrt{T-t})\right] \end{aligned}$$

The cash flows at T and T' are the coefficients of the respective discount factors.

The cash flows for the floorlet can be determined analogously:

- at the beginning of the floorlet period, i.e. at the fixing date T: an *out*going cash flow from the FRA in the amount of

$$-\mathrm{N}(-x')N$$

- at the conclusion of the floorlet period, i.e. at the settlement date T': an *in*coming cash flow in the amount of

$$N\left[\mathrm{N}(-x') + (T'-T)K\mathrm{N}(-x' + \sigma\sqrt{T-t})\right]$$

Swaptions

As shown in Section 18.4.4 a swaption can be interpreted as a bond option. In this interpretation, the fixed side of the swap represents the bond with a coupon equal to the strike rate of the swaption and the (dirty) strike for the bond price is represented by the value of the floater, usually the swap's face value. The cash flows of bond options have already been specified above.

24.2 Interpolation and Cash Flow Mapping

24.2.1 Interpolation

Theoretically, knowledge of the spot rates for any arbitrary lifetime is needed to price financial instruments by discounting cash flows or more generally by means of the martingale property and discounted future expectations. A change in these rates results in a corresponding change in the present value of the instrument. The yield of every zero bond thus represents a risk factor. These represent, in principle, an infinite number of risk factors. These can, however, be estimated through the interpolation between a finite number of spot rates observed in the market. These observed spot rates are sometimes called *vertices.*

The times T_i, $i = 1, ..., n$ denote the maturities corresponding to these vertices. The spot rates for these maturities can thus be obtained from the market. The spot rate at time t for a maturity T between two vertices T_i and T_{i-1} is estimated by estimating at time t a function $R(t, T)$ which at the vertices takes on the observed spot rate values. This procedure is called *interpolation.* The interpolated spot rate for the maturity T is then given by the value of this function at T. There are several plausible ways of constructing such a function, for example, linear interpolation, splines, etc. The simplest possibility is to plot straight lines between neighboring points defined by the interest rates at the vertices. This is referred to as a *linear interpolation.* The spot rate with maturity T is then simply

$$R(t,T) = \frac{T_i - T}{T_i - T_{i-1}} R(t, T_{i-1}) + \frac{T - T_{i-1}}{T_i - T_{i-1}} R(t, T_i)$$
$$\text{with } t = T_0 < T_1 < \cdots < T_{i-1} \leq T < T_i < \cdots < T_n \tag{24.1}$$

For the sake of clarity, the full time dependence of the spot rates has been written here explicitly. The proportion with which the value of a vertex contributes to the interpolated value is thus one minus the relative distance to the respective vertex[1] (relative to the total distance between the vertices).

From this estimated interest rate $R(t, T)$ the discount factor $B_R(t, T)$ corresponding to a maturity T can be calculated for every compounding convention. Thus, the discount factor itself need not be estimated by means of an interpolation. However, we still have to estimate its volatility. The price volatility is by definition, the standard deviation of the logarithmic price change. The logarithmic price changes, however, are the same as yields and the yields of the discount factors are precisely the spot rates. In order to be consistent, the *same* interpolation must used for the price volatility (logarithmic changes) of the discount factors as was used for the spot rates. If the spot rates were estimated with a linear interpolation as in Equation 24.1, then the price

[1] For instance the contribution (i.e. the weight) of $R(t, T_{i-1})$ to the interpolated value $R(t, T)$ is

$$\frac{T_i - T}{T_i - T_{i-1}} = \frac{T_i - T - (T_i - T_{i-1})}{T_i - T_{i-1}} + 1 = 1 - \frac{T - T_{i-1}}{T_i - T_{i-1}}$$

Analogously for the other vertex.

volatilities of the zero bonds must also be linearly interpolated:

$$\sigma(t,T) = \frac{T_i - T}{T_i - T_{i-1}}\sigma(t,T_{i-1}) + \frac{T - T_{i-1}}{T_i - T_{i-1}}\sigma(t,T_i) \tag{24.2}$$

24.2.2 Risk-Based Cash Flow Mapping

Such (or similar) interpolations can be used to construct market parameters (risk factors) in any situation. This is sufficient for the pricing of financial instruments but not for their risk management. In risk management, only those maturities are vertices for which not only the spot rate itself but also its volatility and its correlations with the spot rates at all other vertices are available. Financial data providers generally make the volatilities and correlations for each interest rate curve (usually one or two for each important currency) available daily at the following vertices:

1m	3m	6m	12m	2y	3y	4y	5y	7y	9y	10y	15y	20y	30y

The calculation of the *risk* of a financial instrument whose (expected) cash flows do not occur precisely at these maturities cannot be performed by merely interpolating the spot rates and the volatilities for maturities between these vertices because additional information (correlations, for example) are needed which cannot be "interpolated". To overcome this problem in risk management, we take the opposite approach: Instead of interpolating to fit market parameters to a given instrument (cash flows), we fit the instrument to the given market parameters, i.e. to the given vertices. By means of *cash flow mapping,* the cash flows of the instrument are distributed over the available vertices, generating "artificial cash flows". A cash flow occurring between two vertices is divided into cash flows occurring at the previous and at the following vertex.

The cash flow $C(T)$ at time T is distributed over the two vertices T_{i-1} and T_i between which T occurs. The thus generated "artificial cash flows" (occurring on the dates T_{i-1} and T_i) are expressed in terms of their (percentage) share a and b, respectively, of the original cash flow:

$$C(T) \to \begin{cases} C(T_{i-1}) = a\,C(T) \\ C(T_i) = b\,C(T) \end{cases}$$
$$\text{where} \quad t = T_0 < T_1 < \cdots < T_{i-1} \le T < T_i < \cdots < T_n \tag{24.3}$$

The proportions a and b (two unknowns) can be determined uniquely through two conditions. There are quite a few available candidates for these conditions. At this juncture, we will discuss the two conditions considered to be the most reasonable choices in modern risk management.

- *The present value of the cash flow should remain unchanged under the mapping.* The sum of the two cash flows discounted with the factors $B_R(t,T_i)$ and $B_R(t,T_{i-1})$ should be equal to the original cash flow discounted at the interpolated spot rate $R(t,T)$. This invariance of the present value yields the following condition

$$\begin{aligned} B_R(t,T)C(T) &= B_R(t,T_{i-1})C(T_{i-1}) + B_R(t,T_i)C(T_i) \\ &= B_R(t,T_{i-1})\,a\,C(T) + B_R(t,T_i)\,b\,C(T) \end{aligned}$$

Thus:

$$B_R(t,T) = aB_R(t,T_{i-1}) + bB_R(t,T_i) \tag{24.4}$$

- *The risk should remain unchanged under the mapping.*
  The price risk of the sum of the distributed cash flows should be equal to the risk of the present value of the original cash flow. This condition gives the name *risk-based* cash flow mapping to the mapping introduced here.

The right hand side of Equation 24.4 can be thought of as a portfolio composed of two risk factors which are represented by the *discount factors* with maturities T_i and T_{i-1}. If the risk factors are directly the discount factors (and not the spot rates) then the portfolio is *linear* in the risk factors. It was shown in Section 22.2.2 that in such a situation the *risk* of a portfolio is proportional to the portfolio's *variance*. It was also shown in Section 22.2.2, that this is only possible when the stochastic evolution of the risk factors is approximated over a *finite* time interval δt in the sense of Equation 21.25. With this approximation, the variance of $B_R(t,T_i)$, for example, is simply equal to $B_R(t,T_i)^2\sigma_i^2\delta t$, where σ is the *price volatility* of $B_R(t,T_i)$. Analogous to Equation 22.9, the variance of this portfolio is then

$$\begin{aligned} B_R(t,T)^2\sigma_T^2\delta t = {} & a^2 B_R(t,T_{i-1})^2\sigma_{i-1}^2\delta t \\ & + b^2 B_R(t,T_i)^2\sigma_i^2\,\delta t \\ & + 2ab B_R(t,T_{i-1})B_R(t,T_i)\,\rho_{i,i-1}\,\sigma_{i-1}\,\sigma_i\delta t \end{aligned} \tag{24.5}$$

The volatilities of the discount factors at the vertices, σ_i and σ_{i-1}, are known values as is the correlation $\rho_{i,i-1}$ between the two discount factors. The volatility σ_T is obtained by an interpolation consistent with the interpolation used for the spot rates (for example, by the linear interpolation in Equation 24.2 if Equation 24.1 was used for the spot rates). Thus, only the two weights a and b in Equation 24.5 are unknown.

Equations 24.4 and 24.5 are the two conditions from which the proportionality factors a and b are determined. All variables in these two equations are known with the exception of a and b. Transforming the parameters through

$$\alpha = \frac{B_R(t,T_{i-1})}{B_R(t,T)}a\,, \quad \beta = \frac{B_R(t,T_i)}{B_R(t,T)}b \tag{24.6}$$

simplifies the two conditions substantially:

$$\begin{aligned} 1 &= \alpha + \beta \\ \sigma_T^2 &= \alpha^2\sigma_{i-1}^2 + \beta^2\sigma_i^2 + 2\alpha\beta\,\rho_{i,i-1}\,\sigma_i\,\sigma_{i-1} \end{aligned}$$

From the first of the two equation (the present value has to remain unchanged) it follows immediately that $\beta = 1-\alpha$. Substituting this into the second condition (the risk has to remain unchanged) yields a quadratic equation with only one unknown:

$$\sigma_T^2 = \alpha^2(\sigma_{i-1}^2 + \sigma_i^2 - 2\rho_{i,i-1}\,\sigma_i\,\sigma_{i-1}) + 2\alpha\,(\rho_{i,i-1}\,\sigma_i\,\sigma_{i-1} - \sigma_i^2) + \sigma_i^2$$

The solution of a quadratic equation of the form "$ax^2 + bx + c = 0$" can be found in any compilation of mathematical formulae[2]. The solution is given by

$$\alpha = \frac{\sigma_i^2 - \rho_{i,i-1}\,\sigma_i\,\sigma_{i-1} \pm \sqrt{\sigma_T^2(\sigma_i^2 + \sigma_{i-1}^2 - 2\rho_{i,i-1}\,\sigma_i\,\sigma_{i-1}) - \sigma_i^2\sigma_{i-1}^2(1-\rho_{i,i-1}^2)}}{\sigma_i^2 + \sigma_{i-1}^2 - 2\rho_{i,i-1}\,\sigma_i\,\sigma_{i-1}} \tag{24.7}$$

This value for α (and also $\beta = 1-\alpha$) can be substituted into Equation 24.6 to determine the desired proportionality factors a and b. The mapping in Equation 24.3 is then given explicitly by

$$C(T) \to \begin{cases} C(T_{i-1}) = \alpha \frac{B_R(t,T)}{B_R(t,T_{i-1})} C(T) \\ C(T_i) = (1-\alpha) \frac{B_R(t,T)}{B_R(t,T_i)} C(T) \end{cases}$$
$$\text{with} \quad t = T_0 < T_1 < \cdots < T_{i-1} \leq T < T_i < \cdots < T_n \tag{24.8}$$

Note that α and $\beta = 1 - \alpha$ in Equation 24.7 do *not* directly provide the necessary cash flow proportions at the vertices; these are obtained only after substitution into Equation 24.6 in which α and β were first defined.

At no point did the derivation of Equations 24.7 and 24.8 require knowledge of the interpolation method used to estimate the risk factor or its volatility at time T . The cash flow mapping holds for any arbitrary interpolation. Only when a concrete application is to be calculated is it necessary to decide on an interpolation method, since the interpolation is needed to determine $B_R(t,T)$ and σ_T which are then substituted into Equations 24.7 und 24.8

24.2.3 Duration-Based Cash Flow Mapping

Naturally, other conditions can be called upon to distribute the cash flows on the vertices. Instead of demanding that the present value and the risk remain unchanged, we could require that the present value and the *duration* be invariant under the mapping. This method was frequently used in the past, that is to say, before the age of modern risk management, and is presented here for the sake of completeness.

The condition that the present value be invariant under the mapping is the same as that given above and finds expression in Equation 24.4. The second condition of duration-based mapping is that the Macaulay duration given in Equation 6.8 remains the same. The Macaulay duration of a zero bond (for all forms of compounding) is simply the lifetime of the bond. The condition is therefore given by

$$T - t = \frac{(T_{i-1} - t)B_R(t,T_{i-1})a + (T_i - t)B_R(t,T_i)b}{B_R(t,T)} \tag{24.9}$$

where the rules of the addition of durations discussed in Section 6.4.3 have been observed. This, together with Equation 24.4, now gives the two conditions from which the proportionality parameters a and b are determined. The introduction of α and β

[2]See for instance [21].

as defined in Equation 24.6 reduce these conditions to

$$\begin{aligned} 1 &= \alpha + \beta \\ T - t &= (T_{i-1} - t)\alpha + (T_i - t)\beta \end{aligned}$$

From the first of the two conditions (the present value has to remain unchanged), it follows as before that $\beta = 1 - \alpha$. This, substituted into the second condition (the duration has to remain unchanged), yields an equation with only a single unknown. This equation can be easily solved to obtain

$$\alpha = \frac{T_i - T}{T_i - T_{i-1}} \Rightarrow \beta = 1 - \alpha = \frac{T - T_{i-1}}{T_i - T_{i-1}}$$

Thus, α and β are of the same form as would have been obtained from a linear interpolation, for example, as in Equation 24.1. Despite this fact, the cash flow can not be distributed linearly on the vertices since the relevant proportionality factors are actually a and b as defined in Equation 24.6. These must first be computed to yield the following correct duration-based cash flow mapping:

$$\begin{aligned} C(T) \rightarrow \begin{cases} C(T_{i-1}) = \frac{T_i - T}{T_i - T_{i-1}} \frac{B_R(t,T)}{B_R(t,T_{i-1})} C(T) \\ C(T_i) = \frac{T - T_{i-1}}{T_i - T_{i-1}} \frac{B_R(t,T)}{B_R(t,T_i)} C(T) \end{cases} \\ \text{where } \; t = T_0 < T_1 < \cdots < T_{i-1} \leq T < T_i < \cdots < T_n \end{aligned} \tag{24.10}$$

25

Example of a VaR Computation

In the Excel workbook ValueAtRisk.xls to be found on the CD accompanying this text, several of the value at risk concepts already introduced are applied to explicitly compute the VaR of a concrete portfolio within the delta-normal method. The example is quite dense in the sense that many of the concepts introduced above (as well as several concepts to be presented in later chapters, in particular in Sections 28.3.3 and 28.3.4) are collected in one calculation. However, it is by all means reasonable to present such a summary at this point as it will provide the reader with a complete reference containing all the essential steps for computing a value at risk (at least, the "simple" delta-normal version). We will proceed step by step through the example.

25.1 The Portfolio

The portfolio is presented in Figure 25.1. It is composed of

- A British zero bond (denoted by GBP.R180) with a lifetime of 6 months and a face value of 100,000 British pounds.

- A Japanese zero bond (denoted by JPY.Z06) with a lifetime of 6 years and a face value of 2,000,000 Japanese yen.

- Put options on Japanese 7-year zero bonds (denoted by JPY.Z07) with a face value of 6,000,000 Japanese yen. These puts have a lifetime of 6 months and a strike of 94% (the price of the bonds are also given in percentage points; a value of 100% means that the present value of the bond is equal to its face value). The Black-Scholes price (at a risk-free rate of 3%) of these puts is 0.389% (of the face value), i.e. 23,364 JPY. The delta is -0.2803.

Financial Instrument	Currency	Market Price	Cross Rate to EUR	Principal in Currency	Market Value in EUR
GBP.R180	GBP	97.70%	1.61720 €	£10,000.00	15,800.68 €
JPY.Z06	JPY	96.01%	0.00926 €	2,000,000 JPY	17,777.28 €
Put on JPY.Z07	JPY	0.389%	0.00926 €	6,000,000 JPY	216.30 €
Delta of Put		-0.2803		Sum	33,794.27 €

Figure 25.1: The portfolio.

Risk Factor	Currency	Price		Yield	
		Market Price	Daily Price Vol	Annual Yield	Daily Yield Vol
USD/EUR	USD	$0.9083	0.680%		
USD/GBP	USD	$1.4689	0.404%		
USD/JPY	USD	$0.0084	0.731%		
GBP.R180	GBP	97.70%	0.026%	4.70%	1.10%
JPY.Z05	JPY	97.74%	0.093%	0.46%	4.01%
JPY.Z07	JPY	93.89%	0.177%	0.90%	2.99%
JPY.Z06	JPY	***96.01%***	***0.135%***	***0.68%***	***3.33%***

Figure 25.2: Market data of the portfolio's risk factors given in their original currency.

25.2 Market Data

Prices, daily volatilities (both the price as well as the *yield* volatilities; see Section 28.3.3) and correlations are available for GPG.R180, JPY.Z07 and JPY.Z05 and are presented in Figure 25.2. These three variables are the portfolio's risk factors.

Data for a 6-year zero bond in Japanese yen (JPY.Z06) is not available initially. Its yield and its price volatility were computed from the vertex data for 5 and 7 years by means of a linear interpolation as indicated in Equations 24.1 and 24.2. The price of this zero bond was then calculated from its yield in accordance with Table 3.4 for discrete annual compounding and the yield volatility was determined using Equation 28.18 together with the price volatility. This data is written in bold-faced italics in Figure 25.2 to emphasize the fact that the data for this 6-year bond are not original market values, but have been obtained through interpolation and other calculations.

Naturally, the FX rate of the British pound and the Japanese yen with respect to the reference currency in which the value at risk is to be expressed, are also risk factors for the portfolio (these are also listed in Figure 25.2). The *reference currency* here is the euro (EUR). As will be shown extensively in Section 28.3.4, the *FX risk* is already taken into account by the fact that the given volatilities and correlations with respect to their respective original currencies (here the GBP and JPY) must be transformed in terms of the reference currency. These transformations are determined from the volatilities and correlations of the FX rates concerned. In this way, the FX

Volatilities and Correlations of Risk Factors						
in Original Currency				in EUR		
JPY.Z05	JPY.Z07	GBP.R180		JPY.Z05	JPY.Z07	GBP.R180
0.093%	0.89410	0.17171	JPY.Z05	0.732%	0.76434	0.33993
0.89410	0.177%	0.00987	JPY.Z07	0.76434	0.744%	0.45638
0.17171	0.00987	0.026%	GBP.R180	0.33993	0.45638	0.454%

Figure 25.3: Volatilities and correlations of the risk factors.

Risk-Based Cash Flow Mapping			
Root	1.175E-06	a	0.4539195
α_1	0.4621096	b	0.5500128
α_2	2.69074772	α_2 irrelevant, as > 1	

Figure 25.4: Calculation of the weights needed for the cash flow mapping.

risk is incorporated into the calculation. If these FX rates are not available in the reference currency, but in yet *another* currency (in our example, US dollar), this new currency (with *its* volatilities and correlations) also comes into play. Therefore, the USD/EUR exchange rate appears in Figure 25.2 as well. The transformations of the volatilities and correlations into the reference currency is quite complicated and will receive detailed treatment in Section 28.3.4. The result of all these transformations is presented in Figure 25.3. The volatilities (in the diagonal) and the correlations of the three portfolio risk factors with respect to the *original* currency (as provided by the data provider, for instance) are listed on the left side in Figure 25.3. On the right side, the transformed parameters (in accordance with Section 28.3.4) are displayed in the form needed for the value at risk computation with respect to the reference currency.

25.3 Cash Flow Mapping

Now that all required market data is at our disposal, we can proceed with the actual computation of the value at risk. We begin by distributing the face value of the 6-year zero bond among the JPY.Z07 and JPY.Z05 vertices using the risk-based cash flow mapping given by Equations 24.7 and 24.8. The volatilities and correlations of JPY.Z07 and JPY.Z05 in the *original* currency are used since the mapping occurs entirely within the framework of the original currency of the bonds under consideration (JPY). The cash flow mapping is given in Figure 25.4. This means that 45.4 % of the face value of the 6-year bond is transferred to the 5-year vertex and 55.0% to the 7-year vertex. The weights do not add up to precisely 100% because $\alpha + \beta = 1$, not $a + b$ (see Equation 24.6). The portfolio after the cash flow mapping is presented in Figure 25.5.

Financial Instrument	Currency	Market Price	Cross Rate to EUR	Principal in Currency	Market Value in EUR
GBP.R180	GBP	97.70%	1.61720 €	£10,000.00	15,800.68 €
JPY.Z05	JPY	97.74%	0.00926 €	907,839 JPY	8,215.05 €
JPY.Z07	JPY	93.89%	0.00926 €	1,100,026 JPY	9,562.23 €
Put on JPY.Z07	JPY	0.389%	0.00926 €	6,000,000 JPY	216.30 €
Delta of Put		-0.2803		Sum	33,794.27 €

Figure 25.5: The portfolio after the cash flow mapping.

Value at Risk			
Days 10		Confidence 2.326	
Risk Factor	VaR of Each Position in EUR	Portfolio Delta in EUR with Respect to Each Risk Factor	Portfolio VaR in EUR with Respect to Each Risk Factor
GBP.R180	527.94 €	16,171.97 €	527.94 €
JPY.Z05	442.55 €	8,404.73 €	442.55 €
JPY.Z07	523.51 €	- 5,383.26 €	276.73 €
Put on JPY.Z07	800.23 €	Sum	1,247.21 €
Sum	2,294.23 €	Total VaR	624.05 €

Figure 25.6: Three ways for calculating the value at risk.

25.4 Calculation of Risk

The risk of this portfolio in EUR is now calculated with the transformed volatilities and correlations found in Figure 25.3. In order to separately illustrate the *hedging effects* (the position in the 7-year bond partially hedges the put on the 7-year bond), the *correlation effects* and the *diversification effects*, we calculate three different risk values in Figure 25.6.

Firstly, the value at risk is calculated for each separate *position* with Equation 22.8. If we were to be extremely conservative and not consider either hedging nor correlation nor diversification effects, we could simply add the three values at risk of the separate positions and arrive at a total (very conservative) risk of 2,294.23 euros.

Alternatively, we could calculate the value at risk taking into account the hedge effects. The value at risk of the entire portfolio with respect to each separate *risk factor* is determined, likewise with Equation 22.8 as above. To do so, we first need to calculate the portfolio deltas with respect to the three risk factors. For the bond positions, this is equal to the face value (because the discount factors are treated directly as risk factors) and for the bond option, it is the option delta multiplied by the face value of the option. If different positions depend on the same risk factor, their deltas with respect to that risk factor can simply be added to yield the total delta with respect to that risk factor.

The hedge effect of the put is in this way taken into consideration. The delta of the put on the 7-year bond is negative and substantially reduces the portfolio sensitivity with respect to the 7-year bond[1]. If we wish to include this *hedge effect* of the put in the risk computation while still neglecting the correlation (respectively, the diversification) effects, we would now add the portfolio VaRs with respect to each risk factor to arrive at a total (and still conservative) risk of 1.247,21 euros. The hedge effect of the put position is considerable. The put reduces the value at risk of the portfolio by 1.047 euros to almost half of the original conservative figure calculated when neglecting the hedge effect.

Finally, the value at risk is computed taking both hedge and correlation effects into account, in accordance with Equations 22.10 and 22.11 using the correlations with respect to the reference currency EUR listed in Figure 25.3. The correlation (respectively diversification) effects reduce the value at risk by an additional 623.17 euros to 624.05 euros[2]. Assuming that all model assumptions and approximations made are justified, we can now be 99% confident that the portfolio will lose no more that 624.05 euros in value over the next ten days.

[1] In fact it more than compensates the positive delta of the 7-year bond position since the resulting portfolio delta with respect to the risk factor JPY.Z07 is negative. This is called *over hedging*. However, this is not as bad as the name suggests. Since JPY.Z07 is strongly correlated to JPY.Z05, this negative portfolio delta significantly reduces the *total* risk of the portfolio, when correlation effects are taken into account.

[2] This of course is largely due to the fact that the put "hedges" not only the JPY.Z07 but also the JPY.Z05-risk (i.e. the whole original 6-year bond). This effect can of course only be seen if correlations are taken into account.

26

Backtesting: Checking the Applied Methods

A comparison of the value at risk figures delivered by a risk management system with the actual value changes of a portfolio allows an estimation of the qualitative and quantitative "goodness" of the risk model. Comparisons of realized values with previously calculated values are called *backtesting* procedures.

26.1 Profit-Loss Computations

There are several different *profit-loss* (or *P&L*) methods which can be used for comparison with the value at risk. The differences in these methods reflect the differences in the fundamental "philosophy" behind them.

- The *Dirty* Profit Loss: The actual P&L of the portfolio, including all changes in position, fees paid and received, commission, etc. over the value at risk period are compiled and compared with the value at risk previously calculated. Position changes arise from continued trading during the value at risk period, the maturing of positions in the portfolio (for example, futures and options), the knock-in or knock-out of barrier options, coupon payments of bonds, etc. The effect of continued trading is not, in general, contained in the value at risk model. The dirty P&L is therefore only suitable for evaluating *trading* performance and not for the evaluation of *model* performance.

- The *Cleaned* Profit Loss: The cleaned P&L is calculated in the same way as the dirty P&L but *without* taking position changes into account which result from continued trading during the value at risk period. Furthermore, the payment and receipt of fees and commissions are also omitted from the calculation. However, the cleaned P&L still contains the position changes resulting from the maturity of instruments occurring during the value at risk period (such as options and futures) or other position effects caused by the market (as opposed to the trader)

such as the knock-out or knock-in of barrier options, coupon payments of bonds, etc. The cleaned P&L is therefore suitable to evaluate the model performance of risk models which take account of such maturity effects. The Monte Carlo simulation, for example, allows for such effects, the Variance-Covariance method, on the other hand, does not.

- The *Clean* Profit Loss: Finally, the clean P&L is calculated in the same ways as the cleaned P&L but with reversing the effects of the maturing of positions during the value at risk period. In calculating the clean P&L, the value of the exact same portfolio as that existing upon initial calculation of the value at risk is re-calculated with the new market data observed at the conclusion of the value at risk period. Of course, a record of the initial portfolio positions at the time of the value at risk computation must have been kept. The clean P&L is thus suitable for evaluating the model performance of risk models such as the Variance-Covariance method which do not account for *aging* effects of the positions.

Whichever profit-loss method is chosen, the profit-loss per backtesting period is recorded for the evaluation of the goodness of the value at risk. Additionally, a record is kept for the calculated value at risk of the portfolio per backtesting-period. The data required for backtesting is thus not very large: neither historical time series nor a history of the portfolio positions must be maintained. Only two values per portfolio must be recorded in order to save the history of the portfolio, namely the value at risk and the associated P&L of the portfolio to be compared with the value at risk. In most cases, the P&L will be more favorable than the value at risk, and in others less favorable. Counting the number of times that the P&L is less favorable than the value at risk enables statistical conclusions about the goodness of the utilized model to be drawn. This is the fundamental idea behind backtesting. The backtesting procedure used by the supervising authorities when auditing a bank's internal risk management models in the framework of the Capital Adequacy Directive (CAD II) is presented in the following section.

26.2 The Traffic Light Approach of the Supervising Authorities

26.2.1 Adjusting the Value at Risk (Yellow zone)

The value at risk is a statistical statement. In general, some of the changes in the portfolio's values will be less favorable than the calculated value at risk. Such changes are referred to as *outliers* in the following discussion. At a confidence level c the probability of such an outlier is $1-c$. The *expected* number of outliers in n backtesting periods with respect to this level of confidence is thus

$$\mathrm{E}\,[k] = n(1-c) \tag{26.1}$$

where k denotes the number of outliers observed in the n backtesting periods. The value k is not always equal to its expectation; it is a random variable. A deviation of

the observed k from the expected number of outliers does not necessarily imply that a model is incorrect. Such an observation may be the result of pure chance. This is particularly true for small deviations from the expected value. In such cases, one speaks of the *yellow zone*, in which the supervising authorities will accept the model, but require that the value at risk is increased in the following manner.

The realization of k outliers actually observed in backtesting allows the definition of a new confidence level c' with respect to which the observed number is equal to the expectation:

$$k = n(1 - c') \quad \Rightarrow \quad c' = 1 - k/n \tag{26.2}$$

From the observed outliers, it can be concluded that the value at risk from the model does not correspond to the confidence level c, but to a confidence level c'; or at least that the experimental basis for a confidence c' is greater than for c. The given VaR is then interpreted with respect to a confidence level of c'. A VaR which better corresponds to the *claimed* confidence c is then obtained from the given value of the VaR through multiplication by the ratio of the percentiles (confidence interval bounds) Q_{1-c} and $Q_{1-c'}$ in accordance with Equation 21.19, respectively Equation 22.10:

$$\mathrm{VaR}(c,t,T) \approx \frac{Q_{1-c}}{Q_{1-c'}} \ \mathrm{VaR}(\underbrace{1 - \frac{k}{n}}_{c'}, t, T) \tag{26.3}$$

In order to use this equation, all assumptions and approximations of the delta-normal method must be made. These are:

- The risk factors are random walks, i.e. they are lognormally distributed implying that $Q_{1-c} = Q_{1-c}^{\mathrm{N}(0,1)}$.
- The drifts of the risk factors are all neglected in the calculation.
- The exponential time evolutions of the risk factors are linearly approximated.
- The dependence of the portfolio value on the risk factors is linearly approximated. In particular, the portfolio value is also assumed to be lognormally distributed.

Since, for logarithmic changes, all of the variables under consideration are assumed to be normally distributed, the $Q_{1-c'}$ percentile can be calculated as

$$c' = 1 - \frac{k}{n} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_{1-c'}} e^{-x^2/2} dx$$

For example, if backtesting over 250 periods is performed resulting in the observation of 6 logarithmic portfolio value changes outside of the claimed confidence interval at a

confidence level of 99%, the percentiles Q_{1-c} and $Q_{1-c'}$ are given by

$$0.99 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_{1-c}} e^{-x^2/2} dx \Rightarrow Q_{1-c} \approx 2.326$$

$$1 - \frac{6}{250} = 0.976 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_{1-c'}} e^{-x^2/2} dx \Rightarrow Q_{1-c'} \approx 1.972$$

The confidence level of the value at risk calculated by the model is now assumed to be not 99% as claimed but rather 97.6% as calculated on the basis of the actual events. The value at risk must thus be adjusted by a factor of $Q_{1-c}/Q_{1-c'} = 2.326/1.972 = 1.18$. This now larger value at risk is the value which can, based on actual observations, be relied upon with a confidence of 99%.

26.2.2 Criteria for Rejecting a Model (Red Zone)

Adjusting the value at risk as described above may not be applied for *arbitrary* values of k. It may be only then applied if the difference between the observed value of k and its expectation with resect to the claimed confidence c can be reasonably explained by fluctuations due to the randomness involved. If the number of outliers is too far removed from the expected value, chance is no longer a plausible explanation and the reasons for the deviation lie in all probability on fundamental errors in the model itself. Experts say that the deviation of the measured results from the expected results are *significant*. The supervising authorities say that the model is in the *red zone*.

The field of statistics provides *hypothesis tests* which serve to check whether or not the observed deviation from a claimed value can be plausibly explained by random fluctuations. If not, such a deviation is considered significant and the tested *hypothesis* is in all probability not true. But again, absolute statements cannot be made on the basis of statistics. From such a hypothesis test we can only conclude *with a certain probability* that it was correct to reject (or accept) the hypothesis. The possibility remains that the hypothesis is rejected (accepted) although it is true (false). In statistics, these kind of errors are referred to as *type-I error* (rejection of a correct hypothesis) and *type-II error* (acceptance of a wrong hypothesis).

The hypothesis made when backtesting an internal model is that the observed value changes of a portfolio lie within the confidence interval specified by the calculated value at risk with a probability of c, in other words, that with a probability c the observed value changes are more favorable than the computed value at risk. To check this claim, we test whether the observed portfolio's value changes actually lie within the respective confidence interval. This is done for every VaR period over the entire backtesting time span.

The observed results can be categorized into two possible outcomes: the change in the portfolio's value is either more favorable than the respective VaR or not. This corresponds to the *Bernoulli experiment* described in Section A.4.2. We could associate the event that an actual portfolio change is *more favorable* (or equally favorable) than the VaR with the outcome "tails" when tossing a coin and likewise the event that

the portfolio change is *less favorable* than the VaR to the outcome "heads". The probability of observing k "heads" in n trials (less favorable than the VaR) is binomially distributed (see Equation A.40) where the binomial probability p is the probability of the outcome "heads". In our case here p is then equal to the claimed probability that the portfolio change lies outside of the confidence interval, i.e. $p = 1 - c$. The number of outliers should therefore be binomially distributed with a density

$$B_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \quad p = 1 - c = 1 - \mathrm{N}(Q_{1-c}) \tag{26.4}$$

where $\mathrm{N}(Q_{1-c})$ denotes the probability that a standard normally distributed variable is $\leq 1 - c$, see Equation A.52. Assuming that the model is correct, $B_{n,p}(k)$ is the probability that precisely k outliers are observed. The probability that *at most* k outliers are observed is then

$$\sum_{i=0}^{k} B_{n,p}(i) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \tag{26.5}$$

Again, assuming that the model is correct, the probability of observing *more than* k outliers is then

$$1 - \sum_{i=0}^{k} B_{n,p}(i) = \sum_{i=k+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{26.6}$$

This is equal to the probability that the model is *correct* assuming that k or more outliers are observed. Therefore, this is the probability of making a type-I error (the rejection of a correct model) when the hypothesis is rejected if k or more outliers are observed.

Determining a type-II error (acceptance of a *false* hypothesis) requires that the *true* probabilities for outcomes of the *false* model are known; a situation which almost never happens in practice. To be more specific: If a hypothesis test accepts a model for up to k outliers, then the probability that a *false* model is accepted equals the *true* probability for the event that this *false* model produces k or fewer outliers. Let us consider a simple example: Assume that a false model claims a 99% confidence for its calculated VaR numbers while the *true* confidence for these VaR numbers is only 95%. If we accept that model's VaR-numbers as a 99% VaR as long as only up to 9 out of 250 backtesting periods produce an outlier, we make a type-II error (see Equation 26.5 with $n = 250$, $p = 5\%$) with a probability of

$$\sum_{i=0}^{9} B_{n,p}(i) = \sum_{i=0}^{9} \binom{250}{i} 0.05^i \times (0.95)^{250-i} \approx 19.46\%$$

It should be clear from these considerations that type-II errors only play a minor role in practice since it it very rarely the case that they can be determined in a sensible way.

The supervising authorities make their decision on establishing the limits for the red zone based on the probability of a type-I error (rejection of a correct model). The

model is said to be in the red zone for a number of outliers k if the rejection of the model for *more than* k outliers has a probability for a type-I error of less than 0.01%. Using Equations 26.6 and 26.5, we find this to be the case when the probability (calculated with the model under consideration) of *at most* k outliers is $\geq$ 99.99% under the assumption that the model under consideration is correct[1]. For $n = 250$ backtesting periods, the probability of at most 9 outliers equals 99.975% (see the accompanying Excel Sheet BinomialBacktest.xls). The probability of at most 10 outliers is equal to 99.995% and is thus larger than 99.99%. The red zone established by the supervising authorities for 250 backtesting periods therefore begins at 10 outliers although the probability of making a type-I error (which is the probability for *more than 9* outliers) is 0.025%, i.e. greater than 0.01%. The probability of a type-I error when deciding to reject the model if more than $k = 10$ (in other words, 11 or more) outliers are observed is 0.005%, smaller than the required 0.01%. Nonetheless, 10 outliers within 250 backtesting periods is already deemed to belong to the red zone.

26.2.3 The Green Zone

The boundary of the red zone is the upper boundary of the yellow zone. Analogously, a lower boundary of the yellow zone has been defined. No add-on is required if the observed number of outliers lies below this boundary. This zone is called the *green zone*. The model is said to be in the yellow zone for a number of outliers k if the rejection of the model for *more than* k outliers has a probability for a type-I error of less than 5.00%. Using Equations 26.6 and 26.5, we find this to be the case when the probability of *at most* k outliers is $\geq$ 95% under the assumption that the model is correct. For $n = 250$ backtesting periods for example, the probability of at most 4 outliers is equal to 89.22% (see the accompanying Excel workbook BinomialBacktest.xls). The probability of at most 5 outliers on the other hand is 95.88%. Thus, the supervising authorities establish the boundary for the yellow zone as $k = 5$ for 250 backtesting periods, although the probability of a type-I error in this case is 10.78%, i.e. greater than 5%. For the rejection of the model with *more* than 5 outliers, the probability of a type-I error is 4.12%.

These three zones established by the supervising authorities, motivate the name *traffic light approach.*

26.2.4 Multiplication Factor and Add-On

As a rule, the value at risk calculated with the model must be multiplied by a factor of three even when it is found to be in the green zone for the simple reason that it has not been computed in accordance with the standard methods but by means of an internal model. In the yellow zone, the VaR must *additionally* be multiplied by the ratio of the two percentiles Q_{1-c} and $Q_{1-c'}$ as prescribed in Equation 26.3 where $Q_{1-c'}$ is the value established from Equation 26.2. The multiplication factor for the yellow zone is thus $3Q_{1-c}/Q_{1-c'}$. The amount by which the multiplication factor exceeds the factor

[1]It may seem inconsistent that the supervising authorities establish k as the boundary for the red zone although this k corresponds to a type-I error of greater than 0.01%. This is the rule, however.

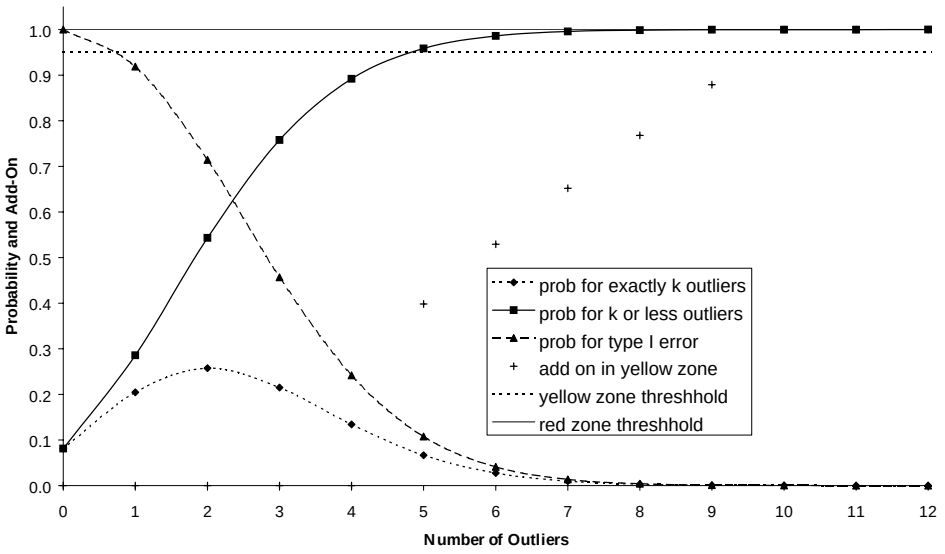


Figure 26.1: Value at risk backtesting by means of a binomial test for 250 backtesting periods and 99% VaR confidence.

3 is referred to as the *add-on*

$$\text{Add On} = 3\frac{Q_{1-c}}{Q_{1-c'}} - 3 \approx \frac{6,978}{Q_{1-c'}} - 3$$

where in the last step the confidence level $c = 99\%$ (required in Capital Adequacy Directive) for the standard normal distribution, $\mathrm{N}(Q_{1-c}) = 1\%$, and consequently $Q_{1-c} \approx 2.326$ was used.

The concepts described above are illustrated in detail in the Excel workbook BinomialBacktest.xls. All of the probabilities mentioned above, the boundaries between the different zones and the add-ons in the yellow zone are computed. The number of backtesting periods as well as the required confidences and the probability thresholds for the boundaries between the zones can be modified and the subsequent effects of these modifications immediately computed. In Figure 26.1, these values are presented for $n = 250$ backtesting periods and a confidence of $c = 99\%$.

In Figure 26.2, the add-ons are again explicitly displayed for the situation in Figure 26.1 (rounded in increments of 0.05). This table can be found in the Capital Adequacy Directive (CAD II). For the red zone, an add-on of one is set and the model is later subjected to a new test.

Despite the multiplication factor, the value at risk computed with internal models is often lower than that calculated in accordance with the standard methods, since the correlation and compensation effects are more accurately taken into account. In many

k	Add-On	Zone
<= 4	0.00	green
<= 5	0.40	yellow
<= 6	0.50	yellow
<= 7	0.65	yellow
<= 8	0.75	yellow
<= 9	0.85	yellow
>=10	1.00	red

Figure 26.2: The table of add-ons for 250 backtesting periods as shown in the text of the German Principle I (corresponding to CAD II).

cases, the VaR of a portfolio computed according to an internal model is, despite the multiplication factor, only half as large as that computed with the standard methods.

Part V

Market Data

27

Interest Rate Term Structures

Interest rate curves are constructed from the prices of bonds traded in the market. In order to construct the *spot rate curve* (also called *spot rate term structure* for *term structure*, for short), for example, the yields of zero bonds for *all* possible maturities are required. Observing an entire array of conditions in order to be consistent with the assumption of an arbitrage-free market, such an interest rate curve can be determined from market data characterizing traded interest rate instruments. Such market data are for instance prices of traded bonds (not necessarily zero bonds), spot rates, par rates, swap rates, etc. All of these variables can be traced back to a single common nucleus. If the market is arbitrage-free (and assuming the same credit-worthiness for all cash flows involved) then for every value date t and maturity date T there exists a *unique* discount factor $B_R(t,T)$. If all possible discount factors are known, then the present value of every instrument or portfolio consisting of cash flows can be determined. Thus, the spot rate curve $R(t,T)$ (which is nothing other than the yields of the discount factors $B_R(t,T)$) has to be constructed in such a way that the observed market prices of interest rate instruments can be reproduced by discounting the future cash flows of the instruments using $B_R(t,T)$.

If a sufficient number of liquid zero bonds were traded in the markets, we would be able to obtain the term structure directly from their market prices. However, there are not enough zero bonds in the market. Instead there are all kinds of other liquid interest rate instruments which not only have different cash flow structures (such as coupon payments) but also are quoted with respect to different compounding and day count conventions and may be subject to different credit risks (different *credit worthiness*). The first step in pricing interest rate products is to construct a sufficient number of discount factors from the quoted prices or yields (of coupon bonds, for example) with identical credit worthiness in order to generate an arbitrage free term structure for use in the valuation.

For example, government bonds (usually coupon bonds) and swaps are usually very liquid instruments and the prices of such instruments are available in the capital

markets. With these prices par rates and swap rates can easily be determined. Swaps are even quoted directly in terms of swap rates[1].

When constructing spot rate curves we are therefore usually posed with the inverse problem of what was presented in Section 6 or Section 16.4.2, namely determining (unknown) spot rates from given par rates or swap rates. This is usually performed with the *bootstrapping* method. In the following discussion we introduce a general bootstrapping procedure which (under certain circumstances) does not require a recursion relation before we present the more traditional bootstrap method.

27.1 Bootstrapping

Bootstrapping methods usually rely on quite restrictive assumptions. One or several of the following assumption are commonly made:

1. The value date t, for which a spot rate is to be computed from the par rate corresponds exactly to the start of a complete coupon period of the given bond, i.e. in the notation of chapter 16, $t = t_m$.

2. All coupon periods of this bond are of equal length.

3. A specified compounding convention, usually discrete annualized compounding or simple compounding, is adopted for the coupon payments.

27.1.1 The General Bootstrapping Equation

In the method presented here, *none* of these assumptions are made. The procedure can be applied to bonds with arbitrary coupon periods and arbitrary value dates t and is valid for all compounding conventions. This is a significant generalization through which the application of this method is substantially simplified as more bonds can be used in the bootstrapping[2].

The equation establishing the relationship between spot and par rates is of course still given by the at-par condition, Equation 6.6. For coupon bonds, this has the explicit form 16.10. This equation must now be solved for the *spot* rates rather than

[1] However, in general government bonds and swaps have different credit risks. So we can generally not use both government bonds and swaps within the same yield curve.

[2] If only one of the above three assumptions is made, namely that the coupon periods are of equal length, a closed form expression for the computation of the spot rates from the par rates can be obtained (see Equation 27.3 below).

the par rates. The at-par condition 16.10 can be written as

$$\begin{aligned} B_K(t_m,t)^{-1} &= \sum_{i=m+1}^{n} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] + B_R(t,t_n) \\ &= \sum_{i=m+1}^{n-1} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] \\ &+ B_R(t,t_n)\left[B_K(t_{n-1},t_n)^{-1}-1\right] + B_R(t,T) \\ &= \sum_{i=m+1}^{n-1} B_R(t,t_i)\left[B_K(t_{i-1},t_i)^{-1}-1\right] + B_R(t,t_n)B_K(t_{n-1},t_n)^{-1} \end{aligned}$$

where $B_R(t,T) = B_R(t,t_n)$. Solving the equation for the zero bond with the longest term t_n yields

$$\underbrace{B_R(t,t_n)}_{\text{Zero Bond}} = B_{K_n}(t_{n-1},t_n)\left(B_{K_n}(t_m,t)^{-1} - \sum_{i=m+1}^{n-1}\underbrace{B_R(t,t_i)}_{\text{Zero Bond}}\left[B_{K_n}(t_{i-1},t_i)^{-1}-1\right]\right)$$
$$\text{for } \; t_m \le t < t_{m+1} < t_{m+2} < \; \cdots \; < t_n = T \;\; , \;\; n \ge m+1 \text{ and } n \ge m+1 \quad (27.1)$$

observing the usual convention that the sum equals zero if its upper limit is less than the lower limit.

In order to emphasize the fact that the par rate K of an instrument depends on the maturity $T = t_n$ without unnecessarily complicating the notation, the par rate has been given the same index as that corresponding to the maturity date, i.e. $K_n := K(t,t_n)$. This K_n is thus the par rate valid at value date t for maturity date t_n.

The zero bond maturing at time t_n is thus obtained from the par rates and recursively from the zero bonds maturing at the coupon payment dates between t and t_n. Thus, the zero bond price for all coupon payment dates prior to t_n must already be known. This is the *bootstrapping* method in its most general form. The equation holds for arbitrary periods (the length of the coupon periods may vary as well), arbitrary value dates t and arbitrary compounding conventions, and even if different compounding methods are used for the par rates and the spot rates.

27.1.2 Bootstrapping for Coupon Periods of Equal Length

If we make only one of the three assumptions listed at the beginning of Section 27.1, i.e. Assumption 2, then we can transform the bootstrapping recursion 27.1 into a closed form expression. For equally long coupon periods, the interest payments from the coupon of a bond over all its periods are equal. This fact can be expressed in the following way

$$B_K(t_{i-1},t_i)^{-1} - 1 = B_K(t_m,t_{m+1})^{-1} - 1 \quad \forall i$$

From which it follows that the term in the square brackets in Equation 27.1 no longer depends on i and can thus be factored out of the sum. Taking advantage of this (and of

Equation 3.4), the bootstrapping equation for coupon periods of equal length reduces to

$$B_R(t,t_n) = B_{K_n}(t_m,t_{m+1}) \left(B_{K_n}(t_m,t)^{-1} - \left[B_{K_n}(t_m,t_{m+1})^{-1} - 1\right] \sum_{i=m+1}^{n-1} B_R(t,t_i) \right)$$
$$= B_{K_n}(t,t_{m+1}) - [1 - B_{K_n}(t_m,t_{m+1}] \sum_{i=m+1}^{n-1} B_R(t,t_i) \tag{27.2}$$

This recursion can be carried out explicitly. We denote the sum over the zero bonds by Σ_n in the following derivation. This sum is non-zero for all n such that $n-1 \geq m+1$:

$$\Sigma_n = \sum_{i=m+1}^{n-1} B_R(t,t_i) \quad \text{for} \quad n > m+1, \; \Sigma_n = 0 \quad \text{for} \quad n \leq m+1$$

Since the same time periods repeatedly appear in the argument of the coupon discount factors, their inclusion in the notation provides no additional information. We thus replace the expressions for the coupon discount factors by

$$B_{K_n} := B_{K_n}(t,t_{m+1}) \quad , \quad B^V_{K_n} := B_{K_n}(t_m,t_{m+1})$$

for a partial or a complete coupon period, respectively. Substituting this notation into Equation 27.2 allows us to simply write

$$B_R(t,t_n) = B_{K_n} - \left(1 - B^V_{K_n}\right) \Sigma_n$$

Separating the sum Σ_n into the sum over the first $n-m-1$ terms plus the last term and exploiting the bootstrapping equation yields a recursion relation for Σ:

$$\Sigma_n = B_R(t,t_{n-1}) + \underbrace{\sum_{i=m+1}^{n-2} B_R(t,t_i)}_{\Sigma_{n-1}}$$
$$= \underbrace{B_{K_{n-1}} - (1 - B^V_{K_{n-1}})\Sigma_{n-1}}_{B_R(t,t_{n-1})} + \Sigma_{n-1}$$
$$= B_{K_{n-1}} + B^V_{K_{n-1}} \Sigma_{n-1}$$

Repeated application of this recursion relation for Σ produces

$$\begin{aligned}
\Sigma_n &= B_{K_{n-1}} + B^V_{K_{n-1}} \Sigma_{n-1} \\
&= B_{K_{n-1}} + B^V_{K_{n-1}} (B_{K_{n-2}} + B^V_{K_{n-2}} \Sigma_{n-2}) \\
&= B_{K_{n-1}} + B^V_{K_{n-1}} B_{K_{n-2}} + B^V_{K_{n-1}} B^V_{K_{n-2}} (B_{K_{n-3}} + B^V_{K_{n-3}} \Sigma_{n-3}) \\
&= \cdots \\
&= \sum_{j=1}^{k} B_{K_{n-j}} \prod_{i=1}^{j-1} B^V_{K_{n-i}} + \prod_{i=1}^{k} B^V_{K_{n-i}} \Sigma_{n-k}
\end{aligned}$$

This procedure ends when $k = n - m - 1$, since Σ is by definition equal to zero for $k \geq n - m - 1$:

$$\Sigma_{n-(n-m-1)} = \Sigma_{m+1} = \sum_{i=m+1}^{(m+1)-1} B_R(t, t_i) = 0$$

The expression for Σ upon completion of the recursion is thus

$$\Sigma_n = \sum_{j=1}^{n-m-1} B_{K_{n-j}}(t, t_{m+1}) \prod_{i=1}^{j-1} B_{K_{n-i}}(t_m, t_{m+1})$$

observing the usual convention that a sum is equal to zero and a product is equal to one if the upper limit is less than the lower limit. Renaming the index variables as

$$j \to x := n - j \quad \text{and} \quad i \to y := n - i$$

and then rearranging the indices of the sum and the product so that they are performed in reverse order we obtain

$$\Sigma_n = \sum_{x=m+1}^{n-1} B_{K_x}(t, t_{m+1}) \prod_{y=x+1}^{n-1} B_{K_y}(t_m, t_{m+1})$$

Renaming the indexes again as $x = j$ and $y = i$, and substituting this Σ_n into the original Equation 27.2 finally yields a closed form expression for the zero bond discount factor:

$$\begin{aligned} B_R(t, t_n) =& B_{K_n}(t, t_{m+1}) - [1 - B_{K_n}(t_m, t_{m+1})] \\ & \times \sum_{j=m+1}^{n-1} B_{K_j}(t, t_{m+1}) \prod_{i=j+1}^{n-1} B_{K_i}(t_m, t_{m+1}) \end{aligned} \tag{27.3}$$

The right-hand side of this equation is expressed entirely in terms of the *par* rates! This equation makes it unnecessary to determine the zero rates up to the last coupon payment date t_{n-1} in order to determine the zero rate up to t_n. The zero rates can be computed *directly* from the par rates. Equation 27.3 is valid for arbitrary value dates t and holds for all compounding conventions, even if *different* compounding methods are applied for zero rates and par rates. The only assumption made for Equation 27.3 was that all complete coupon periods between t_{m+1} and t_n are of equal length (the period between t and t_{m+1} may be incomplete).

The reader should explicitly verify[3] that the results obtained from the closed form Equation 27.3 actually agree with those obtained using the general bootstrapping method, Equation 27.1, for coupon periods of equal length (i.e. with 27.2). In the accompanying Excel workbook BOOTSTRAPPING.XLS, a demonstration of the computation of the spot rates from a given data set of par rates is provided, first using the conventional bootstrapping method Equation 27.2, and then using the closed form expression Equation 27.3 (see also Figure 27.1). Simple, annual compounding was used

[3]Taking $m = 0$, $n = 1, 2, 3$ should be sufficient to develop confidence in Equation 27.3.

Spot Date t	Jan-18-01
Length of Current Coupon Period (30/360)	0.7667
Length of Full Coupon Period (30/360)	1.0000
Time to Maturity T (Act/365f)	9.769863

Input		Par Rate Discount Factor		Closed Form Method			General Classic
Coupon Date	Par Rate 30/360	for Full Coupon Period	for Running Coupon Period	j	Product over i	"j-th" Term in Sum	Spot Rate Discount Factor for Coupon Date
Okt-24-01	5.27%	0.9499182	0.9611500	1	0.64650	0.621384	0.961149963
Okt-24-02	5.36%	0.9491268	0.9605287	2	0.68115	0.654267	0.911631904
Okt-24-03	5.44%	0.9484067	0.9599631	3	0.71821	0.689453	0.863340097
Okt-24-04	5.51%	0.9477775	0.9594688	4	0.75778	0.727067	0.816581604
Okt-24-05	5.58%	0.9471491	0.9589750	5	0.80006	0.767242	0.771211400
Okt-24-06	5.65%	0.9465215	0.9584818	6	0.84527	0.810174	0.727245423
Okt-24-07	5.72%	0.9458948	0.9579890	7	0.89362	0.856076	0.684695030
Okt-24-08	5.76%	0.9455073	0.9576842	8	0.94512	0.905127	0.645121740
Okt-24-09	5.81%	0.9451200	0.9573796	9	1.00000	0.957380	0.607191755
Okt-24-10	5.85%	0.9447331	0.9570752	**Sum over j =**		**6.988169**	0.570860836

Zero Bond with Maturity	**Okt-24-10**	**0.570861**	**0.570861**
Spot Rate for Maturity	**Okt-24-10**	**0.057382**	**0.057382**

Figure 27.1: Bootstrapping for coupon periods of equal length according to the methods represented by Equations 27.2 and 27.3. The results agree exactly.

for the coupon payments with the 30/360 day count convention; for the zero rate, continuous compounding was assumed with the Act/365 day count convention. The most important intermediate steps are indicated. The results from both methods agree exactly. The same input data are used here as will be used in Section 27.1.3 on the classical bootstrapping method (which holds only for complete coupon periods and for simple compounding over the periods). As expected, we obtain identical results with the more general Equations 27.2 and 27.3 if we use the same day count and compounding conventions as in classical bootstrapping. Note, however, that Equations 27.2 and 27.3 also hold when the first period is incomplete (*stub*), and for arbitrary compounding methods as well.

All this is demonstrated in Figure 27.1, which presents the still quite general case for which the value date is not the beginning of a coupon period and different compounding and day count conventions are used for the spot and par rates. For the special case that the value date happens to coincide with the beginning of a coupon period, i.e. $t = t_m$ (Assumption 1 at the beginning of Section 27.1), the closed form equation 27.3 reduces to

$$B_R(t_m, t_n) = B_{K_n}(t_m, t_{m+1}) - [1 - B_{K_n}(t_m, t_{m+1})] \sum_{j=m+1}^{n-1} \prod_{i=j}^{n-1} B_{K_i}(t_m, t_{m+1})$$

For *annual* coupon payments with discrete compounding, the coupon discount factors over entire 1-year periods are simply $B_{K_i} = 1/(1 + K_i)$ and the closed form expression becomes

$$\begin{aligned} B_R(t_m, t_n) &= \frac{1}{1+K_n} - \left[1 - \frac{1}{1+K_n}\right] \sum_{j=m+1}^{n-1} \prod_{i=j}^{n-1} \frac{1}{1+K_i} \\ &= \frac{1}{1+K_n} - \left[\frac{1+K_n-1}{1+K_n}\right] \sum_{j=m+1}^{n-1} \prod_{i=j}^{n-1} \frac{1}{1+K_i} \\ &= \frac{1}{1+K_n} - \left[\frac{K_n}{1+K_n}\right] \sum_{j=m+1}^{n-1} \prod_{i=j}^{n-1} \frac{1}{1+K_i} \end{aligned}$$

and finally

$$B_R(t_m, t_n) = \frac{1}{1+K_n} \left(1 - K_n \sum_{j=m+1}^{n-1} \prod_{i=j}^{n-1} \frac{1}{1+K_i}\right)$$

This equation, which holds only for the special case of annual coupon payments over periods of exactly one year (Assumption 2) and discrete compounding (Assumption 3) starting at the beginning of a complete coupon period (Assumption 1), was already given in [70]. But even in the more general case relying only on Assumption 2, the closed form expression in Equation 27.3 eliminates the need for a recursion if the coupon periods are of identical length.

27.1.3 Classical Bootstrapping

The bootstrap method commonly presented in the literature assumes that the value date t falls at the start of a complete coupon period, $t = t_m$ (Assumption 1), in addition to applying the simple compounding method (Assumption 3) over the coupon periods

$$B_{K_n}(t_{i-1}, t_i) = \frac{1}{1 + K_n (t_i - t_{i-1})}$$

where K_n denotes the coupon rate for the bond with maturity $T = t_n$. Substituting into the general equation 27.1, we obtain the bootstrapping equation in its best-known form

$$B_R(t, t_n) = \frac{1 - K_n \sum \tau_i B_R(t, t_i)}{1 + K_n \tau_n} \tag{27.4}$$

Here, $\tau_i = t_i - t_{i-1}$ denotes the length of the coupon period between t_{i-1} and t_i in accordance with the observed day count convention. Thus, the period lengths τ_i can by all means vary, for instance depending on whether a coupon payment date happens to coincide with a holiday or the like. Assumption 2 at the beginning of Section 27.1 is *not* necessary for the classical bootstrapping method represented by Equation 27.4.

Money Market Rates

Since different instruments are used when computing the interest rates associated with the various terms in the construction of a term structure, an initial classification of the market segments is necessary: at the short end of the yield curve, *money market rates* and money market futures such as *euro futures* are normally applied. Since daily, weekly and monthly money market rates are, in principle, zero coupon bonds, the quoted rates $R(t, t_i)$ with maturity t_i can be used to compute the discount factors:

$$B_R(t, t_i) = \frac{1}{1 + \tau_i R(t, t_i)} \tag{27.5}$$

Money Market Futures

Liquid money market futures are usually preferred to money market rates. Money market futures are discussed in great detail in Section 15.5.2. In order to determine the discount factor for the first future with maturity t_i and an interest period from t_i until t_{i+1} we must first compute $B_R(t, t_i)$. This *stub rate* can frequently be estimated from the surrounding money market rates. Since the (forward neutral) expectation of the future 3-month rate (not, however the forward rate) can be read directly from the quotation of the 3-month money market future in the market, a *convexity adjustment* as described in Section 15.5.2 (either according to the Approximation 15.26 or by means of a term structure model) must be taken into account when calculating the implied forward discount factor $B_R(t_i, t_{i+1})$. These forward discount factors can then be used to compute spot discount factors through simple arbitrage consideration as in Equation 3.3. If the futures contracts form a *future strip*, i.e. if the interest periods form one continuous time interval without overlap, then with each calculated discount factor we can calculate the next forward discount factor from the corresponding money market future quote. We again emphasize at this point that money market futures can only be used for the calculation of discount factors in combination with a convexity adjustment. For this reason we simply use money market rates directly as input in the example below to avoid unnecessary complications.

Swap Rates

For longer terms, determining the term structure is much more difficult. The quoted yields for government bonds, mortgage bonds or the swap rates (to name three important levels of credit worthiness) are *not* zero coupon rates and can thus not be used immediately to find the discount factors $B_R(t, t_i)$. Furthermore, the mortgage bond markets usually offer too small a selection for the methods introduced below since the several term segments are comparatively illiquid and in addition, not enough bonds are available whose coupons match their yields (par bonds). In contrast, *swaps* are highly liquid and are quoted with small bid-ask spreads in all terms up to ten years (and often more). A swap is nothing other than the combination of a *fixed leg* or *coupon strip* whose payments are calculated on the basis of the fixed swap rate, and a *floating leg* whose payments are dependent on a floating rate referred to as a *LIBOR strip* below. If the swap rate is chosen so that the present values of both strips are equal, then the

coupon strip corresponds to a fictitious coupon bond (financed with a floating rate money market transaction) with maturity t_n and with a coupon K_n lying exactly on the yield curve for the credit worthiness of the swap, i.e. the coupon strip corresponds to a par bond. Instead of the par rates in the above bootstrapping equation 27.4, the corresponding *swap rates* are therefore frequently used in practice.

With this bootstrapping equation, we can determine the spot rate discount factor for the term t_n from the preceding discount factors and the coupon rate K_n. If the discount factor for one year is determined from the money market rates or futures prices, then the two year discount factor can be computed from the swap rate for two years $K_2 = K(t_2)$. Once this discount factor is available it can in turn be used to compute the next discount factor from the next swap rate, and so on. This *bootstrapping* procedure thus generates the discount factors for all terms for which the swap rates exist[4]. Note that all instruments used in the bootstrapping must have the same credit worthiness; the swaps must be as credit worthy as the futures and the money market rates.

In summary we note that from given money market rates, futures prices and swap rates, the arbitrage-free discount factors at the corresponding maturities can be determined through bootstrapping. Since the payments of other instruments to be priced may occur between these *vertices* of the interest rate curve, an additional step, namely an *interpolation*, must be made to estimate the discount factors between theses vertices.

Example

In the Excel workbook BOOTSTRAPPING.XLS, the traditional bootstrapping method is explicitly performed for the example of the yield curve at October 24th, 2000 (see Figure 27.2). The *business day convention* is *modified following* under consideration of the *Frankfurt holiday calender* and *day count conventions* have been observed which are typical of the respective markets: *Act/360* for the money market and *30/360* for the swap market (see Section 3.1.1).

We will now walk through the example step by step:

- Determining the times to the maturity of the benchmark instruments.
- Filling the "Swap-Gaps".
- Determining the discount factors from the money market rates.
- Determining the discount factors for the final maturity of each of the quoted swap rates.
- Determining the zero bond yields from the discount factors for any desired compounding method and day count convention.

[4]If a swap rate should be missing, it can be estimated from the neighboring rates by an interpolation (e.g. linear interpolation).

		Input		Spot Date Okt-24-00				spot rate	
		Market Rate	Maturity Date	Time to Maturity in Years	Length of Period	$\Sigma\tau_i B_R(t,t_i)$	discount factor	conti nuous	simple
money market	O/N	4.740%	**Okt-25-00**	0.0028	Ending at Swap Maturity		0.999868	4.806%	4.740%
	1m	4.835%	Nov-24-00	0.0861			0.995854	4.892%	4.996%
	2m	4.890%	**Dez-27-00**	0.1778			0.991382	4.936%	4.968%
	3m	5.010%	Jan-24-01	0.2556			0.987359	5.047%	5.121%
	4m	5.025%	**Feb-26-01**	0.3472			0.982851	5.051%	5.149%
	5m	5.040%	**Mrz-26-01**	0.4250			0.979029	5.056%	5.073%
	6m	5.085%	Apr-24-01	0.5056			0.974937	5.090%	5.142%
	9m	5.145%	Jul-24-01	0.7583			0.962449	5.117%	5.202%
	12m	5.200%	Okt-24-01	1.0139	1.00000		0.949918	5.138%	5.272%
swap market	2y	5.360%	Okt-24-02	2.0000	1.00000	0.9499182	0.900801	5.224%	
	3y	5.440%	Okt-24-03	3.0000	1.00000	1.8507196	0.852922	5.303%	
	4y	5.510%	**Okt-25-04**	4.0028	1.00278	2.7036415	0.806469	5.370%	
	5y	5.580%	Okt-24-05	5.0000	0.99722	3.5123512	0.761630	5.443%	
	6y	***5.650%***	Okt-24-06	6.0000	1.00000	4.2718654	0.718069	5.517%	
	7y	5.720%	Okt-24-07	7.0000	1.00000	4.9899342	0.675914	5.593%	
	8y	***5.763%***	Okt-24-08	8.0000	1.00000	5.6658477	0.636762	5.638%	
	9y	***5.807%***	**Okt-26-09**	9.0056	1.00556	6.3026098	0.599040	5.687%	
	10y	5.850%	**Okt-25-10**	10.0028	0.99722	6.9049777	0.563203	5.736%	

Figure 27.2: Determining the yield curve using classical bootstrapping. The bold-faced yields in the column "Market Rate" are interpolated (swap gaps). The bold-faced data in the column "Maturity Date" are adjusted according to the business day convention.

Determination of the Times to Maturity of the Benchmark Instruments First, using the business day convention as indicated in Section the actual maturity date for each of the benchmark instruments is determined (see under the column "Maturity Date" in Figure 27.2). Then, for each benchmark instrument, the associated day count convention in accordance with Section 3.1.1 is used to establish the time between the current spot date (at which the spot rate curve is to be determined) and the maturity date in years; this value can be found under the column "Time to Maturity in Years" in Figure 27.2. We can clearly see that because of the day count convention, the times to maturity do not agree exactly with the "lifetimes". For example, $t_i - t$ for three months is not exactly 0.25 or for twelve months not exactly one year. For example the time to maturity for the six month rate is according to the Act/360 convention

$$(24.04.2001 - 24.10.2000)/360 = 0,5055555556$$

Filling the "Swap Gaps" Swap rates are usually not available in the market for each of the maturity dates required for the bootstrapping; for example, in Figure 27.2, the market data for maturities of 5, 7 and 10 years are quoted but not for 6, 8 and 9 years. According to the bootstrapping equation, a discount factor for each coupon

payment date is needed to determine the discount factor at the maturity date of a swap. Since a 10-year swap in our example has annual coupon payments, the discount factors for the coupon payment dates in year 8 and 9 must also be known. This holds accordingly for the 7 year swap which has coupon payments in 6 years. These *swap gaps* can be filled through interpolation. One of the most common methods for swap rates is the *linear interpolation* between vertices (see Equation 24.1) as described in Section 24.2.1. This interpolation method was used to estimate the 6, 8 and 9-year swap rates listed in the column "Market Rate" in Figure 27.2. These values have been written in bold in order to differentiate them from data obtained directly from the market.

Determination of the Discount Factors from the Money Market Rates We now begin with the calculation of the discount factors, independent of all conventions and compounding methods (the column "Discount Factor" in Figure 27.2) from which zero bond yields can later be derived for any arbitrary compounding and day count convention. The discount factors in the short-term range of less than one year (money market) are easily determined from the quoted money market yields using linear compounding, i.e. using Equation 27.5. The term corresponding to the required day count convention Act/360 can be found in the column "Time to Maturity in Years". For instance the 6-month discount factor becomes

$$\frac{1}{1 + 0.5055555556 \times 5.085\%} = 0.9749368119$$

The Discount Factors for the Maturity of Each of the Quoted Swap Rates Since we now have a complete set of swap rates for the term between two and ten years, the discount factors for each maturity date can be determined via bootstrapping. We begin with the 2-year swap rate

$$B_R(t, t_2) = \frac{1 - K_2 \tau_1 B_R(t, t_1)}{1 + K_2 \tau_2}$$

Here, τ_2 is the duration (in years) of the last swap period of the 2-year swap, i.e. the length of time from Oct-24-2001 until Oct-24-2002 in the swap market convention 30/360. Correspondingly τ_1 is the duration of the first swap period of the 2-year swap, i.e. the length of time from Oct-24-2000 until Oct-24-2001, *likewise in the swap market convention 30/360*. The duration 1.01388889 of this time span already computed in the money market convention Act/360 should *not* be applied here! In the swap market convention, this time span yields a value of exactly 1.000000.

Furthermore, the discount factor for the first swap period $B_R(t, t_1)$ is needed for the computation of the second $B_R(t, t_2)$. This has already been computed from the 12-month money market rate. The12-month discount factor corresponds exactly to the first coupon payment date of the 2-year swap, namely 24.10.2001. In total, we obtain

$$B_R(t, t_2) = \frac{1 - 5.360\% \times 1.0000000 \times 0.9499182015}{1 + 5.360\% \times 1.0000000} = 0.9008014279$$

This $B_R(t,t_2)$ as well as $B_R(t,t_1)$ can now be substituted into the equation for the computation of $B_R(t,t_3)$.

$$B_R(t,t_3) = \frac{1 - K_3 \sum_{i=1}^{2} \tau_i B_R(t,t_i)}{1 + K_3 \tau_3}$$

The calculation of the sum of the zero bond discount factors $\sum \tau_i B_R(t,t_i)$ is presented in a separate column in Figure 27.2.

The final result, the zero bond discount factors, are finally computed and presented in the column "Discount Factor". After having uniquely determined the discount factors for each time point, we are now able to find the arbitrage-free prices of all structures with arbitrary cash flows. The important and advantageous property of the discount factors is that they are *independent of conventions*. While the money market rates and the swap rates are quoted on the basis of various different bank holiday and day count conventions, the discount factors in the term structure are *unique*. Once they are generated, it is possible to price any arbitrary cash flow structures.

Determination of the Zero Bond Yields from the Discount Factors After the bootstrappings, the actual spot rate term structure is obtained from the discount factors in any desired compounding method as indicated in Table 3.4. For this, each time to maturity is calculated in terms of the desired day count convention. The column "Spot Rate Continuous" shows as an example the zero bond yields (the spot rates) at all vertices for the day count convention Act/365 and exponential compounding.

27.2 Interpolations

In the last section, bootstrapping was used to determine the discount factors for various maturities from a series of quoted money market and swap rates. The time points established in the term structure curve from the various terms are called *vertices*. Now it is quite probable that an existing instrument which is to be priced with the aid of the term structure has cash flow dates which do not correspond to vertices. In order to determine the discount factors for these terms, we must rely on an appropriate interpolation procedure.

Interpolation of the term structure curve is a very broad and controversial issue. Different philosophies deviate from one another as to whether spot rates, discount factors or even the forward rates should be interpolated. Many of these different points of view assume from the outset an interpolation using theoretical functions whose parameters are determined with the aid of fitting procedures.

These models can often only be realized with the implementation of complicated numerical procedures. Furthermore, the user of these procedures must have very exact knowledge of the parameters used and the implications these parameters have on the results. Small deviations within a certain range of the term structure curve may lead to considerable errors in the valuation. On the other hand, there exist simple and efficient interpolation schemes which at the very least guarantee that the vertices of the term structure curve are hit exactly. For instance, linear interpolation of the zero bond

rates corresponds (for continuous compounding) to *exponential interpolation* for the discount factors. The implementation of this exponential procedure is just as simple and efficient as linear interpolation. For a discount factor with maturity T between two neighboring vertices t_i and t_{i+1} (sometimes called the *surroundings*), the exponential interpolation is performed as follows:

$$\begin{aligned}
\lambda &= \frac{t_{i+1} - T}{t_{i+1} - t_i} \quad \forall\ t_{i+1} > T > t_i \\
\frac{\ln\ B_R(t,T)}{T-t} &= \lambda \frac{\ln\ B_R(t,t_i)}{t_i - t} + (1-\lambda)\frac{\ln\ B_R(t,t_{i+1})}{t_{i+1}-t} \Longrightarrow \\
B_R(t,T) &= \exp\left[(T-t)\left(\lambda \frac{\ln\ B_R(t,t_i)}{t_i - t} + (1-\lambda)\frac{\ln\ B_R(t,t_{i+1})}{t_{i+1}-t}\right)\right]
\end{aligned}$$

28

Volatility

28.1 Implied Volatilities

28.1.1 Smiles and Volatility Indices

The most up-to-date estimation for the volatility of a risk factor can be obtained when the risk factor is the underlying of a liquid option. Since the option is liquid, it is traded with a volume which ensures that the bid-ask spreads are sufficiently small to define a precise option price. The volatility can then be determined from the pricing model valid for the valuation of the option in the market if all other factors influencing the price of the option (interest rates, price of the underlying, etc.) are known. The parameter volatility is then varied in the model until the it yields the market price of the option. In this way, the *implied volatility* is determined. If there is agreement in the market as to which model should be used for the determination of the option price, then there is a one-to-one relation between option price and this implied volatility; if one of these values is known, the other can be uniquely determined. This method is so common in the market that many options are quoted directly in terms of their implied volatility instead of their price.

Often, the assumptions on which the pricing models are based (for example, the assumption of continuous trading) are satisfied reasonably well for options *at* the money, but not for options which are far in or out of the money. This has the effect that the pricing model yields different implied volatilities for options that are not at the money. If the implied volatilities are plotted as a function of the strike price, the curve obtained is not the graph of a constant function as assumed by the Black-Scholes model, for example. For most options we observe the effect that the implied volatilities increase with increasing distance from the at-the-money strike (in other words, when the option is increasingly in or out of the money); the resulting volatility curve tends to look like a *smile*. This expression has come to be the accepted term for describing the implied volatility graphed as a function of the *strike* and is independent of the form of the curve.

Options with different *maturities* also show different implied volatilities even if the strike is the same. As with interest rates, the dependence of the volatility on the time

to maturity of the option is known as the *volatility term structure.*

The variables maturity date and strike price are *absolute* variables and as such are not very suitable for the comparison of the implied volatilities of different options on the same underlying; the variables *time to* maturity and *relative* strike price are preferable. The relative strike is the strike expressed as a percentage of the current underlying price. A call with a strike of 84 euros on an underlying with a spot price of 80 euros is 5% out of the money and therefore has a relative strike of -5%. A corresponding put would have a relative strike of +5%. This relative strike is called *moneyness.* The moneyness is thus defined as the relative distance between the strike K and the at-the-money point

$$\text{moneyness} = \frac{K - S(t)}{S(t)}$$

where $S(t)$ is the current spot price[1] of the underlying. The implied volatility of options on the same underlying (for example, the DAX) as a function of time to maturity and relative strike is called the *volatility surface.* The volatility surfaces obtained from liquid option prices can be used to price illiquid options or OTCs on the same underlying.

For options on interest rates or interest rate instruments such as a cap or an option on a Bund future, the implied volatility depends on a third parameter. This is the time to maturity of the underlying. For such options, a complete volatility surface is determined by the volatility as a function of the options' maturities and the maturities of the underlying as well as the options' relative strikes. The construction of such structure requires many liquid options. Therefore, these structures usually contain a large number of gaps. The gaps on the volatility structure of an underlying, i.e. the implied volatilities for the parameter combination for which no liquid options are available, are usually closed with clever (or courageous!) interpolation schemes.

An example of an implied volatility published by an exchange, a *volatility index,* is the *VDAX* published since December 5, 1994 by the DTB (later renamed Eurex). This index refers to one single point of the volatility structure of the DAX, namely the volatility of the DAX for a (fictitious) at the money option with a time to maturity of 45 days. Not the current DAX value is considered at the money but rather the forward price of the DAX in 45 days. This forward price is available since the prices of DAX futures contracts are known (or can be determined from the prices of DAX options using the put-call parity Equation 7.7). The volatility of the fictitious ATM option is then determined as follows:

- For each maturity date, only those option series are considered whose strikes are at most 100 points away from the ATM point. For a series of options with strikes increasing in 25 point increments, this allows a maximum of 16 options (eight calls and eight puts) per maturity date.

[1] Sometimes the at-the-money-point is defined to be at the current *forward* price $S(t,T)$ of the underlying, where T is the maturity of the option (*at-the-money forward*). Correspondingly, the moneyness is then defined as the relative distance between the strike and this forward price.

- The theoretical option price is calculated using the Black-Scholes Equation 9.7 with the same volatility for all options with the same maturity date. This volatility is then varied as an input parameter until the sum of the squared difference between the theoretical option prices per maturity date attain a minimum value (*least squares fit*). In this way, a fitted implied volatility for each maturity date is obtained.
- The actual VDAX at time t is then obtained by interpolating the implied volatilities for the two surrounding maturities T' and T'' of the maturity given by $T = t + 45$ as follows

$$\text{VDAX}(t) = \sqrt{\frac{T'' - T}{T' - T}\sigma(T')^2 + \frac{T - T'}{T'' - T'}\sigma(T'')^2}$$
$$\text{where} \quad T' \leq T < T'' \quad , \quad T = t + 45 \text{ Days}$$

28.2 Local Volatility Surfaces

The prices of traded options contain a great deal more information about the underlying than just the implied volatility. For general underlyings modeled as Ito processes of the form 3.16 in Section 3.4, the transition probabilities and the *Dupire volatilities*, also referred to as *local volatilities*, can be derived by differentiating the option price. The reader is advised to become familiar with Sections 3.4 and 8.2 as the material in these sections are essential for a sound comprehension of what follows.

28.2.1 Implicit Transition Probabilities

The central relation holding for all Ito processes of the form 3.16 between the transition probability p defined in Equation 3.27a and the price V of a financial instrument (an option, for example) has already been indicated in Equation 8.16 of Section 8.2:

$$V(S,t,T) = B_r(t,T)\int_{-\infty}^{\infty} f(S',T)p(S',T\,|S,t\,)dS' \tag{28.1}$$

where the function $f(S',T)$ corresponds to the payoff of the financial instrument at maturity T. This equation can be inverted to determine the transition probabilities p from the prices of liquid options observed in the market; in doing so we determine the *implied transition probabilities.* We assume here that the market prices of a series of liquid plain vanilla options (both calls and puts) with different strikes K and maturities T are available.

For a call with a payoff profile $f(S', t' = T) = \max(S' - K, 0)$, Equation 28.1 becomes

$$V_{\text{call}} = B_r(t,T)\int_{K}^{\infty} (S' - K)p(S',T\,|t,S\,)dS'$$

The price of the call is now differentiated with respect to the limit of integration K for the purpose of extracting p from the integral. This yields the integrand evaluated at the (lower) limit of integration $S' = K$ and, since K itself appears in the integrand, a

second term arises, involving the integral of the derivative of the integrand with respect to the strike:

$$\frac{\partial V_{\text{call}}}{\partial K} = -B_r(t,T)\underbrace{(K-K)}_{0}p(K,T\,|S,t\,) - B_r(t,T)\int_K^\infty p(S',T\,|t,S\,)dS'$$

Thus

$$\frac{\partial V_{\text{call}}}{\partial K} = -B_r(t,T)\int_K^\infty p(S',T\,|t,S\,)dS' \tag{28.2}$$

For a plain vanilla put with a payoff profile $f(S') = \max(K-S',0)$ we proceed analogously to obtain

$$\frac{\partial V_{\text{put}}}{\partial K} = +B_r(t,T)\int_{-\infty}^K p(S',T\,|t,S\,)dS'$$

A further derivative with respect to K yields the same result for both puts and calls:

$$\frac{\partial^2 V}{\partial K^2} = B_r(t,T)p(K,T\,|t,S\,) \tag{28.3}$$

This can now be easily solved for the transition probability p[2]

$$p(K,T\,|t,S\,) = B_r(t,T)^{-1}\frac{\partial^2 V}{\partial K^2} \tag{28.4}$$

In this way, the implicit transition probabilities can (at least theoretically) be obtained from the observed option prices (with fixed t and S, the strike and maturity date being variable). Through Equation 28.4 we can determine (at time t and known S) the transition probability p associated with the points (K,T) for which call or put prices exist and thus (after an appropriate interpolation between these points) an *implied transition probability surface.* Noted that not only the price but also the *second derivative* of the price function with respect to the strike must be known. Since the numerical determination of derivatives on the basis of measured data is quite prone to error, this method can only be implemented in practice if an analytic formula for the observed option prices as a function of the strike price is available; the derivative with respect to the strike can then be performed analytically. For example, the derivative with respect to the strike price can be calculated from the Black-Scholes equation for the price of a call[3]

$$\frac{\partial V_{\text{call}}}{\partial K} = -B_r(t,T)N(x-\sigma\sqrt{T-t})$$

with x as in Equation 9.5. The second derivative is then

$$\frac{\partial^2 V}{\partial K^2} = \frac{B_r(t,T)}{K\,\sigma\sqrt{T-t}}\mathrm{N}'(x-\sigma\sqrt{T-t})$$

[2]Observe that in the derivation of this equation we assumed certain (plain vanilla) payoff profiles for the options but we did not assume any specific underlying process. Thus, everything is valid for plain vanilla calls and puts on arbitrary underlyings of the form 3.16.

[3]Similar derivatives have already been calculated in Section 13.6.2.

Because of the put-call parity, the second derivative of the put price can immediately be calculated by taking the derivative with respect to K on both sides of the equation $\frac{\partial V_{\text{put}}}{\partial K} = \frac{\partial V_{\text{call}}}{\partial K} + B_r(t,T)$; this shows again that the second derivative with respect to K of the put and call are identical. Hence, for the Black-Scholes case, the implicit transition probabilities are given by

$$p(K,T\,|t,S\,) = \frac{\exp\left\{-\frac{1}{2}\left(x-\sigma\sqrt{T-t}\right)^2\right\}}{K\sqrt{2\pi\sigma^2(T-t)}} \tag{28.5}$$

The observed option prices no longer appear explicitly in this expression but the volatilities do. These are the implied Black-Scholes volatilities which yield the observed option price when substituted into the Black Scholes pricing formula. The implied Black-Scholes volatilities are nothing other than the observed option prices quoted in another form; the Black-Scholes formula is the mapping between the two ways of quoting (on the one hand the option price, on the other the implied volatility). In the Excel workbook ImpliedVolAndTransProb.xls, the computation of the implied Black-Scholes volatilities from quoted ODAX options (from February 2000) are explicitly performed by means of the *Newton-Raphson method* along with the subsequent calculation of the transition probabilities.

In this way, we have constructed the implicit transition probabilities enabling, theoretically, the determination of the price of other (OTC) options via Equation 28.1. However, we need a range of S' (the first argument of p in Equation 28.1) from $-\infty$ to ∞. The first argument of p corresponds to the strike, when determining p via Equation 28.4. Therefore we need traded options for strikes ranging from $-\infty$ to ∞ and thus cannot proceed without extrapolation arguments.

Observe that Equation 28.5 agrees exactly with the transition probability 3.32 for the process 3.21 given earlier. Thus, everything is consistent. This can be seen by making the obvious associations $t' = T$ and $S' = K$ in 3.32 and substituting the risk-neutral drift $\widetilde{\mu}$ (see Equation 11.25 in connection with Equation 3.23):

$$\widetilde{\mu} = \frac{1}{T-t}\ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right) \tag{28.6}$$

28.2.2 Implicit Local Volatility Surfaces

The *local volatilities* can be determined in a similar manner as illustrated above for finding the implicit transition probabilities from observed market prices of derivatives. These are to be distinguished from the Black-Scholes volatilities (which are nothing other than a way of quoting option prices where the mapping from one quotation to the other is defined by the Black-Scholes formula). The method introduced here goes back to a procedure developed by Dupire. *Local volatilities* are therefore also referred to as *Dupire volatilities.*

We again assume that market prices $V(S,t,K,T)$ of a series of plain vanilla options (calls and puts) on an underlying with price $S(t)$ are known at time t for various strikes K and maturities $t' = T$. Starting from the central, general equation 28.1, we now take

the derivative with respect to T and then use the forward equation 3.28 to substitute for $\partial p/\partial T$. For example, for a call with a payoff profile given by $f(S',T) = \max(S'-K,0)$:

$$\begin{aligned}\frac{\partial V_{\text{call}}}{\partial T} &= \frac{\partial B_r(t,T)}{\partial T}\frac{V_{\text{call}}}{B_r(t,T)} + B_r(t,T)\int_K^\infty (S'-K)\frac{\partial p(S',T\,|S,t\,)}{\partial T}dS' \\ &= \frac{\partial \ln B_r(t,T)}{\partial T}V_{\text{call}} \\ &+ B_r(t,T)\int_K^\infty (S'-K)\left[\frac{1}{2}\frac{\partial^2}{\partial S'^2}\left[b(S',T)^2 p\right] - \frac{\partial}{\partial S'}\left[a(S',T)p\right]\right]dS' \\ &= \frac{\partial \ln B_r(t,T)}{\partial T}V_{\text{call}} + B_r(t,T)(\frac{1}{2}I_1 - I_2)\end{aligned}$$

The derivative with respect to the variable of integration in the integrand suggests that we proceed by integrating by parts.

$$\begin{aligned}I_1 &\equiv \int_K^\infty (S'-K)\frac{\partial^2(b^2p)}{\partial S'^2}dS' \\ &= (S'-K)\frac{\partial(b^2p)}{\partial S'}\Big|_K^\infty - \int_K^\infty \frac{\partial(S'-K)}{\partial S'}\frac{\partial(b^2p)}{\partial S'}dS'\end{aligned}$$

A reasonable assumption would be that, starting from S, the probability p (and its derivatives) for reaching S' in a finite time $(T-t)$ should converge towards zero very fast for $S' \to \infty$. The upper limit $S' = \infty$ of the first term on the right-hand side then makes no contribution to the evaluation of the integral. The expression evaluated at the lower limit $S' = K$ is obviously equal to zero as well. Since $\partial(S'-K)/\partial S' = 1$, the above expression for I_1 reduces to

$$I_1 = -\int_K^\infty \frac{\partial(b^2p)}{\partial S'}dS' = -\,b^2p\big|_K^\infty = b(K,T)^2 p(K,T\,|S,t\,)$$

where we have again made use of the fact that $p(S',T\,|S,t\,)$ converges towards zero very fast as $S' \to \infty$ and that therefore the upper limit contributes nothing to the integral. Proceeding analogously with the second integral I_2 yields after integration by parts

$$\begin{aligned}I_2 &\equiv \int_K^\infty (S'-K)\frac{\partial(ap)}{\partial S'}dS' \\ &= \underbrace{(S'-K)ap\big|_K^\infty}_{0} - \int_K^\infty \underbrace{\frac{\partial(S'-K)}{\partial S'}}_{1} a\,p\,dS' \\ &= -\int_K^\infty a(S',T)p(K,T\,|S,t\,)dS'\end{aligned}$$

In total, $\partial V/\partial T$ becomes

$$\begin{aligned}\frac{\partial V_{\text{call}}}{\partial T} &= \frac{\partial \ln B_r(t,T)}{\partial T}V_{\text{call}} \\ &+ B_r(t,T)\left[\frac{1}{2}b(K,T)^2 p(K,T\,|S,t\,) + \int_K^\infty a(S',T)p(K,T\,|S,t\,)dS'\right]\end{aligned}$$

Analogously for puts

$$\frac{\partial V_{\text{put}}}{\partial T} = \frac{\partial \ln B_r(t,T)}{\partial T} V_{\text{put}} + B_r(t,T)\left[\frac{1}{2}b(K,T)^2 p(K,T\,|S,t) - \int_{-\infty}^{K} a(S',T)p(K,T\,|S,t)dS'\right]$$

where an analogous assumption has been made that $p(S',T\,|S,t)$ for $S' \to -\infty$ converges towards zero fast enough and that, consequently, the lower limit makes no contribution to the integral.

As yet, our deliberations hold for every general Ito process of the form 3.16. From now on, however, an explicit stochastic process must be assumed in order to continue with the computation. For the special case of a *risk-neutral* random walk, i.e. if the parameters are chosen as in Equation 8.1 with a volatility $\sigma = \sigma(S,t)$ and a risk-neutral drift, Equation 28.6, we obtain for the call

$$\frac{\partial V_{\text{call}}}{\partial T} = \frac{\partial \ln B_r(t,T)}{\partial T} V_{\text{call}} + B_r(t,T)\frac{1}{2}\sigma(K,T)^2 K^2 p(K,T\,|S,t) + \frac{B_r(t,T)}{T-t}\ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right)\int_K^{\infty} p(S',T\,|S,t)S'dS'$$

and for the put

$$\frac{\partial V_{\text{put}}}{\partial T} = \frac{\partial \ln B_r(t,T)}{\partial T} V_{\text{put}} + B_r(t,T)\frac{1}{2}\sigma(K,T)^2 K^2 p(K,T\,|S,t) - \frac{B_r(t,T)}{T-t}\ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right)\int_{-\infty}^{K} p(S',T\,|S,t)S'dS'$$

Using Equations 28.1 and 28.2, the remaining integral for the call can be written as

$$\begin{aligned}\int_K^{\infty} p(S',t'\,|S,t)S'dS' &= \int_K^{\infty} p(S',t'\,|S,t)(S'-K)dS' \\ &\quad + K\int_K^{\infty} p(S',t'\,|S,t)dS' \\ &= B_r(t,T)^{-1}V_{\text{call}} - KB_r(t,T)^{-1}\frac{\partial V_{\text{call}}}{\partial K}\end{aligned}$$

and analogously for the corresponding integral for the put

$$\begin{aligned}\int_{-\infty}^{K} p(S',t'\,|S,t)S'dS' &= -\int_{-\infty}^{K} p(S',t'\,|S,t)(K-S')dS' \\ &\quad + K\int_{-\infty}^{K} p(S',t'\,|S,t)dS' \\ &= -B_r(t,T)^{-1}V_{\text{put}} + KB_r(t,T)^{-1}\frac{\partial V_{\text{put}}}{\partial K}\end{aligned}$$

The remaining p is replaced by the expression in Equation 28.4, which holds for both puts and calls, to obtain the same differential equation for call and put prices with respect to the variables K and T

$$\begin{aligned}\frac{\partial V}{\partial T} &= \frac{\partial \ln B_r(t,T)}{\partial T} V + \frac{1}{2}\sigma\left(K,T\right)^2 K^2 \frac{\partial^2 V}{\partial K^2} \\ &+ \frac{1}{T-t}\ln\left(\frac{B_q(t,T)}{B_r(t,T)}\right)\left(V - K\frac{\partial V}{\partial K}\right)\end{aligned} \tag{28.7}$$

For continuous compounding this simplifies further to

$$\frac{\partial V}{\partial T} = \frac{1}{2}\sigma\left(K,T\right)^2 K^2 \frac{\partial^2 V}{\partial K^2} - (r-q)K\frac{\partial V}{\partial K} - qV$$

This equation can be easily solved for $\sigma\left(K,T\right)^2$ thereby allowing implicit local volatilities to be calculated from the option prices observed in the market. As for the implicit transition probabilities described above, Equation 28.7 can be used (at time t for known values of S) to determine the value of $\sigma\left(K,T\right)$ for all points (K,T) for which the price of a call or put is observable in the market; an appropriate interpolation yields a *local volatility surface*. Of course, we need more than just the option price to complete this construction; the derivatives with respect to the strike price and the maturity must be known as well. Here we *always* must rely on the *numerical* determination of these derivatives. This is because σ (as opposed to p) itself appears in any analytical pricing formula (as for example in the Black-Scholes formula) and substitution of such price formulas into the above equation thus ultimately just yields the trivial[4] identity "$\sigma = \sigma$". As mentioned before, such numerical computations of (second!) derivatives are prone to error.

The volatilities obtained through numerical differentiation of the option prices appearing in Equation 28.7 are particularly unstable for options that are either well in or well out of the money since the (especially error prone!) second derivative $\partial^2 V/\partial K^2$ is quite close to zero for these values of K. Since isolating $\sigma\left(K,T\right)^2$ involves dividing by this very small (and error prone) second derivative, extremely large errors can be expected for the resulting local volatility. We can avoid at least this problem (i.e. the problem of dividing by a very small variable) by expressing the option prices in terms of their implied Black-Scholes volatilities $\widehat{\sigma}$ (which are just a different quotation of option prices). We then obtain an expression for the local volatility with a denominator of the form $1+\varepsilon$ with $\varepsilon > 0$. For example, for continuous compounding and with x as given in Equation 9.5

$$\sigma\left(K,T\right)^2 = \frac{\widehat{\sigma} + 2(T-t)\widehat{\sigma}\frac{\partial\widehat{\sigma}}{\partial T} + 2rK(T-t)\widehat{\sigma}\frac{\partial\widehat{\sigma}}{\partial K}}{\left[1 + Kx\sqrt{T-t}\frac{\partial\widehat{\sigma}}{\partial K}\right]^2 + K^2(T-t)\widehat{\sigma}\left[\frac{\partial^2\widehat{\sigma}}{\partial K^2} - x(\frac{\partial\widehat{\sigma}}{\partial K})^2\sqrt{T-t}\right]}$$

[4] This can be shown quite easily for continuous compounding with the help of Equations 9.6, 28.2 and 28.3. The one additional differentiation needed is

$$\begin{aligned}\frac{\partial V_{\text{call}}}{\partial T} &= -qe^{-q(T-t)}SN(x) + re^{-r(T-t)}KN(x-\sigma\sqrt{T-t}) \\ &+ e^{-r(T-t)}\,\frac{\sigma\left(K,T\right)}{2\sqrt{T-t}}K\,N'(x-\sigma\sqrt{T-t})\end{aligned}$$

But even in this form the local volatility remains very instable as long as the derivatives are approximated numerically from market data[5]. Practitioners usually attempt to smooth the Black-Scholes volatility surface by using one of several approaches. One example is to fit an analytic function (for example, a polynomial) $u(K,T)$ to the observed Black-Scholes volatilities. The derivatives of this function can then be calculated analytically when constructing the *Dupire local volatility surface.*

Having constructed the local volatility surface, the valuation of further options (OTCs) can be accomplished with a finite difference scheme, for example, as described in Section 10. The Excel workbook FiniteDifferenceMethod.xls accompanying in this book is conceived so that it is immediately obvious at what stage a volatility surface enters into the computation.

28.3 Volatility Transformations

28.3.1 Transformation between Relative and Absolute Volatility

The volatility of a risk factor with a price $S(t)$ at time t is usually measured as the standard deviation of the *logarithmic* price changes $\delta \ln S(t)$ (e.g. implicitly extracted from the option prices as discussed in Sections 28.1 and 28.2 or else through the analysis of historical time series as in Section 29) or obtained from data providers. For the *absolute* risk, in other words, for the actual monetary amount of a potential loss, the change of the price itself is needed and not its logarithm. Furthermore, the functional relationships (the sensitivities, for instance) between the price of a financial instrument and the risk factor S are usually given as functions of S and not in terms of $\ln S$. Therefore, there are many reasons for transforming the volatility of logarithmic changes into one of absolute changes $\delta S(t)$. In linear approximation, this is quite simply accomplished through[6]

$$\begin{aligned}
\delta \ln S(t) \equiv \ln S(t+\delta t) - \ln S(t) &= \ln\left(\frac{S(t+\delta t)}{S(t)}\right) \\
&= \ln\left(\frac{S(t) + \overbrace{[S(t+\delta t) - S(t)]}^{\delta S(t)}}{S(t)}\right) \\
&= \ln\left(1 + \frac{\delta S(t)}{S(t)}\right) = \frac{\delta S(t)}{S(t)} - \frac{1}{2}\left(\frac{\delta S(t)}{S(t)}\right)^2 \pm \cdots \\
&\approx \frac{\delta S(t)}{S(t)}
\end{aligned}$$

[5]Such a numerical approximation is done by approximating the differential quotient with a difference quotient, i.e. by $\frac{\partial V}{\partial K} \approx \frac{V_1(K_1) - V_2(K_2)}{K_1 - K_2}$, for instance, where V_1 and V_2 are two observed market prices, etc. An analogous procedure is performed for the Black-Scholes volatilities.

[6]The following expansion of the logarithm function is used: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \pm \cdots$

and hence,

$$\delta S(t) \approx S(t)\, \delta \ln S(t) \tag{28.8}$$

Thus, in linear approximation, *relative* (i.e. percentage) price changes are equal to *logarithmic* price changes

$$\frac{S(t+\delta t) - S(t)}{S(t)} \equiv \frac{\delta S(t)}{S(t)} \approx \delta \ln S(t) \tag{28.9}$$

where the approximation becomes exact for $\delta t \to 0$. This, incidentally, is exactly the point which provides an intuitive picture of the logarithm: The logarithmic change over a time span is nothing other than the corresponding relative change if this time span is infinitesimally short (see also Section 3.3.2).

According to Equation 28.8, the absolute price change follows from the logarithmic change multiplied by the current market price $S(t)$, called the *current level*. This is obviously consistent with the change in a risk factor over a finite time interval as given by Ito's lemma (see Equations 3.25 and 3.26) as can be seen by Taylor-expanding the exponential function appearing there for small δt. We have seen this approximation several times already in this book, see for example Equation 21.16.

Note that $S(t)$ is known at time t and as such is not a random variable. On the right-hand side of Equation 28.8, only $\delta \ln(S(t))$ is a random variable. Therefore, the variance of the absolute price change is simply

$$\begin{aligned} \operatorname{var}[\delta S(t)] &\approx \operatorname{var}\left[\underbrace{S(t)}_{\text{Known}} \quad \underbrace{\delta \ln S(t)}_{\text{Random Variable}} \right] \\ &= S(t)^2 \operatorname{var}[\delta \ln S(t)] \\ &= S(t)^2 \sigma^2(t) \delta t \end{aligned} \tag{28.10}$$

where in the last step the variance of the logarithmic price changes over a time span δt in accordance with the definition in Equation 3.11 was expressed in terms of the volatility σ (in %) over this time span.

28.3.2 Summation of Volatilities

A sum (a portfolio) of *multiple* risk factors S_i, containing N_i of each of the risk factors S_i has a value given by

$$V(t) \equiv \sum_{i=1} N_i S_i(t)$$

The volatility σ_V of the value V of this portfolio is defined exactly as was done for individual risk factors as the standard deviation[7] of the *logarithmic* value changes of

[7] More precisely: the standard deviation per square root of the time over which the change occurs.

the portfolio. Equation 28.8 can be applied to approximate the logarithmic changes of V as well as of the individual risk factors, yielding

$$V(t)\,\delta\ln(V(t)) \approx \delta V(t) = \delta\left(\sum_{i=1}^{n} N_i S_i(t)\right) = \sum_{i=1}^{n} N_i \delta S_i(t)$$
$$\approx \sum_{i=1}^{n} N_i S_i(t)\,\delta\ln(S_i(t))$$

and hence,

$$\delta\ln\left(V(t)\right) \approx \sum_{i=1}^{n} \underbrace{\left(\frac{N_i S_i(t)}{V(t)}\right)}_{\text{Weight}} \underbrace{\delta\ln\left(S_i(t)\right)}_{\text{Random Variable}} \tag{28.11}$$

The logarithmic change of the sum of risk factors is thus the weighted sum over the logarithmic changes of each of the individual risk factors. The weight of each individual risk factor is equal to the contribution (in %) of the current value of the pertinent risk factor to the total current value of the portfolio. It is thus a *present value weighting,* similar to that appearing in the *duration* in Equation 6.8.

Each of the current prices $S_i(t)$ are known at time t and as such are not random variables. The logarithmic changes of the individual risk factors, on the other hand, are random variables. From Equation A.17, we know that the *variance* of a sum of random variables is equal to the sum of the *covariances* of these random variables:

$$\begin{aligned}\operatorname{var}\left[\delta\ln\left(V\right)\right] &\approx \frac{1}{V^2}\operatorname{var}\left[\sum_{i=1}^{n} N_i S_i\,\delta\ln\left(S_i\right)\right] \\ &= \frac{1}{V^2}\sum_{i,j=1}^{n} N_i N_j\,S_i\,S_j\operatorname{cov}\left[\delta\ln\left(S_i\right),\delta\ln\left(S_j\right)\right]\end{aligned} \tag{28.12}$$

We now express the variances of the logarithmic price changes over a time span of length δt again as in Equation 3.11 in terms of the corresponding volatilities and use Equation A.14 for the covariance:

$$\operatorname{var}\left[\delta\ln\left(V\right)\right] = \sigma_V^2\delta t,\ \operatorname{var}\left[\delta\ln\left(S_i\right)\right] = \sigma_i^2\delta t$$
$$\operatorname{cov}\left[\delta\ln\left(S_i\right),\delta\ln\left(S_j\right)\right] = \rho_{i,j}\,\sigma_i\sqrt{\delta t}\,\sigma_j\sqrt{\delta t}$$

Thus, we finally obtain the desired relation between the volatility σ_V of the sum of risk factors and the volatilities σ_i of the individual risk factors:

$$\sigma_V^2 \approx \sum_{i,j=1}^{n} \underbrace{\left(\frac{N_i S_i(t)}{V(t)}\right)}_{\text{Weight } i}\sigma_i\,\rho_{i,j}\,\sigma_j\underbrace{\left(\frac{N_j S_j(t)}{V(t)}\right)}_{\text{Weight } j} \tag{28.13}$$

28.3.3 Transformation between Yield and Price Volatility

As was discussed in detail in Section 22.2, the same holds (in linear approximation) for a portfolio with sensitivities $\Delta_i = \partial V/\partial S_i$ with respect to the risk factor S_i with the obvious association $\Delta_i = N_i$. This substitution leads from Equation 28.13 immediately to Equation 22.9.

Specifically, a change in the price V of a financial instrument with a sensitivity Δ_i with respect to a single risk factor S_i is in linear approximation given by

$$\delta V(t) \approx \Delta_i(t)\, \delta S_i(t) \approx \Delta_i(t)\, S_i(t)\, \delta \ln\left(S_i(t)\right) \tag{28.14}$$

The first "approximately equal" sign is from the delta-normal approximation for the price of the financial instrument as a function of the risk factor. The second is a result of the linear approximation 28.8 for the logarithm of the risk factor change. The linear approximation of the logarithmic change in V is likewise equal to the relative change $\delta V/V$. Dividing Equation 28.14 by V yields the relationship between the *logarithmic* change of V and that of S:

$$\delta \ln\left(V(t)\right) \approx \frac{\delta V(t)}{V(t)} \approx \frac{\Delta_i(t)}{V(t)}\, S_i(t)\, \delta \ln\left(S_i(t)\right) \tag{28.15}$$

At time t, all of the variables appearing in the equation are known except for the changes δV and $\delta \ln(S)$. The variance of this equation is thus

$$\mathrm{var}\left[\delta \ln\left(V(t)\right)\right] \approx \frac{\Delta(t)^2}{V(t)^2}\, S_i(t)^2 \mathrm{var}\left[\delta \ln\left(S_i(t)\right)\right]$$

Expressing the variances through the volatilities as in Equation 3.11, we obtain the relationship between the volatility σ_V of the price of a financial instrument induced by the volatility σ_i of a risk factor and the volatility of the risk factor itself:

$$\sigma_V(t) = \left|\frac{\Delta_i(t)}{V(t)}\, S_i(t)\right| \sigma_i(t) \tag{28.16}$$

The absolute value sign indicates explicitly that the volatility, being the square root of the variance, is always positive.

The correlation of *another* risk factor S_j with this financial instrument (also dependent on S_i) is the same as its correlation with the risk factor S_i itself since

$$\begin{aligned}
&\sigma_V\, \sigma_j\, \rho_{V\,j}\, \delta t \\
&= \mathrm{cov}\left[\delta \ln\left(V\right),\ \delta \ln\left(S_j\right)\right] = \mathrm{cov}\left[\frac{\Delta_i}{V} S_i\, \delta \ln\left(S_i\right),\ \delta \ln\left(S_j\right)\right] \\
&= \frac{\Delta_i}{V} S_i\, \mathrm{cov}\left[\delta \ln\left(S_i\right),\ \delta \ln\left(S_j\right)\right] = \underbrace{\frac{\Delta_i}{V} S_i\, \sigma_i}_{\sigma_V}\, \sigma_j\, \rho_{ij}\, \delta t \\
&= \sigma_V\, \sigma_j\, \rho_{ij}\, \delta t
\end{aligned}$$

and hence

$$\rho_{V\,j} = \rho_{ij} \tag{28.17}$$

The transformation between the yield and price volatility is a special case of Equation 28.16. Data services provide both the volatilities σ_R for interest rates $R(t,T)$, called *yield volatilities*, as well as volatilities σ_B for the prices (*price volatilities*) of the corresponding zero bonds $B_R(t,T)$ for the various terms $T-t$. The transformation between these two volatilities is accomplished with Equation 28.16, where the relative sensitivity Δ/V for bonds is, as is well known, the (negative) *modified duration*, Equation 6.10. The transformation for zero bonds specifically is thus

$$\begin{aligned}\sigma_B(t,T) &= D_{\text{mod}} R(t,T)\,\sigma_R(t,T) \\ &= \begin{cases} \frac{R(t,T)(T-t)}{1+R(t,T)}\sigma_R(t,T) & \text{Discrete Compounding} \\ \frac{R(t,T)(T-t)}{1+R(t,T)(T-t)}\sigma_R(t,T) & \text{Simple Compounding} \end{cases}\end{aligned} \tag{28.18}$$

where Table 16.1 in Section 16.1 provides the explicit form for the two most frequently used compounding methods.

28.3.4 Currency Transformation of Volatilities and Correlations

The General Transformations for all Risk Factors

Transformation of the Volatility For prices (as opposed to interest rates), the value of the risk factor (the price) is always quoted with reference to a currency, even when quoted in %. A German government bond quoted at, say, 98% has a value of 98% of its face value in euros; the Nikkei index has a value corresponding to Japanese yen (JPY). Volatilities and correlations of a price depend heavily on the currency in which the price is quoted. This is because a price quoted in a particular currency which is not the desired *reference currency* or *base currency* not only entails the risk of value changes in the quoted currency but also the additional uncertainty of the exchange rate with the reference currency. It will be shown that the conversion of a price volatility into another currency is similar (but not exactly the same!) as the determination of the volatility of a portfolio with two risk factors; the first being the price in the quoted currency and the second the risk factor of the exchange rate with the reference currency. It will prove to be somewhat more difficult to transform the correlations. The explicit transformations for the volatilities and the correlations will be given below.

Let S be the price of a risk factor in a ("foreign") currency other than the reference currency and D the exchange rate with respect to the reference currency (i.e. D units of the reference currency are received for one unit of the foreign currency). The value of the risk factor expressed in terms of the reference currency is simply the product DS of the exchange rate and S. We now calculate the variance of the logarithmic change of this value in the reference currency. We make use of the property of the logarithm which turns products into sums so that the variance of the resulting sum of random

variables can be expressed as the sum of the covariances as in Equation A.17:

$$\begin{aligned}
&\text{var}\left[\delta \ln\left(D(t)S(t)\right)\right] \\
&= \text{var}\left[\underbrace{\delta \ln(D(t))}_{\text{Random Variable } X_1} + \underbrace{\delta \ln\left(S(t)\right)}_{\text{Random Variable } X_2}\right] = \sum_{a,b=1}^{2} \text{cov}\left[X_a, X_b\right] \\
&= \text{var}\left[\delta \ln\left(D(t)\right)\right] + \text{var}\left[\delta \ln\left(S(t)\right)\right] + 2\text{cov}\left[\delta \ln\left(D(t)\right), \delta \ln\left(S(t)\right)\right]
\end{aligned}$$

With these variances, it follows immediately from Equation 3.11 that the volatility of the product DS is given by

$$\sigma_{DS} = \sqrt{\sigma_D^2 + \sigma_S^2 + 2\rho_{S,D}\,\sigma_D\,\sigma_S} \tag{28.19}$$

This is the volatility transformation: on the left-hand side, we have the desired volatility in the reference currency, on the right, the volatility with respect to the foreign currency, the volatility of the exchange rate and the correlation between them.

While Equation 28.13 gives the volatility of a *sum* of risk factors, Equation 28.19 is the volatility of the *product* of risk factors where the exchange rate D and the price S each represent a risk factor.

Although the results are quite similar, there are two fundamental differences:

- In contrast to Equation 28.13, no approximation of the logarithm was made in the derivation of Equation 28.19. It is therefore an exact equality.
- In Equation 28.19, only the volatilities and correlations appear and not the current levels of the risk factors D and S.

Let us consider the following example: a portfolio composed of a British bond with a face value of $N =$GBP 100,000 whose price and volatility is quoted in British Pounds (GBP). The risk of the portfolio *expressed in GBP* is simply the product of the given daily volatility $\sigma = 0.026\%$ and the current level 99.704% of the face value N (multiplied by the square root of the liquidation period and the percentile for the confidence level c):

$$\begin{aligned}
\text{VaR}(c,t,T) &= Q_{1-c}\sqrt{T-t}\; S\, N\, \sigma_S \\
&= Q_{1-c}\sqrt{T-t} \times 97.704\% \times N \times 0.00026 \\
&= Q_{1-c}\sqrt{T-t} \times 25.38\,\text{GBP}
\end{aligned}$$

The risk is now to be converted into USD. The transformation in Equation 28.19 intuitively means that a further risk factor must be taken into consideration, the USD/GBP exchange rate D. This value and its daily volatility are: $D = 1.469$ USD/GBP and $\sigma_D = 0.404\%$. The correlation between the two risk factors is $\rho_{S,D} = -0.11574$. The value at risk in USD (with respect to a confidence level of c and a liquidation period $T-t$) is essentially the product of the current portfolio value in USD and the portfolio volatility (with respect to USD). The portfolio's value is just the product (and not the

sum; herein lies the reason for the difference between this and Equation 28.13) of the two risk factors S and D. The portfolio volatility with respect to USD is given by the transformation in Equation 28.19. Thus, the value at risk in USD becomes

$$\begin{aligned}
\mathrm{VaR}(P_a, t, T) &= Q_{1-c}\sqrt{T-t}D(t)S(t)N\sigma_{SD} \\
&= Q_{1-c}\sqrt{T-t}D(t)S(t)N\sqrt{\sigma_D^2 + \sigma_S^2 + 2\rho_{S,D}\,\sigma_D\,\sigma_S} \\
&= Q_{1-c}\sqrt{T-t} \times 97,704\,\mathrm{GBP} \times 1.469\frac{\mathrm{USD}}{\mathrm{GBP}} \\
&\times \sqrt{0.00404^2 + 0.00026^2 - 2 \times 0.11574 \times 0.00404 \times 0.00026} \\
&= Q_{1-c}\sqrt{T-t} \times 577.39\,\mathrm{USD}
\end{aligned}$$

This is very different from the (wrong) result obtained had we (incorrectly) converted the GBP-based value at risk into the value at risk in USD by simply multiplying it by the current GBP/USD exchange rate. This would have resulted in a completely incorrect VaR of "25.38 GBP $\times$1.469 USD/GBP = 37.28 USD". Thus, the risk of the position mainly stems from the fact that it is a position in a foreign currency (as seen from USD). The specific price risk of the bond with respect to its home currency only plays a minor role.

Transformation of the Correlation Having presented the transformation of the volatility in detail, we now consider the correlations. The transformation for the correlations are somewhat more complex since two risk factors are involved, these being the two factors to which the correlation refers. We must differentiate between two cases: a risk factor is given in the reference currency and the other in the "foreign" currency, or both risk factors are quoted in the foreign currency.

Case 1 The risk factor with price S_1 is given in the foreign currency, the second risk factor S_2 in the reference currency. The foreign currency has the exchange rate D with respect to the reference currency (D units of the reference currency are received for one unit in the foreign currency). The price of the first risk factor in the reference currency is therefore equal to DS_1. We now consider the covariance of the logarithmic value changes of these prices with those of the second risk factor. Exploiting the linearity of the covariance (Equation A.13), it is easy to see that

$$\begin{aligned}
&\mathrm{cov}\left[\delta \ln(D\,S_1),\ \delta \ln(S_2)\right] \\
&= \mathrm{cov}\left[\delta \ln(D) + \delta \ln(S_1),\ \delta \ln(S_2)\right] \\
&= \mathrm{cov}\left[\delta \ln(D),\ \delta \ln(S_2)\right] + \mathrm{cov}\left[\delta \ln(S_1),\ \delta \ln(S_2)\right]
\end{aligned}$$

With Equation A.14, the covariances can be expressed in terms of the volatilities and the correlations:

$$\rho_{DS_1,\,S_2}\sigma_{DS_1}\,\sigma_{S_2} = \rho_{D,\,S_2}\,\sigma_D\,\sigma_{S_2} + \rho_{S_1,\,S_2}\,\sigma_{S_1}\,\sigma_{S_2}$$

Applying Equation 28.19 for the volatility of the product DS_1 finally yields the desired transformation

$$\rho_{DS_1,\,S_2} = \frac{\rho_{D,\,S_2}\,\sigma_D\,\sigma_{S_2} + \rho_{S_1,\,S_2}\,\sigma_{S_1}\,\sigma_{S_2}}{\sigma_{S_2}\sqrt{\sigma_D^2 + \sigma_{S_1}^2 + 2\,\rho_{D,\,S_1}\,\sigma_D\,\sigma_{S_1}}} \tag{28.20}$$

Case 2 The risk factor with a price S_1 is quoted in a foreign currency, the second risk factor S_2 is likewise given in a foreign currency. These two foreign currencies may be different. The foreign currency associated with S_1 has exchange rate D_1 with respect to the reference currency, the second foreign currency has exchange rate D_2. Then the price of the first risk factor in the reference currency equals D_1S_1. Correspondingly the price of the second risk factor is D_2S_2. As before, we consider the covariance of the logarithmic value changes of the prices

$$\begin{aligned}
&\operatorname{cov}\left[\delta \ln\left(D_1S_1\right), \delta \ln\left(D_2S_2\right)\right] \\
&= \operatorname{cov}\left[\delta \ln(D_1) + \delta \ln\left(S_1\right), \delta \ln(D_2) + \delta \ln\left(S_2\right)\right] \\
&= \operatorname{cov}\left[\delta \ln(D_1), \delta \ln(D_2)\right] + \operatorname{cov}[\delta \ln\left(S_1\right), +\delta \ln\left(S_2\right)] \\
&+ \operatorname{cov}\left[\delta \ln(D_1), \delta \ln\left(S_2\right)\right] + \operatorname{cov}[\delta \ln\left(S_1\right), \delta \ln(D_2)]
\end{aligned}$$

Applying Equation A.14, the covariances again can be expressed in terms of the volatilities and the correlations:

$$\begin{aligned}
\rho_{D_1S_1,\,D_2S_2}\sigma_{D_1S_1}\,\sigma_{D_2S_2} = {} & \rho_{D_1,\,D_2}\,\sigma_{D_1}\,\sigma_{D_2} + \rho_{D_1,\,S_2}\,\sigma_{D_1}\,\sigma_{S_2} \\
& + \rho_{D_2,\,S_1}\,\sigma_{D_2}\,\sigma_{S_1} + \rho_{S_1,\,S_2}\,\sigma_{S_1}\,\sigma_{S_2}
\end{aligned}$$

Using Equation 28.19 for the volatilities of the products D_1S_1 and D_2S_2 finally yields the desired transformation

$$\begin{aligned}
&\rho_{D_1S_1,\,D_2S_2} \\
&= \frac{\rho_{D_1,\,D_2}\sigma_{D_1}\,\sigma_{D_2} + \rho_{D_1,\,S_2}\sigma_{D_1}\,\sigma_{S_2} + \rho_{D_2,\,S_1}\sigma_{D_2}\,\sigma_{S_1} + \rho_{S_1,\,S_2}\sigma_{S_1}\,\sigma_{S_2}}{\sqrt{\sigma_{D_1}^2 + \sigma_{S_1}^2 + 2\,\rho_{D_1,\,S_1}\,\sigma_{D_1}\,\sigma_{S_1}}\,\sqrt{\sigma_{D_2}^2 + \sigma_{S_2}^2 + 2\,\rho_{D_2,\,S_2}\,\sigma_{D_2}\,\sigma_{S_2}}}
\end{aligned} \tag{28.21}$$

This is the correlation with respect to the reference currency between both of the risk factors S_1 and S_2 quoted in their respective currencies.

We are frequently confronted with the special case of this equation in which the two foreign currencies are the same, implying $D_1 = D_2 = D$. In this case the transformation reduces to

$$\rho_{DS_1,\,DS_2} = \frac{\sigma_D^2 + \left(\rho_{D,\,S_2}\sigma_{S_2} + \rho_{D,\,S_1}\sigma_{S_1}\right)\sigma_D + \rho_{S_1,\,S_2}\sigma_{S_1}\,\sigma_{S_2}}{\sqrt{\sigma_D^2 + \sigma_{S_1}^2 + 2\,\rho_{D,\,S_1}\,\sigma_D\,\sigma_{S_1}}\,\sqrt{\sigma_D^2 + \sigma_{S_2}^2 + 2\,\rho_{D,\,S_2}\,\sigma_D\,\sigma_{S_2}}} \tag{28.22}$$

Transformations for the Exchange Rates and Cross Rates

The transformations described above must often be applied in practice. All the exchange rates quoted in the data sets of a data provider like RiskMetrics$^{\text{TM}}$, for example,

are usually stated by reference to a base currency, e.g. USD. An investor whose reference currency is, for example, EUR must therefore convert the listed exchange rate volatilities and correlations from USD to EUR as the *base currency* before they can be used for risk management. But not only are the risk factors to be converted, a transformation must also be performed from one base currency (USD) into another (EUR). Here, *cross rates* come into play. For a USD-based investor, the Yen/EUR exchange rate, for example, is a cross rate. Cross rates are obtained by dividing the desired rate (in this case USD/JPY) by the exchange rate of the new base currency (for example USD/EUR). To illustrate this concept further, consider the following examples of exchange rates with respect to the USD

$$D_i(t) = \frac{USD}{EUR}\,, \quad D_j(t) = \frac{USD}{SFR}$$
$$D_n(t) = \frac{USD}{JPY}\,, \quad D_m(t) = \frac{USD}{SEK}$$

From these, cross rates for arbitrary combinations can be established as follows:

$$D_{ji}(t) = \frac{D_j(t)}{D_i(t)} = \frac{EUR}{SFR}$$
$$D_{ni}(t) = \frac{D_n(t)}{D_i(t)} = \frac{EUR}{JPY}$$
$$D_{jn}(t) = \frac{D_j(t)}{D_n(t)} = \frac{JPY}{SFR}\,, \text{ etc.}$$

Thus, the subtlety in transforming volatilities and correlations of cross rates lies in the fact that, as opposed to the previous section, not the products but the *quotients* of risk factors (exchange rates) must be considered. The quotient of two values a and b can, of course, be seen as the product of a and $1/b$. The arguments of the previous section can then be applied. Consider the volatilities and correlations of an "inverse risk factor" (exchange rate) $1/D$:

$$\begin{aligned}\text{var}\left[\delta\,\ln\left(\frac{1}{D(t)}\right)\right] &= \text{var}\left[-\delta\,\ln(D(t))\right] = (-1)^2\text{var}\left[\delta\,\ln(D(t))\right]\\ \text{cov}\left[\delta\,\ln\left(\frac{1}{D(t)}\right), \delta\,\ln(S(t))\right] &= \text{cov}\left[-\delta\,\ln(D(t)), \delta\,\ln(S(t))\right]\\ &= -\text{cov}\left[\delta\,\ln(D(t)), \delta\,\ln(S(t))\right]\end{aligned}$$

This implies that $1/D$ has the same volatility as D, and that the correlations of $1/D$ with an arbitrary risk factor S is just the negative correlation of D with S (the correlation between two inverse prices is then the same as the correlation between the prices themselves):

$$\sigma_{\frac{1}{D}} = \sigma_D \quad , \quad \rho_{\frac{1}{D},S} = -\rho_{D,S}\,, \quad \rho_{\frac{1}{D},\frac{1}{S}} = \rho_{D,S} \tag{28.23}$$

The transformations of the previous section can now be directly applied. If σ_i denotes the volatility of D_i, σ_{ji} the volatility of the cross rate $D_{ji} = D_j/D_i$, etc., the transformations for the exchange rates are given explicitly by:

Price Volatilities and Correlations in Original Currency						
	USD/EUR	USD/JPY	USD/GBP	JPY.Z05	JPY.Z07	GBP.R180
USD/EUR	**0.680%**	0.49105	0.76692	0.13130	0.03701	-0.01863
USD/JPY	0.49105	**0.731%**	0.51365	-0.02040	-0.02092	-0.07130
USD/GBP	0.76692	0.51365	**0.404%**	0.12458	0.10580	-0.11574
JPY.Z05	0.13130	-0.02040	0.12458	**0.093%**	0.89410	0.17171
JPY.Z07	0.03701	-0.02092	0.10580	0.89410	**0.177%**	0.00987
GBP.R180	-0.01863	-0.07130	-0.11574	0.17171	0.00987	**0.026%**

Figure 28.1: The original price volatilities and correlations. The data are for relative daily changes.

- From Equation 28.19 (with the substitution $D = D_j$, $S = 1/D_i$), the volatility of a cross rate is:

$$\sigma_{ji} = \sqrt{\sigma_j^2 + \sigma_i^2 - 2\rho_{ji}\,\sigma_j\,\sigma_i} \tag{28.24}$$

- From Equation 28.20 (with the substitution $D = D_j$, $S_1 = 1/D_i$, $S_1 = S$), the correlation of a cross rate with an arbitrary risk factor S is:

$$\rho_{ji,S} = \frac{\rho_{j,S}\sigma_j\,\sigma_S - \rho_{i,S}\sigma_i\,\sigma_S}{\sigma_S\,\sqrt{\sigma_j^2 + \sigma_i^2 - 2\,\rho_{j,i}\,\sigma_j\,\sigma_i}} \tag{28.25}$$

- From Equation 28.21 (with the substitution $D_1 = D_j$, $D_2 = D_n$, $S_1 = 1/D_i$, $S_2 = 1/D_m$), the correlation between two cross rates $D_{ji} = D_j/D_i$ and $D_{nm} = D_n/D_m$ is:

$$\rho_{ji,nm} = \frac{\rho_{j,n}\sigma_j\,\sigma_n - \rho_{j,m}\sigma_j\,\sigma_m - \rho_{i,n}\sigma_i\,\sigma_n + \rho_{i,m}\sigma_i\,\sigma_m}{\sqrt{\sigma_j^2 + \sigma_i^2 - 2\,\rho_{j,i}\,\sigma_j\,\sigma_i}\;\sqrt{\sigma_n^2 + \sigma_m^2 - 2\,\rho_{n,m}\,\sigma_n\,\sigma_m}} \tag{28.26}$$

- Similar to Equation 28.22, the correlation of two cross rates with respect to the same base currency can be obtained by setting $m = i$:

$$\rho_{ji,ni} = \frac{\rho_{j,n}\sigma_j\,\sigma_n - (\rho_{j,i}\sigma_j\; + \rho_{i,n}\,\sigma_n)\,\sigma_i + \sigma_i^2}{\sqrt{\sigma_j^2 + \sigma_i^2 - 2\,\rho_{j,i}\,\sigma_j\,\sigma_i}\;\sqrt{\sigma_n^2 + \sigma_i^2 - 2\,\rho_{n,i}\,\sigma_n\,\sigma_i}} \tag{28.27}$$

In the accompanying Excel workbook ValueAtRisk.xls, applications of the transformations introduced above are demonstrated explicitly. The original data as it would be obtained from a data provider is shown in Figure 28.1. The transformations of these date into EUR-based data proceeds in three steps:

1. The volatilities and correlations of the cross rates are determined in accordance with Equations 28.24 and 28.26.

Price Volatilities and Correlations in EUR					
	EUR/JPY	EUR/GBP	JPY.Z05	JPY.Z07	GBP.R180
EUR/JPY	**0.713%**	0.49406	-0.14609	-0.05673	-0.05533
EUR/GBP	0.49406	**0.452%**	-0.08604	0.03906	-0.07564
JPY.Z05	-0.14609	-0.08604	**0.732%**	0.76434	0.33993
JPY.Z07	-0.05673	0.03906	0.76434	**0.744%**	0.45638
GBP.R180	-0.05533	-0.07564	0.33993	0.45638	**0.454%**

Figure 28.2: The volatilities and correlations with respect to EUR.

Price Volatilities and Correlations in USD						
	USD/EUR	USD/JPY	USD/GBP	JPY.Z05	JPY.Z07	GBP.R180
USD/EUR	**0.680%**	0.49105	0.76692	0.50501	0.48829	0.76987
USD/JPY	0.49105	**0.731%**	0.51365	0.99192	0.97158	0.51183
USD/GBP	0.76692	0.51365	**0.404%**	0.52664	0.52665	0.99794
JPY.Z05	0.50501	0.99192	0.52664	**0.735%**	0.99058	0.52632
JPY.Z07	0.48829	0.97158	0.52665	0.99058	**0.748%**	0.52515
GBP.R180	0.76987	0.51183	0.99794	0.52632	0.52515	**0.402%**

Figure 28.3: The volatilities and correlations with respect to USD.

2. Subsequently, the correlations between these cross rates and the risk factors JPY.Z05 (Japanese 5-year zero bond), JPY.Z07 (Japanese 7-year zero bond) and GBP.R180 (British 6-month zero bond) are determined as indicated in Equation 28.25.

3. Then the correlations of the risk factors with one another are transformed via Equation 28.21 (or Equation 28.22 for the correlation between the two bonds). For this, we need the (just found) EUR-based correlations and volatilities of the cross rates and of the risk factors with the cross rates. Therefore, this step can only be completed *after* the first two steps have been performed (Equations 28.24, 28.26, 28.25).

The results of these transformations are presented in Figure 28.2 and were already used in Figure 25.3 of Section 25 for calculating the value at risk.

Note that the original data refer to *different* currencies: the exchange rates all refer to the USD, the Japanese bonds to JPY and the British bond to GBP. Thus, even if the base currency was USD, the volatilities and correlations of the Japanese and British bonds must be converted. This is accomplished using Equation 28.19 for the volatilities, Equation 28.20 for the correlations between the bonds and the exchange rates (for S_2 the exchange rate is to be substituted) and Equations 28.21 and Equation 28.22 for correlations of the bonds with one another. The corresponding results are presented in Figure 28.3.

29

Market Parameter from Historical Time Series

Having shown in the previous sections how statistical parameters such as the volatility can be obtained *implicitly* from the prices of derivatives traded in the market, we now proceed with what is perhaps the more natural approach, which builds directly on the definition of the statistical values, namely the analysis of historical time series. *Time series analysis* is a broad topic in the field of *statistics* whose application here will be limited to those areas which serve the purposes of this book. A much more general and wide-reaching presentation can be found in [72], for example.

29.1 Historical Yields, Volatility and Correlation

When analyzing both historical and simulated *time series* of a risk factor, the logarithm of the ratio between the new and old price is measured, which is in linear approximation the same as the *relative* price change, see Equation 28.9. From the expectation and the variance of this variable the *historical volatility* σ and the *historical mean return* μ can be determined by recording the relative changes over a period of length δt along a historical (or simulated) path:

$$\begin{aligned} \mu = \frac{1}{\delta t}\,\mathrm{E}\left[X\right] \quad , \quad \sigma^2 = \frac{1}{\delta t}\,\mathrm{var}\left[X\right] \\ \text{where} \quad X = \ln\left(\frac{S(t+\delta t)}{S(t)}\right) \approx \frac{S(t+\delta t) - S(t)}{S(t)} \end{aligned} \tag{29.1}$$

The *historical correlation* between two price processes is likewise determined from the relative changes. With Y denoting the change of the second price (analogous to X as defined above), the correlation is given by

$$\rho = \frac{\mathrm{var}\left[XY\right]}{\sqrt{\mathrm{var}\left[X\right]\mathrm{var}\left[Y\right]}}$$

These equations make it apparent that a procedure is required enabling the determination of the expectation and variance from historical (or simulated) data. It is well known from statistics that from n measurements X_i, which are realizations of a random variable X, the *mean* $\langle X \rangle$ (and more generally the mean $\langle f(X) \rangle$ of a function of X) can be computed in the following way

$$\langle X \rangle = \frac{1}{n} \sum_{i=1}^{n} X_i \quad , \quad \langle f(X) \rangle = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

The *law of large numbers* states that these means approximate the desired *expectations* and variances as follows[1]:

$$\begin{aligned} &\langle f(X) \rangle \overset{n \to \infty}{\longrightarrow} \mathrm{E}\left[f(X)\right] \\ &\left\langle f(X)^2 \right\rangle - \langle f(X) \rangle^2 \overset{n \to \infty}{\longrightarrow} \frac{n-1}{n} \operatorname{var}\left[f(X)\right] \end{aligned} \tag{29.2}$$

This result may seem trivial or the difference between the actual parameters (the right-hand side) and the measured approximation (left-hand side) may be unclear to the reader. This difference, however, is fundamental: the theoretical value (the right-hand side) can never be precisely known; it can only be *estimated* more or less exactly through the computation of means of measured data. Such "means of measured data" are called *estimators* in statistics to distinguish them from the true values. Examples of estimators are the left-hand sides in Equation 29.2 or the right-hand sides in Equations 29.5 and 29.9.

The expectation and variance of X are needed for the description of the risk factor X. For this, the means of the realizations of the random variables X and X^2 are needed. The determination of the *error* made in making these estimates requires X^4 as well (see Section 29.3). Any good analysis of times series of historical data (or Monte Carlo simulations) should therefore include the computation of at least the following means:

$$\langle X \rangle = \frac{1}{n} \sum_{i=1}^{n} X_i \ , \ \left\langle X^2 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \ , \ \left\langle X^4 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} X_i^4 \tag{29.3}$$

If multiple correlated prices are involved, the following means for each pair of prices are also necessary:

$$\langle XY \rangle = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i \ , \ \left\langle (XY)^2 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} X_i^2 Y_i^2 \ , \ \left\langle (XY)^4 \right\rangle = \frac{1}{n} \sum_{i=1}^{n} X_i^4 Y_i^4 \tag{29.4}$$

From these values, historical estimates for the mean return, volatility and correlation

[1] The factor $(n-1)/n$ of the variance is necessary if the estimator for the variance is to be *unbiased*. See any introductory statistics textbook for more on this subject.

can be obtained:

$$\mu = \frac{1}{\delta t}\,\mathrm{E}\,[X] \approx \frac{1}{\delta t}\,\langle X\rangle \tag{29.5}$$

$$\sigma = \frac{1}{\sqrt{\delta t}}\sqrt{\mathrm{var}\,[X]} \approx \frac{1}{\sqrt{\delta t}}\sqrt{\frac{n}{n-1}}\,\sqrt{\langle X^2\rangle - \langle X\rangle^2}$$

$$\rho = \frac{\mathrm{var}\,[XY]}{\sqrt{\mathrm{var}\,[X]}\,\sqrt{\mathrm{var}\,[Y]}} \approx \frac{\left\langle (XY)^2\right\rangle - \langle XY\rangle^2}{\sqrt{\langle X^2\rangle - \langle X\rangle^2}\,\sqrt{\langle Y^2\rangle - \langle Y\rangle^2}}$$

29.2 Autocorrelation and Autocovariance

Up to now, it has been tacitly assumed that the n measurements in a time series come from *independent* observations. Independence is an assumption which often cannot be made, in particular in time series analysis (for example, in the case of moving average methods, see Section 31.4). Daily changes in a moving 30-day price average, for example (the mean of prices observed over the previous 30 days is computed) will remain small from one day to the next since in each daily adjustment, only the oldest price is replaced by the most recently observed value, the other 29 prices in the average remain the same. The measurement of such a variable is highly correlated with the measurement made on the previous day. Such a correlation of one and the same variable *with itself* is called *autocorrelation*[2]. Just as *correlation* measures the statistical influence between two *different* random processes, *auto*correlation measures the statistical influence a process has *on itself*. The autocorrelation is defined by

$$\varrho(t,h) = \frac{\mathrm{cov}(X_{t+h}, X_t)}{\mathrm{cov}(X_t, X_t)} = \frac{\mathrm{cov}(X_{t+h}, X_t)}{\mathrm{var}(X_t)} \tag{29.6}$$

where the definition of the *autocovariance* is completely analogous to the definition of the covariance between two different random variables (see Equation A.10)

$$\begin{aligned}\mathrm{cov}(X_{t+h}, X_t) &= \mathrm{E}\,[(X_{t+h} - \mathrm{E}\,[X_{t+h}])\,(X_t - \mathrm{E}\,[X_t])] \\ &= \mathrm{E}\,[X_{t+h}X_t] - \mathrm{E}\,[X_{t+h}]\,E\,[X_t]\end{aligned} \tag{29.7}$$

We arrive at the final equality by merely taking the product in the previous expression and using the linearity of the expectation. The resulting equation corresponds to A.11. Definition 29.7 immediately establishes the relation between the autocovariance and the variance

$$\mathrm{cov}(X_t, X_t) = \mathrm{E}[(X_t - \mathrm{E}\,[X_t])(X_t - \mathrm{E}\,[X_t])] = \mathrm{var}(X_t)$$

which has already been used in establishing the last equality in Equation 29.6.

The autocorrelation in Equation 29.6 also corresponds to the ordinary correlation known from statistics. If the time series is stationary[3] (in particular, if the variance of

[2] Autocorrelations do not appear merely in certain measurement methods but are by definition inherent to non-Markov processes, i.e. for processes whose current value is influenced by past values. See Section 30.1 for more on this topic.

[3] This means intuitively that the parameters describing the time series are time independent, see Section 30.

the process remains constant), $\mathrm{var}(X_t) = \mathrm{var}(X_{t+h})$ holds and ρ can be written in a form corresponding to Equation A.14:

$$\varrho(h) = \frac{\mathrm{cov}(X_{t+h}, X_t)}{\sqrt{\mathrm{var}(X_{t+h})}\sqrt{\mathrm{var}(X_t)}}$$

This is just the definition of the correlation of the random variable X_{t+h} with the random variable X_t, irrespective of whether these random variables belong to the same time series.

29.3 Error Estimates

Estimating the error in the measured values is essential for evaluating their meaningfulness. It is therefore insufficient to provide a value as the result of making observations since the actual theoretical value is not obtained (and will never be obtained) from measured data. It is more appropriate to find an interval on the basis of observed data within which the theoretical value lies.

A claim such as "the historical price volatility is 20% per year" is not very meaningful if nothing is said about the error associated with such a claim, for example 0.5% or 50%. A valid statement on the other hand might be "the historical price volatility is $20\% \pm 3\%$ per year". This means that the actual volatility lies somewhere between 17% and 23% per year.

In this section, several simple methods for determining the statistical error are introduced and its calculation will be demonstrated explicitly for the volatility and correlation. We begin by assuming that data in a time series (referred to as *observed values* or *measurements*) are pairwise independent and in consequence uncorrelated. Finally, we will briefly show how to test whether measurements are independent and how to proceed in the case that they are not, i.e. how to account for autocorrelations. Naturally, the subject is quite technical. Error estimation is theoretically quite simple but lengthy. Nonetheless, anyone who wishes to conduct a serious analysis of historical or simulated data should understand and apply this material.

The determination of expectations is accomplished through calculating the mean of the observed values as in Equation 29.5. For a sufficiently large number n of measurements

$$\mathrm{E}[X] \approx \langle X \rangle \quad , \quad \mathrm{var}[X] \approx \frac{n}{n-1}\left(\langle X^2 \rangle - \langle X \rangle^2\right) \tag{29.9}$$

holds. The central question is: what is the error[4] involved in estimating these parameters as above? The n observations are realizations of the random variables X_i, $i = 1, ...n$. The mean is the weighted sum over these random variables X_i and as such is again a random variable. The statistical error will be defined as the *standard*

[4]There are two different types of error. The first is the *statistical error*, which is a consequence of the fact that only a finite number of measurements are taken. The second is the *systematic error*. These are errors arising from a fundamental mistake (for example, a programming error in a Monte Carlo simulation). The failure to decrease with an increasing number of measurements is characteristic of a systematic error (as opposed to the statistical error). Only the statistical error will be dealt with in this section.

deviation of this new random variable, or equivalently, the error is the square root of the variance of the mean. A fundamental result from the field of statistics is

$$\langle X\rangle = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \tag{29.10}$$
$$\mathrm{var}\left[\langle X\rangle\right] = \frac{1}{n}\mathrm{var}\left[X\right] \qquad \text{if} \quad \mathrm{E}\left[X_i X_j\right] = 0 \quad \text{for} \quad i \neq j$$

The variance of the *mean* of uncorrelated, identically distributed random variables is equal to the variance of the random variable itself divided by the number of observations. This result combined with the approximation in Equation 29.9 for the variance yields the statistical error (denoted below by the symbol δ) defined as the standard deviation of the mean

$$\delta\langle X\rangle \equiv \sqrt{\mathrm{var}\left[\langle X\rangle\right]} = \sqrt{\frac{1}{n}\mathrm{var}\left[X\right]} \approx \sqrt{\frac{\langle X^2\rangle - \langle X\rangle^2}{n-1}} \tag{29.11}$$

This holds in general for the mean of a function of the random variable X:

$$\delta\langle f(X)\rangle = \sqrt{\frac{1}{n}\,\mathrm{var}\left[f(X)\right]} \approx \sqrt{\frac{\langle f(X)^2\rangle - \langle f(X)\rangle^2}{n-1}}$$

The statistical error in the estimation of the mean return as defined in Equation 29.5 is thus[5]

$$\delta\mu \approx \frac{1}{\delta t\sqrt{n-1}}\sqrt{\langle X^2\rangle - \langle X\rangle^2} \tag{29.12}$$

The mean of a function is to be distinguished from the function of the mean if the function is not linear: $\langle f(X)\rangle \neq f\left(\langle X\rangle\right)$. Thus, the *error of a function* of an uncertain value z (in the case under discussion, $z = \langle X\rangle$) cannot be computed directly in general. A Taylor series expansion of the function can provide assistance in such cases. The *propagation of error* can be obtained from this Taylor series:

$$f = f\left(z\right) \quad \text{with} \quad z = z \pm \delta z \quad \Rightarrow \quad \delta f = \left|\delta z \frac{\partial f}{\partial z}\right| + \frac{1}{2}\left|(\delta z)^2 \frac{\partial^2 f}{\partial z^2}\right| + \cdots$$

The vertical lines in the above equation indicate the absolute value. This can be generalized for functions of multiple variables with associated errors. Such expressions quickly become very lengthy. Restricting our consideration to the first (linear) terms, we obtain what is known as the *quadratic error propagation.*

$$f = f\left(z_1, z_2, \ldots, z_k\right) \quad \text{with} \quad z_i = z_i \pm \delta z_i \quad \Rightarrow$$
$$\delta f \approx \left|\sum_{i=1}^{k} \delta z_i \frac{\partial f}{\partial z_i}\right| = \sqrt{\sum_{i=1}^{k}\left(\delta z_i \frac{\partial f}{\partial z_i}\right)^2} \tag{29.13}$$

[5]Here, the δ in δt denotes the length of a time interval betweem two data points in the time series and *not* the "error in t".

This law will be required in order to determine the error involved in measuring the variance since

$$\begin{aligned}\mathrm{var}\,[X] &\approx \frac{n}{n-1}\left(\langle X^2\rangle - \langle X\rangle^2\right) = f(z_1,z_2) \quad \text{with}\\ z_1 &= \langle X\rangle\,,\ z_2 = \langle X^2\rangle\,,\ f(z_1,z_2) = \frac{n}{n-1}\left(z_2 - z_1^2\right) \quad \Rightarrow\\ \frac{\partial f}{\partial z_1} &= -\frac{2n}{n-1}z_1\,,\quad \frac{\partial f}{\partial z_2} = \frac{n}{n-1}\end{aligned}$$

The error of f, calculated with quadratic error propagation, is then

$$\begin{aligned}\delta f &\approx \sqrt{\left(\delta z_1 \frac{\partial f}{\partial z_1}\right)^2 + \left(\delta z_2 \frac{\partial f}{\partial z_2}\right)^2}\\ &= \sqrt{\left(\delta z_1 \frac{2n}{n-1} z_1\right)^2 + \left(\delta z_2 \frac{n}{n-1}\right)^2}\\ &= \frac{n}{n-1}\sqrt{4z_1^2\left(\delta z_1\right)^2 + \left(\delta z_2\right)^2}\end{aligned}$$

Since z_1 and z_2 are means of X and $f(X) = X^2$ respectively, their errors are respectively,

$$\begin{aligned}z_1 = \langle X\rangle \ &\Rightarrow\ \delta z_1 = \delta\langle X\rangle \approx \frac{1}{\sqrt{n-1}}\sqrt{\langle X^2\rangle - \langle X\rangle^2}\\ z_2 = \langle X^2\rangle \ &\Rightarrow\ \delta z_2 = \delta\langle X^2\rangle \approx \frac{1}{\sqrt{n-1}}\sqrt{\langle X^4\rangle - \langle X^2\rangle^2}\end{aligned}$$

Substituting these into the expression for δf yields

$$\begin{aligned}\delta f &\approx \frac{1}{\sqrt{n-1}}\frac{n}{n-1}\sqrt{4\langle X\rangle^2\left(\langle X^2\rangle - \langle X\rangle^2\right) + \left(\langle X^4\rangle - \langle X^2\rangle^2\right)}\\ &= \frac{1}{\sqrt{n-1}}\frac{n}{n-1}\sqrt{\langle X^4\rangle - \langle X^2\rangle^2 + 4\langle X^2\rangle\langle X\rangle^2 - 4\langle X\rangle^4}\end{aligned}$$

This is the statistical error made in measuring the variance var$[X]$. In order to determine this error when analyzing historical time series, the means of X, X^2 and X^4 must be measured.

Likewise, the error of the volatility σ can be determined through the following deliberations

$$\begin{aligned}\sigma\,[X] \ &\equiv \frac{1}{\sqrt{\delta t}}\sqrt{\mathrm{var}\,[X]} \approx \frac{1}{\sqrt{\delta t}}\sqrt{\frac{n}{n-1}}\sqrt{\langle X^2\rangle - \langle X\rangle^2} = \frac{1}{\sqrt{\delta t}}g(z_1,z_2)\\ \text{where } z_1 = \langle X\rangle\ ,\quad z_2 &= \langle X^2\rangle\ ,\quad g(z_1,z_2) = \sqrt{\frac{n}{n-1}}\sqrt{z_2 - z_1^2} \quad \Rightarrow\\ \frac{\partial g}{\partial z_1} &= -\sqrt{\frac{n}{n-1}}\frac{z_1}{\sqrt{z_2 - z_1^2}}\ ,\quad \frac{\partial g}{\partial z_2} = \frac{1}{2}\sqrt{\frac{n}{n-1}}\frac{1}{\sqrt{z_2 - z_1^2}}\end{aligned}$$

An analogous calculation as above yields

$$\delta\,\sigma \approx \frac{1}{\sqrt{\delta t}}\frac{1}{\sqrt{n-1}}\sqrt{\frac{n}{n-1}}\,\frac{\sqrt{\langle X^4\rangle - \langle X^2\rangle^2 + 4\,\langle X^2\rangle\,\langle X\rangle^2 - 4\,\langle X\rangle^4}}{2\sqrt{\langle X^2\rangle - \langle X\rangle^2}} \tag{29.14}$$

for the error in the measured volatility. The expression for the error of the correlation between two prices X and Y is even longer as it is represented by a function of six means:

$$\rho[X,Y] \equiv \frac{\mathrm{var}[XY]}{\sqrt{\mathrm{var}[X]\mathrm{var}[Y]}} \approx \frac{\langle (XY)^2\rangle - \langle XY\rangle^2}{\sqrt{\langle X^2\rangle - \langle X\rangle^2}\sqrt{\langle Y^2\rangle - \langle Y\rangle^2}} = \rho(z_1,z_2,z_3,z_4,z_5,z_6)$$

$$\text{where}\quad z_1 = \langle XY\rangle,\quad z_2 = \langle (XY)^2\rangle,\quad z_3 = \langle X\rangle,\quad z_4 = \langle X^2\rangle,\quad z_5 = \langle Y\rangle,\quad z_6 = \langle Y^2\rangle$$

$$\Longrightarrow \frac{\partial\rho}{\partial z_1} = \frac{-2z_1}{\sqrt{z_4 - z_3^2}\sqrt{z_6 - z_5^2}},\quad \frac{\partial\rho}{\partial z_2} = \frac{1}{\sqrt{z_4 - z_3^2}\sqrt{z_6 - z_5^2}}$$

$$\frac{\partial\rho}{\partial z_3} = \rho\frac{z_3}{z_4 - z_3^2},\quad \frac{\partial\rho}{\partial z_4} = \frac{-\rho}{2}\frac{1}{z_4 - z_3^2},\quad \frac{\partial\rho}{\partial z_5} = \rho\frac{z_5}{z_6 - z_5^2},\quad \frac{\partial\rho}{\partial z_5} = \frac{-\rho}{2}\frac{1}{z_6 - z_5^2}$$

Starting with this expression, simple but tedious algebra leads to a somewhat lengthy expression for the error when measuring correlation:

$$\begin{aligned}\partial\rho &\approx \sqrt{\sum_{i=1}^{6}\left(\delta z_i\frac{\partial\, g}{\partial\, z_i}\right)^2} \\ &\approx \frac{1}{\sqrt{n-1}}\,\frac{\langle (XY)^2\rangle - \langle XY\rangle^2}{\sqrt{\langle X^2\rangle - \langle X\rangle^2}\sqrt{\langle Y^2\rangle - \langle Y\rangle^2}} \\ &\times \sqrt{\begin{array}{c}\frac{4\langle XY\rangle^2}{\langle (XY)^2\rangle - \langle XY\rangle^2} + \frac{\langle (XY)^4\rangle - \left\langle \langle XY\rangle^2\right\rangle^2}{\langle \langle (XY)^2\rangle - \langle XY\rangle^2\rangle^2} + \frac{\langle X\rangle^2}{\langle X^2\rangle - \langle X\rangle^2} \\ + \frac{\langle X^4\rangle - \left\langle X^2\right\rangle^2}{4(\langle X^2\rangle - \langle X\rangle^2)^2} + \frac{\langle Y\rangle^2}{\langle Y^2\rangle - \langle Y\rangle^2} + \frac{\langle Y^4\rangle - \langle Y^2\rangle^2}{4(\langle Y^2\rangle - \langle Y\rangle^2)^2}\end{array}}\end{aligned} \tag{29.15}$$

Figure 29.1 illustrates the application of Equation 29.5 for the measurement of the mean return and the volatility from a data set with $n = 250$ observations and the estimation of their statistical errors in accordance with Equations 29.12 and 29.14. In addition to the relative price changes X, the second and fourth powers are measured. Using these means, the above equations are used to compute the mean return and the volatility as well as the errors involved in estimating these two parameters.

The data set was generated by a simulated random walk with a yield of 6.00% and a volatility of 20.00%. The measured values could thus be compared with the "true values" (a luxury naturally not at our disposal when using historical data). The true values lie within the error of the measurement.

The error is naturally large since the number of measurements taken is so small. As the number of measurements gets bigger, we see that the error decreases as the inverse of square root of the number of measurements; a tenfold decrease in the statistical error can thus only be accomplished if a sample size 100 times as large is placed at our disposal.

	Yield per δt	Vol per δt
Simulated	**6.00%**	**20.00%**
Measured	**5.99%**	**19.70%**
Error	**1.25%**	**1.00%**

x	x^2	x^4	
Mean Values			
0.05986719	0.04223756	0.00502627	
Data			**n**
-0.070894477	0.005026027	2.52609E-05	1
-0.014768650	0.000218113	4.75733E-08	2
0.011417976	0.00013037	1.69964E-08	3
-0.066321993	0.004398607	1.93477E-05	4
-0.113237822	0.012822804	0.000164424	5
0.089718194	0.008049354	6.47921E-05	6
0.200262728	0.04010516	0.001608424	7
0.329164502	0.108349270	0.011739564	8
-0.003557309	1.26544E-05	1.60135E-10	9
-0.066032319	0.004360267	1.90119E-05	10
0.266072855	0.070794764	0.005011899	11
-0.173980700	0.030269284	0.00091623	12
0.193003141	3.72502E-02	0.001387578	13
-0.044716037	0.001999524	3.9981E-06	14
-0.038558758	0.001486778	2.21051E-06	15
-0.033664767	0.001133317	1.28441E-06	16
.....	.....	.....	...
.....	.....	.....	...

Figure 29.1: Measuring the yield, the volatility and their errors from a (simulated) data series of 250 "measurements". The parameters used to simulate the data set are shown for comparison.

It is essential to be aware that errors are also only statistical quantities. Therefore we can *not* be *sure* that the interval defined by the error actually contains the true value of the estimated parameter; there is only a certain *probability* that this is the case. If we have reason to believe that the measured estimator, such as the mean return, is normally distributed, then there is approximately 68% probability that the true parameter will lie within the error interval, since the error is defined as *one* standard deviation. Thus, the probability for the true value to lie outside the range obtained from the statistical error is approximately 32% in this case. If this uncertainty is too large, we could of course define the statistical error to correspond to two or three standard deviations or more. We only need to multiply the error in the derivations above by the desired multiplicative factor. For example, if we define the error as two standard deviations and the measured estimator is normally distributed, then the probability that the true value will lie within the new error interval is 95.4%.

As is clear from the above, we need the probability distribution of the measured estimator if we want to assign confidence levels to error intervals. It can by no means be taken for granted that the measured estimator has the same distribution as the

random variables in the time series. If, for instance, the random variables X in the time series are *uniformly* distributed on a finite interval $[a, b]$ then the estimator for μ as defined in Equation 29.5 is (for large n) approximately *normally* distributed (because of the *central limit theorem*, see Section A.4.3). If, however, the random variables X in the time series are *normally* distributed, then the estimator for μ is also *normally* distributed. However, the estimator for the *variance* is in this case a χ^2-distributed variable. This can be seen as follows: according to Equations 29.9 and 29.3, the estimator for the variance is

$$\operatorname{var}[X] \approx \frac{n}{n-1}\left(\langle X^2\rangle - \langle X\rangle^2\right) = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \langle X\rangle^2\right)$$

The X_i are all normally distributed and the mean $\langle X\rangle$ is also normally distributed in this case. Therefore the estimator for the variance is a sum of *squared* normally distributed random variables and as such χ^2-distributed (see Section A.4.6).

29.3.1 Treatment of Autocorrelations

The methods described above for the determination of the statistical error hold only for *uncorrelated* measurements. It is not always certain that this is the case when dealing with measurements obtained from time series. Autocorrelations as described in Section 29.2 are quite likely to arise. In this case the error considerations must be modified significantly. To do so, we require the autocorrelation time which specifies the number of time steps needed between two measurements in order to guarantee that the two measurements are (at least approximately) uncorrelated. The *autocorrelation time* τ is defined through the autocorrelation Equation 29.6 in the following manner

$$\tau(t) \equiv \frac{1}{2}\sum_{h=-\infty}^{\infty} \varrho(t,h) = \frac{1}{2}\sum_{h=-\infty}^{\infty} \frac{\operatorname{cov}(X_{t+h}, X_t)}{\operatorname{cov}(X_t, X_t)} \tag{29.16}$$

As a rule, only *stationary* time series are investigated (see Section 30). Then ϱ is only dependent on the time difference h between the observations and not on the time point t. The autocorrelation time is then likewise independent of t. For *un*correlated observations we have $\varrho(t,h) = \delta_{h,0}$ (where $\delta_{h,0}$ again denotes the Kronecker delta) and therefore simply $\tau = 1/2$.

If the number n of observations is much larger than the autocorrelation time τ it can be shown that

$$\delta\langle X\rangle \approx \sqrt{\frac{2\tau}{n}\operatorname{var}[X]} \approx \sqrt{\frac{2\tau}{n-1}\left[\langle X^2\rangle - \langle X\rangle^2\right]} \quad \text{for} \quad n >> \tau \tag{29.17}$$

holds for the mean error of X. This reduces to Equation 29.11 if the measurements are uncorrelated since then we have $\tau = 1/2$.

The autocorrelation time (and thus the autocorrelations) must be measured if the error in Equation 29.17 is to be determined. An estimator for the *autocovariance* is

given by

$$\operatorname{cov}(X_{t+h}, X_t) \approx \frac{1}{n-|h|} \sum_{\substack{i,j \\ i-j=|h|}} X_i X_j - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \tag{29.18}$$

For $h = 0$ this estimator is similar to Equation 29.9 but the factor $n/(n-1)$ is not reproduced (for large n however, this factor is very close to 1):

$$\operatorname{var}[X_t] = \operatorname{cov}(X_t, X_t) \approx \frac{1}{n} \sum_{\substack{i,j \\ i=j}} X_i X_j - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 = \langle X_t^2 \rangle - \langle X_t \rangle^2$$

As can be seen from Equation 29.16, the autocorrelations between $h = -\infty$ and $h = -\infty$ must in theory be measured for the calculation of the autocorrelation *time*. In practice, a suitable *cutoff* $(n-\tilde{n})$ can be chosen to limit the sum in Equation 29.16 to a finite one, neglecting the autocorrelations that are so small that they contribute almost nothing to the sum. Substituting the estimator for the autocovariance, Equation 29.18, then yields the estimator for the autocorrelation time.

$$\begin{aligned} \tau &\approx \frac{1}{2} \sum_{h=-(n-\tilde{n})}^{(n-\tilde{n})} \frac{\operatorname{cov}(X_{t+h}, X_t)}{\operatorname{cov}(X_t, X_t)} \qquad \text{with } \tau << \tilde{n} << n \\ &\approx \frac{1}{2} \frac{1}{\langle X^2 \rangle - \langle X \rangle^2} \sum_{h=\tilde{n}-n}^{n-\tilde{n}} \left[\frac{1}{n-|h|} \sum_{\substack{i,j \\ i-j=|h|}} X_i X_j - \langle X \rangle^2 \right] \end{aligned}$$

All this substituted into Equation 29.17 finally provides a possibility of estimating the error of a mean taking autocorrelations into account:

$$\delta(\langle X \rangle) \approx \sqrt{\frac{1}{n-1} \sum_{h=\tilde{n}-n}^{n-\tilde{n}} \left[\frac{1}{n-|h|} \sum_{\substack{i,j \\ i-j=|h|}} X_i X_j - \langle X \rangle^2 \right]}$$

There are thus two possibilities of taking the autocorrelations into consideration:

- We wait for at least as long as the autocorrelation time before taking a second measurement; the measurements would then be uncorrelated.
- We do *not* wait for the autocorrelation time to pass (because we do not know how long it is, for instance) and use the above expression for the error of the correlated observations.

30

Time Series Modeling

Time series analysis goes a significant step further than merely determining statistical parameters from observed time series data (such as the variance, correlation, etc.) as described above. Indeed, it is primarily used as a tool for deriving *models* describing the time series concerned. Estimators such as those appearing in Equation 29.5 are examples of how *parameters* can be estimated which are subsequently used to model the stochastic process governing the time series (for example, a random walk with drift μ and volatility σ). Building a model which "explains" and "describes" the time series data is the principle goal of time series analysis. The object is thus to interpret a series of observed data points $\{X_t\}$, for example a historical price or volatility evolution (in this way acquiring a fundamental understanding of the process) and to *model* the processes underlying the observed historical evolution. In this sense, the historical sequence of data points is interpreted as just one *realization* of the time series process. The parameters of the process are then estimated from the available data and can subsequently be used in making *forecasts*, for example.

As much structure as possible should be extracted from a given data sequence and then transferred to the model. Let $\{\widehat{X}_t\}$ be the time series generated by the model process (called the *estimated* time series). The difference between this and the actually observed data points $\{X_t\}$ are called *residues* $\{X_t - \widehat{X}_t\}$. These should consist of only "noise", i.e. they should be unpredictable random numbers.

In order to be able to fit a time series model, the "raw data", i.e. the sequence of historical data points, must sometimes undergo a *pre-treatment*. In this procedure, *trends* and *seasonal components* are first eliminated and a change may be made to the scale of the data, so that the resulting sequence is a *stationary time series*[1]. A stationary time series is characterized by the *time* invariance of its expectation, variance and covariance. In particular, the expectation and variance are constant. Without loss of generality, the expectation can be assumed to be zero since it can be eliminated during the pre-treatment through a *centering of the time series.* This is accomplished by subtracting the mean $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ from every data point in the time series $\{X_t\}$.

[1] More precisely, we are dealing with a *weakly stationary* time series in what follows.

As just discussed, the *stationarity* of a time series implies $\mathrm{E}[X_t] = \mathrm{E}[X]\ \forall t$ and the autocovariance Equation 29.7 becomes

$$\mathrm{cov}(X_{t+h}, X_t) = \mathrm{E}[X_{t+h}\ X_t] - \mathrm{E}\,[X]\,\mathrm{E}[X] = \mathrm{E}[X_{t+h}\ X_t] \tag{30.1}$$

The final equality in the above equation holds if the time series has been centered in the pre-treatment. We will always assume this to be the case. Furthermore, the autocovariance and autocorrelation (just as the variance) are independent of t if the time series is stationary, and therefore depend only on the *time lag* h. We frequently write

$$\gamma(h) := \mathrm{cov}(X_{t+h}, X_t)$$

Likewise, if the time series is stationary we have $\varrho(t,h) = \varrho(h)$ in the autocorrelation Equation 29.6. The following useful symmetry relations can be derived directly from the stationarity of the time series (this can be shown by substituting t with $t' = t - h$):

$$\gamma(-h) = \gamma(h) \quad , \quad \varrho(-h) = \varrho(h) \tag{30.2}$$

From Definition 29.7, we can immediately obtain an estimate[2] for the autocorrelation and the autocovariance of a stationary data sequence

$$\widehat{\gamma}(h) = \widehat{\mathrm{cov}}(X_{t+h}, X_t) = \frac{1}{T}\sum_{t=1}^{T-h}(X_{t+h} - \langle X\rangle)(X_t - \langle X\rangle)\ , \quad \widehat{\varrho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \tag{30.3}$$

for $h \in \mathbb{N}_0$. The autocovariances (and autocorrelations) are usually computed for at most $h \leq 40$. Note that h has to be substantially smaller than T in all cases; the estimation is otherwise too inexact[3].

Of course, we can fit *different* time series models to a stationary time series (after having undergone a pre-treatment if necessary) and then compare their goodness of fit and forecasting performance. Thus the following three general steps must be taken when modeling a given sequence of data points

1. Pre-treatment of the data sequence to generate a stationary series (elimination of trend and seasonal components, transformation of scale, etc.).
2. Estimation and/or fitting of the time series model and its parameters.
3. Evaluation of the goodness of fit and forecasting performance on the basis of which a decision is made as to whether the tested model should be accepted or a new model selected (step 2).

Figure 30.1 shows the daily relative change (returns) of the FTSE Index taken from the daily data from Jan-01-1987 through Apr-01-1998 (2,935 days). This sequence of data points is defined as

$$X_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \tag{30.4}$$

[2]In the following material, we will distinguish the estimator of a parameter from the parameter itself with a "hat" notation.

[3]The fact that only T appears in the denominator in Equation 30.3 instead of $T-h$, as one might expect, guarantees that the estimator for the covariance matrix $[\hat{\gamma}(i-j)]_{i,j=1}^{T}$ is automatically positive definite.

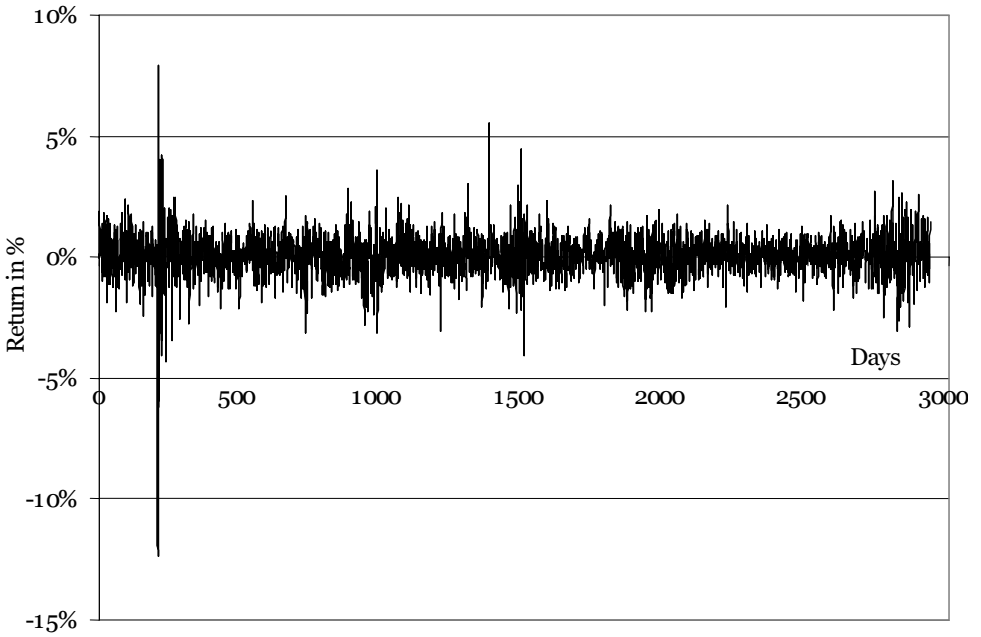


Figure 30.1: Daily returns of the FTSE index as an example of a stationary data series. The crash in October 1987 is clearly visible.

where $\{Y_t\}$ represents the original data sequence of FTSE values. The data set $\{X_t\}$ consists of 2,934 values. According to Equation 28.9, the relative changes in Equation 30.4 are approximately equal to the difference of the logarithms if the daily changes are sufficiently small:

$$X_t \approx \ln(Y_t) - \ln(Y_{t-1}) \tag{30.5}$$

This is the first difference of the logarithm of the original sequence of FTSE index values. The above example represents a typical pre-treatment procedure performed on the data. Instead of the original data $\{Y_t\}$, which is by no means stationary (drift$\neq 0$ and variance increase with time as $\sim \sigma t$), we generate a stationary data sequence as in Equation 30.5 through standard transformations in time series analysis. Specifically in our case, what is known as *Box-Cox scaling* (taking the logarithm of the original data) was performed and subsequently the first differences were calculated for the purpose of trend elimination. Stationary time series data like these are then used in the further analysis, in particular, when fitting a model to the data.

The above example should provide sufficient motivation for the pre-treatment of a time series. The interested reader is referred to Section 33.1 for further discussion of pre-treating time series data to generate stationary time series. We will assume from now on that the given time series have already been pre-treated, i.e. potential trends and seasonal components have already been eliminated and scaling transformations have already been performed appropriately, so that the resulting data sequences are

stationary. Such a stationary time series is given by a sequence of random variables $\{X_t\}$, $t \in \mathbb{N}$.

30.1 Stationary Time Series and Autoregressive Models

This chapter introduces a basic approach in time series analysis employing a specific time series model, called autoregressive model. We then continue by extending the results to the case of a time-dependent variance (GARCH model) which finds application in modeling *volatility clustering* in financial time series. This technique is widely used in modeling the *time evolution* of volatilities.

Rather than working under the idealized assumption of time-continuous processes, the processes modeled in this chapter are truly discrete in time. The discussion is geared to the needs of the user. We will forgo mathematical rigor and in most cases the proofs of results will not be given. Not taking these "shortcuts" would increase the expanse of this chapter considerably and shift the focus of the book to that comparable to a mathematical treatise. However, the attempt will be made to provide thorough reasoning for all results presented.

A process for modeling a time series of stock prices, for example, has already been encountered in this text: the random walk. An important property of the random walk is the *Markov property.* Recall that the Markov property states that the next step in a random walk depends solely on its current value, but not on the values taken on at any previous times. If such a Markov process is unsatisfactory for modeling the properties of the time series under consideration, an obvious generalization would be to allow for the influence of past values of the process. Processes whose current values can be affected by values attained in the past are called *autoregressive.* In order to characterize these processes, we must first distinguish between the *unconditional* and *conditional* variance denoted by $\mathrm{var}[X_t]$ and $\mathrm{var}[X_t|X_{t-1}, \ldots, X_1]$, respectively. The *un*conditional variance is the variance we are familiar with from previous chapters, whereas the conditional variance is the variance of X_t *under the condition* that $X_{t-1}, \ldots, X_1$ have occurred. Analogously, we must differentiate between the *conditional* and *unconditional* expectation denoted by $\mathrm{E}[X_t]$ and $\mathrm{E}[X_t|X_{t-1}, \ldots, X_1]$, respectively, where the last is the expectation of X_t *under the condition* that $X_{t-1}, \ldots, X_1$ have occurred. There is no difference between the two when the process under consideration is independent of its history.

30.1.1 AR(p) Processes

Having made these preparatory remarks and definitions, we now want to consider processes whose current values are influenced by one or more of their predecessors. If, for example, the effect of the p previous values of a time series on the current value is *linear*, the process is referred to as an *autoregressive process of order p,* and denoted by AR(p). The general autoregressive process of p^{th} order makes use of p process values

in the past to generate a representation of today's value, or explicitly

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots \phi_p X_{t-p} + \varepsilon_t \\ &= \sum_{i=1}^{p} \phi_i X_{t-i} \; + \varepsilon_t \,, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2) \end{aligned} \tag{30.6}$$

The changes ε_t here are independent of all previous time series values $X_s,\ s < t$, and thus represent an injection of truly new information into the process. In particular, this means that $\mathrm{cov}[X_i, \varepsilon_j]$ is always zero. The conditional variance and conditional expectation of the process are

$$\begin{aligned} \mathrm{E}[X_t | X_{t-1}, \ldots, X_1] &= \sum_{i=1}^{p} \phi_i X_{t-i} \\ \mathrm{var}[X_t | X_{t-1}, \ldots, X_1] &= \mathrm{var}[\varepsilon_t] = \sigma^2 \end{aligned} \tag{30.7}$$

It can be shown that stationarity is guaranteed if the zeros z_k of the *characteristic polynomial*[4]

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \tag{30.8}$$

lie outside of the closed unit disk, i.e. when the norm $|z_k|$ is larger than 1 for all zeros z_k. In particular, if the process is stationary then the *un*conditional expectation and variance have the following properties: $\mathrm{E}[X_t] = \mathrm{E}[X_{t-i}]$ and $\mathrm{var}[X_t] = \mathrm{var}[X_{t-i}]$. Exploiting this, we can easily calculate explicit expressions for the *un*conditional expectation and variance. The unconditional expectation $\mathrm{E}[X_t]$ is

$$\mathrm{E}[X_t] = \mathrm{E}[\sum_{i=1}^{p} \phi_i X_{t-i} \; + \varepsilon_t] = \sum_{i=1}^{p} \phi_i \underbrace{\mathrm{E}\left[X_{t-i}\right]}_{\mathrm{E}[X_t]} + \underbrace{\mathrm{E}\left[\varepsilon_t\right]}_{0} = \mathrm{E}\left[X_t\right] \sum_{i=1}^{p} \phi_i$$

In the first step we have simply used Definition 30.6 for X_t. The second step is merely the linearity of the expectation operator. In the third step we have finally used the decisive properties of the process, namely stationarity of the expectation and randomness of the residues. The result is therefore

$$\mathrm{E}[X_t] \left(1 - \sum_{i=1}^{p} \phi_i \right) = 0$$

This implies that the unconditional expectation must be zero since stationarity guarantees that the sum of the ϕ_i is *not* equal to one[5].

[3] The notation ε_t will always indicate independent, identically $\mathrm{N}(0, \sigma^2)$-distributed random variables. Another common definition is $\varepsilon_t \sim \mathrm{W}(0, \sigma^2)$, where W stands for *white noise*. This is a somewhat more general statement and is used in reference to random variables which are not normally distributed as well.

[4] This polynomial plays a central role in the theory of time series.

[5] This can be shown using the characteristic polynomial.

The unconditional variance can be computed using similar arguments

$$\begin{aligned}
\mathrm{var}[X_t] &= \mathrm{var}[\sum_{i=1}^{p} \phi_i X_{t-i} \; + \varepsilon_t] \\
&= \sum_{i,j=1}^{p} \phi_i \phi_j \mathrm{cov}[X_{t-i}, X_{t-j}] + \sum_{i=1}^{p} \phi_i \mathrm{cov}[X_{t-i}, \varepsilon_t] + \mathrm{var}[\varepsilon_t] \\
&= \sum_{i,j=1}^{p} \phi_i \phi_j \mathrm{cov}[X_{t-i}, X_{t-j}] + 0 + \sigma^2 \\
&= \mathrm{var}[X_t] \sum_{i,j=1}^{p} \phi_i \phi_j \varrho(i-j) + \sigma^2
\end{aligned}$$

where we used Equation A.17 and – in the last step – Definition 29.6 for stationary processes. Solving for $\mathrm{var}[X_t]$ yields immediately

$$\mathrm{var}[X_t] = \frac{\sigma^2}{1 - \sum_{i,j=1}^{p} \phi_i \phi_j \varrho(i-j)} \tag{30.9}$$

An expression for the autocorrelation function ϱ of the process can be obtained by multiplying both sides of Equation 30.6 by X_{t-h} and taking the expectation. Here stationarity is used in form of Equations 30.2 and 30.1:

$$\begin{aligned}
\varrho(h) = \varrho(-h) &= \frac{\mathrm{cov}(X_{t-h}, X_t)}{\mathrm{cov}(X_t, X_t)} = \frac{\mathrm{E}(X_{t-h}, X_t)}{\mathrm{E}(X_t^2)} \\
&= \frac{1}{\mathrm{E}(X_t^2)} \mathrm{E}(X_{t-h}, \sum_{i=1}^{p} \phi_i X_{t-i} \; + \varepsilon_t) \\
&= \frac{1}{\mathrm{E}(X_t^2)} \sum_{i=1}^{p} \phi_i \mathrm{E}(X_{t-h}, X_{t-i}) + \frac{1}{\mathrm{E}(X_t^2)} \underbrace{\mathrm{E}(X_{t-h}, \varepsilon_t)}_{0} \\
&= \sum_{i=1}^{p} \phi_i \frac{\mathrm{E}(X_{t-h+i}, X_t)}{\mathrm{E}(X_t^2)} = \sum_{i=1}^{p} \phi_i \frac{\mathrm{E}(X_{t-(h-i)}, X_t)}{\mathrm{E}(X_t^2)}
\end{aligned}$$

and thus

$$\varrho(h) = \sum_{i=1}^{p} \phi_i \varrho(h-i) \tag{30.10}$$

These are the *Yule-Walker equations* for the autocorrelations ϱ. The autocorrelations can thus be computed recursively by setting the initial condition $\varrho(0) = 1$. Consider the following example of an AR(2) process:

$$\varrho(1) = \phi_1 \varrho(1-1) + \phi_2 \varrho(1-2) = \phi_1 1 + \phi_2 \varrho(1) \Rightarrow \quad \varrho(1) = \frac{\phi_1}{1-\phi_2}$$

$$\varrho(2) = \phi_1 \varrho(1) + \phi_2 \varrho(0) = \frac{\phi_1^2}{1-\phi_2} + \phi_2 \text{ , and so on.}$$

Here, the symmetry indicated in Equation 30.2 was used together with Equation 30.10. Substituting these autocorrelations into Equation 30.9 finally yields the unconditional variance of an AR(2) process:

$$\begin{aligned}\mathrm{var}[X_t] &= \frac{\sigma^2}{1-\phi_1^2\varrho(1-1)-\phi_1\phi_2\varrho(1-2)-\phi_2\phi_1\varrho(2-1)-\phi_2^2\varrho(2-2)} \\ &= \frac{\sigma^2}{1-\phi_1^2-\phi_2^2-2\phi_1\phi_2\varrho(1)} = \frac{\sigma^2}{1-\phi_1^2-\phi_2^2-2\phi_1^2\phi_2/(1-\phi_2)}\end{aligned}$$

In practice, however, the autocorrelations are usually computed from the data series itself with the aid of Equation 30.3, and not from the coefficients ϕ_i. After all, the data sequence is given and *not* the ϕ_i, $i = 1, 2, \ldots, p$ which themselves must be estimated.

The Autoregressive Process of First Order

We now consider the most simple case, namely $p = 1$. Explicitly, the *autoregressive process of first order AR(1)* is defined as

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2) \tag{30.11}$$

The stationarity condition for this process implies that $|\phi| < 1$ since Equation 30.8 states simply that

$$1 - \phi z = 0 \quad \text{for some } z \text{ where } |z| > 1$$

The conditional variance and conditional expectation of the process are

$$\begin{aligned}\mathrm{E}[X_t|X_{t-1}, \ldots, X_1] &= \phi X_{t-1} \\ \mathrm{var}[X_t|X_{t-1}, \ldots, X_1] &= \mathrm{var}[\varepsilon_t] = \sigma^2\end{aligned}$$

The *un*conditional expectation is equal to zero as was shown above to hold for general AR(p) processes. The *un*conditional variance can be calculated as

$$\begin{aligned}\mathrm{var}[X_t] &= \mathrm{var}[\phi X_{t-1} + \varepsilon_t] \\ &= \phi^2\mathrm{var}[X_{t-1}] + \phi\mathrm{cov}[X_{t-1}, \varepsilon_t] + \mathrm{var}[\varepsilon_t] \\ &= \phi^2\mathrm{var}[X_t] + 0 + \sigma^2 \implies \\ \mathrm{var}[X_t] &= \frac{\sigma^2}{1-\phi^2}.\end{aligned} \tag{30.12}$$

Recursively constructing future values via Equation 30.11 starting from X_t yields

$$\begin{aligned}X_{t+h} &= \phi X_{t+h-1} + \varepsilon_{t+h} \\ &= \phi^2 X_{t+h-2} + \phi\varepsilon_{t+h-1} + \varepsilon_{t+h} \\ &\cdots \\ &= \phi^h X_t + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t+h-i}\end{aligned}$$

The autocovariance of the AR(1) thus becomes explicitly

$$\begin{aligned}\operatorname{cov}(X_{t+h}, X_t) &= \operatorname{cov}(\phi^h X_t + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t+h-i}, X_t) \\ &= \phi^h \operatorname{cov}(X_t, X_t) + \sum_{i=0}^{h-1} \phi^i \underbrace{\operatorname{cov}(\varepsilon_{t+h-i}, X_t)}_{0} \\ &= \phi^h \operatorname{var}[X_t] \\ &= \phi^h \frac{\sigma^2}{1-\phi^2}\end{aligned}$$

The auto*correlation* is therefore simply ϕ^h, and as such is an exponentially decreasing function of h. The same result can of course be obtained from the Yule-Walker equations

$$\varrho(h) = \sum_{i=1}^{1} \phi_i \varrho(h-i) = \phi\varrho(h-1) = \phi^2\varrho(h-2) = \cdots = \phi^h \underbrace{\varrho(h-h)}_{1}$$

It is worthwhile to consider a *random walk* from this point of view. A (one-dimensional) random walk is by definition constructed by adding an independent, identically distributed random variable (*iid*, for short) with variance σ^2 to the last value of attained in the walk. The random walk can thus be written as

$$X_t = X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

It follows from this definition that the conditional variance of the random walk is σ^2 and the expectation equals zero. The random walk corresponds to an AR(1) process with $\phi = 1$. This contradicts the stationarity criterion $|\phi| < 1$! The random walk is therefore a *non*-stationary AR(1) process. The non-stationarity can be seen explicitly by considering the *un*conditional variance:

$$\begin{aligned}\operatorname{var}[X_t] &= \operatorname{var}[X_{t-1} + \varepsilon_t] \\ &= \operatorname{var}[X_{t-1}] + \operatorname{cov}[X_{t-1}, \varepsilon_t] + \operatorname{var}[\varepsilon_t] \\ &= \operatorname{var}[X_{t-1}] + 0 + \sigma^2\end{aligned}$$

Thus, for all $\sigma \neq 0$ we have $\operatorname{var}[X_{t-1}] \neq \operatorname{var}[X_t]$, i.e. the process *cannot* be stationary. Therefore we cannot obtain a closed from expression similar to Equation 30.12 for the unconditional variance (this can also be seen from the fact that if $\phi = 1$, Equation 30.12 would imply a division by zero). The unconditional variance can, however, be determined recursively

$$\operatorname{var}[X_t] = k\sigma^2 + \operatorname{var}[X_{t-k}]$$

Assuming from the outset that a value $X_{t=0}$ is known (and because it is known, has zero variance) we obtain the well-known property of the random walk

$$\operatorname{var}[X_t] = t\sigma^2$$

The variance is thus time dependent; this is a further indication that the random walk is not stationary. Since the variance is linear in the time variable, the standard deviation is proportional to the square root of time. This is the well-known *square root law* for scaling the volatility with time.

Another special case of an AR(1) process is *white noise* which has an expectation equal to zero and constant variance. It is defined by

$$X_t = \varepsilon_t$$

The random variables$\{\varepsilon_t\}$ are *iid* random variables with variance σ^2. This corresponds to the AR(1) process with $\phi = 0$. The stationarity criterion $|\phi| < 1$ is satisfied and the above results for the stationary AR(1) process can be applied with $\phi = 0$, for example, $\mathrm{cov}(X_{t+h}, X_t) = 0$ and $\mathrm{var}(X_t) = \sigma^2$.

30.1.2 Univariate GARCH(p, q) Processes

The conditional variance of the AR(p) processes introduced above was always a constant function of time; in each case it was equal to the variance of ε_t. This, however, is not usually the case for financial time series. Take, for example, the returns of the FTSE data set in Figure 30.1. It is clear to see that the variance of the data sequence is not constant as a function of time. On the contrary, the process goes through both calm and quite volatile periods. It is much more probable that large price swings will occur close to other large price swings than to small swings. This behavior is typical of financial time series and is referred to as *volatility clustering* or simply *clustering*. A process which is capable of modeling such behavior is the *GARCH(p, q) process*, which will be introduced below. The decisive difference between GARCH and AR(p) processes is that not only past values of X_t are used in the construction of a GARCH process, but past values of the *variance* enter into the construction as well. The GARCH(p, q) process is defined as

$$X_t = \sqrt{H_t}\varepsilon_t \quad \text{with} \quad H_t = \alpha_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2, \quad \varepsilon_t \sim \mathrm{N}(0,1) \qquad (30.13)$$

where the $\{\varepsilon_t\}$ are *iid* standard normally distributed. The $\{\varepsilon_t\}$ are independent of X_t. The time series$\{X_t\}$ is, in principle, white noise $\{\varepsilon_t\}$ with a time-dependent variance which is determined by the coefficients $\{H_t\}$. These H_t take into consideration the past values of the time series *and* the variance. If the $\{X_t\}$ are large (distant from the equilibrium value which is in this case zero as $\mathrm{E}[\varepsilon_t] = 0$), then so is $\{H_t\}$. For small values $\{X_t\}$ the opposite holds. In this way, clustering can be modeled. The order q indicates how many past values of the time series $\{X_t\}$ influence the current value H_t. Correspondingly, p is the number of past values of the variance itself which affects the current value of H_t. In order to ensure that the variance is positive, the parameters

must satisfy the following conditions:

$$\begin{aligned} \alpha_0 &\geq 0 \\ \beta_1 &\geq 0 \\ \sum_{j=0}^{k} \alpha_{j+1} \beta_1^{k-j} &\geq 0, \quad k = 0, \ldots, q-1 \end{aligned} \tag{30.14}$$

This implies that $\alpha_1 \geq 0$ always holds, the other α_i however, may be negative. Furthermore, the time series $\{X_t\}$ should be (weakly) stationary to prevent it from "drifting away". The following condition is sufficient to guarantee this stationarity:

$$\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j < 1 \tag{30.15}$$

The two most important properties of this process pertain to the conditional expectation and the conditional variance

$$\begin{aligned} \mathrm{E}[X_t | X_{t-1}, \ldots, X_1] &= 0 \quad \text{and} \\ \mathrm{var}[X_t | X_{t-1}, \ldots, X_1] &= H_t = \alpha_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 \end{aligned} \tag{30.16}$$

The first equation holds because $\mathrm{E}[\varepsilon_t] = 0$, the second because $\mathrm{var}[\varepsilon_t] = 1$. The H_t are thus the conditional variances of the process. The conditional expectation (under the condition that all X up to time $t-1$ are known) of H_t is simply H_t itself since no stochastic variable ε appears in Equation 30.13 where H_t is defined, and thus

$$\mathrm{E}[H_t | X_{t-1}, \ldots, X_1] = H_t = \alpha_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 \tag{30.17}$$

H is thus always known one time step in advance of X. This may seem trivial but will be quite useful in Section 31.2 when making volatility forecasts.

The *un*conditional variance is by definition

$$\begin{aligned} \mathrm{var}[X_t] &= \mathrm{E}[X_t^2] - \mathrm{E}[X_t]^2 \\ &= \mathrm{E}[H_t \varepsilon_t^2] - \mathrm{E}[\sqrt{H_t}\varepsilon_t]^2 \\ &= \mathrm{E}[H_t]\mathrm{E}[\varepsilon_t^2] - (\mathrm{E}[\sqrt{H_t}]\mathrm{E}[\varepsilon_t])^2 \end{aligned}$$

where in the last step we have made use of the fact that $\{\varepsilon_t\}$ are uncorrelated with $\{H_t\}$. Furthermore, since the $\{\varepsilon_t\}$ are *iid* $\mathrm{N}(0,1)$ distributed

$$\mathrm{E}[\varepsilon_t] = 0 \quad \text{and} \quad \mathrm{E}[\varepsilon_t^2] = \mathrm{E}[\varepsilon_t^2] - 0 = \mathrm{E}[\varepsilon_t^2] - (\mathrm{E}[\varepsilon_t])^2 = \mathrm{var}[\varepsilon_t] = 1$$

and therefore

$$\mathrm{var}[X_t] = \mathrm{E}[H_t] = \alpha_0 + \sum_{i=1}^{p} \beta_i \mathrm{E}[H_{t-i}] + \sum_{j=1}^{q} \alpha_j \mathrm{E}[X_{t-j}^2]$$

Just as the ε_t, the X_t have zero expectation, which also implies that $\mathrm{E}[X_t^2] = \mathrm{var}[X_t]$. Under consideration of this relation and the stationarity (constant variance), all of the expectations involving squared terms in the above equation can be written as the variance of X_t :

$$\mathrm{E}[X_{t-j}^2] = \mathrm{var}[X_{t-j}] = \mathrm{var}[X_t]$$
$$\mathrm{E}[H_{t-i}] = \mathrm{var}[X_{t-j}] = \mathrm{var}[X_t]$$

This leads to the following equation for the unconditional variance:

$$\mathrm{var}[X_t] = \alpha_0 + \sum_{i=1}^{p} \beta_i \mathrm{var}[X_t] + \sum_{j=1}^{q} \alpha_j \mathrm{var}[X_t] \Longleftrightarrow$$
$$\mathrm{var}[X_t] = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} =: \widetilde{\alpha}_0 \tag{30.19}$$

This unconditional variance can of course also be estimated from the observed data, i.e. as usual through the computation of the empirical variance estimator over a large number of realizations of $\{X_t\}$.

The GARCH(p, q) process can be expressed in terms of the unconditional variance $\widetilde{\alpha}_0$ as follows:

$$\begin{aligned} H_t &= \alpha_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 \\ &= \alpha_0 \frac{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 \\ &= \widetilde{\alpha}_0 - \widetilde{\alpha}_0 \sum_{i=1}^{q} \alpha_i - \widetilde{\alpha}_0 \sum_{j=1}^{p} \beta_j + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 \\ &= \widetilde{\alpha}_0 + \sum_{i=1}^{p} \beta_i (H_{t-i} - \widetilde{\alpha}_0) + \sum_{j=1}^{q} \alpha_j (X_{t-j}^2 - \widetilde{\alpha}_0) \end{aligned} \tag{30.20}$$

The conditional variance H_t can thus be interpreted as the unconditional variance $\widetilde{\alpha}_0$ plus the sum of the distances from this unconditional variance. If all α_j and β_i are greater than zero (which is always the case for a GARCH(1,1) process), this form of the conditional variance has another interpretation:

The β_i terms cause a kind of *persistence* of the variance which serves to model the *volatility clustering* phenomenon: the greater H_{t-i} becomes in comparison to the long-term expectation $\widetilde{\alpha}_0$ (the unconditional variance), the greater the positive contribution of these terms to H_t; the H_t tend to get even larger. Conversely, for values of H_{t-i} which are smaller than $\widetilde{\alpha}_0$ the contribution of these terms become negative and thus H_t will tend to get even smaller.

The terms involving α_i describe the *reaction* of the volatility to the the process itself. Values X_{t-j}^2 larger than $\widetilde{\alpha}_0$ favor a growth in the variance; values X_{t-j}^2 smaller than $\widetilde{\alpha}_0$ result in a negative contribution and thus favor a decline in the variance. If

the process itself describes a price *change*, as is common in the financial world, this is precisely the effect that strong price changes tend to induce growth in volatility.

Overall, these properties lead us to expect that GARCH models are indeed an appropriate choice for modeling certain phenomena observed in the financial markets (in particular, volatility clustering and the reaction of the volatility to price changes). In practical applications, we often set $p = 1$ and even $q = 1$. It has been shown in the literature that significantly better results are not achieved with larger values of p and q and thus the number of parameters to be estimated would be unnecessarily increased.

30.1.3 Simulation of GARCH Processes

One of the examples to be found on the Excel workbook GARCH.XLS on the CD-ROM accompanying this book is the generation (simulation) of a *GARCH(1,1) process*. The first simulated value X_1 of the time series is obtained, according to Equation 30.13, from a realization of a standard normal random variable followed by multiplication of this number by $\sqrt{H_1}$. Subsequently, H_2 is computed from the values now known at time $t = 1$. Then, a realization X_2 is generated from a standard normal random variable and multiplied by $\sqrt{H_2}$. This procedure is repeated until the end of the time series is reached.

In order to generate a GARCH(p, q) process, q values of the X process

$$X_{-q+1}, X_{-q+2}, \ldots, X_0$$

and p values of the conditional variance

$$H_{-p+1}, H_{-p+2}, \ldots, H_0$$

must be available in order to be able to compute the first conditional variance H_1 as indicated in Equation 30.13. The choice of these initial values is not unique but the orders of magnitude of the time series values and the variances should at least be correct. The unconditional expectation $\mathrm{E}[X_t]$ and the unconditional variance $\mathrm{var}[X_t]$ are therefore good candidates for this choice. The first values of the generated time series should then be rejected (as a rule, 50 values are sufficient), since they still include the above "initial conditions". After taking several steps, realizations of the desired GARCH process can be generated. Figure 30.2 illustrates a simulated GARCH process.

Such simulated time series can be implemented to test optimization methods which have the objective of "finding" parameters from the simulated data series which have been previously used for the simulation. After all, if a data set is given (real or simulated), the parameters of a model have to be determined. Methods for doing this are the subject of the next section.

30.2 Calibration of Time Series Models

All of the time series models introduced above include parameters which may be varied for the purpose of fitting the model "optimally" to the time series data. We represent

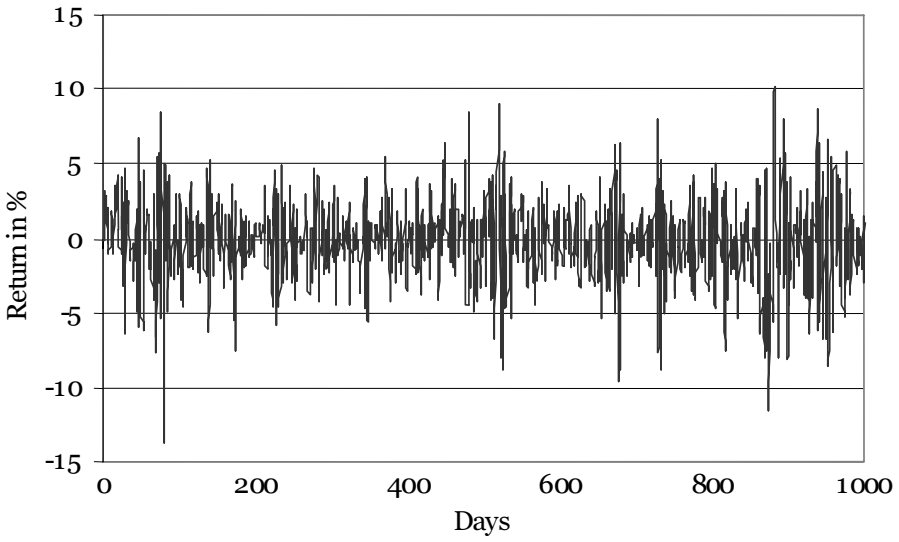


Figure 30.2: Simulated GARCH(1,1) process. The first 100 values have not been used. Clustering can clearly be observed.

these parameters as a parameter vector θ. For an AR(p) process, the free parameters are the ϕ_i and σ^2 while the GARCH(p, q) has the free parameters α_i and β_i. Thus

$$\begin{array}{ll} \theta = (\phi_1, \phi_2, \ldots, \phi_p, \sigma^2) & \text{for AR}(p) \\ \theta = (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \beta_2, \ldots \beta_p) & \text{for GARCH}(p, q) \end{array}$$

A widely used estimation procedure for the determination of unknown parameters in statistics is the *maximum likelihood estimator*. This procedure selects the parameter values which maximize the likelihood of the model being correct. These are just the parameter values which maximize the *probability* (called the *likelihood*) that the values observed will be realized by the assumed model. Using the model, the probability is expressed as a function of the parameters θ. Then this probability function is maximized by varying the parameter values. The parameter values for which the probability function attains a maximum corresponds to a "best fit" of the model to the given data sequence. They are the most probable parameter values given the information available (i.e. given the available time series). This procedure will now be performed explicitly for both an AR(p) and a GARCH(p,q) process.

30.2.1 Parameter Estimation for AR(p) Processes

The likelihood for the AR(p) process is obtained as follows: from Equations 30.6 and 30.7 we can see that if we assume an AR(p) process with a parameter vector

$$\theta = (\phi_1, \phi_2, \ldots, \phi_p, \sigma^2)$$

then X_t has the normal distribution

$$\mathrm{N}(\sum_{i=1}^{p} \phi_i X_{t-i}, \sigma^2)$$

The conditional probability for one single observed value of X_t (also called the *conditional likelihood* of X_t) is thus

$$\begin{aligned} &L_\theta(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_{t-p}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[X_t - \sum_{i=1}^{p} \phi_i X_{t-i} \right]^2 \right\} \end{aligned}$$

The total likelihood for all T measured data points is in consequence of the independence of ε_t simply a product of all conditional likelihoods:

$$\begin{aligned} L_\theta(X_1, X_2, \ldots, X_T) &= \prod_{t=1}^{T} L_\theta(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_{t-p}) \\ &= \frac{1}{(2\pi)^{T/2}\, \sigma^T} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2\sigma^2} \left[X_t - \sum_{i=1}^{p} \phi_i X_{t-i} \right]^2 \right\} \end{aligned}$$

Observe that for the likelihoods of the first data points X_t where $t < p + 1$,a further p data points $\{X_0, X_{-1}, \ldots, X_{-p+1}\}$ are required in advance. The extent of the data sequence needed is thus a data set encompassing $T + p$ data points.

Maximizing this likelihood through the variation of the parameters $\phi_1, \phi_2, \ldots, \phi_p$ and σ^2, we obtain the parameters $\{\phi_1, \phi_2, \ldots, \phi_p, \sigma^2\}$ which, under the given model assumptions[6], actually maximizes the (model) probability that the observed realization $\{X_t\}$ will actually appear. It is, however, simpler to maximize the *logarithm* of the likelihood (because of the size of the terms involved and the fact that sums are more easily dealt with than products). Since the logarithm function is strictly monotone increasing, the maximum of the likelihood function is attained for the same parameter values as the maximum of the logarithm of the likelihood function. The *log-likelihood* function for the AR(p) process is given by

$$\mathcal{L}_\theta = -\frac{T}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{p} \phi_i X_{t-i} \right]^2$$

[6]The model assumption is that the time series was generated by an AR(p) process.

$\phi_1, \phi_2, \ldots, \phi_p$ appear only in the last expression (the sum), which appears with a negative sign in the log-likelihood function. The values of $\phi_1, \phi_2, \ldots, \phi_p$ which maximize the log-likelihood function therefore minimize the expression

$$\sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{p} \phi_i X_{t-i} \right]^2 \tag{30.21}$$

This, however, is just a sum of the quadratic deviations. The desired parameter estimates $\{\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_p\}$ are thus the solution to a least squares problem. The $\widehat{\phi}_t$ can thus be determined independently from the variance σ^2. The estimation of the variance is obtained from simple calculus by taking the derivative of the log-likelihood function with respect to σ^2 and setting the resulting value equal to zero (the optimal ϕ_i, namely the $\widehat{\phi}_i$, have already been substituted into the log-likelihood function):

$$\begin{aligned}
\frac{\partial \mathcal{L}_\theta}{\partial \sigma^2} &= -\frac{T}{2} \frac{\partial \ln(2\pi\sigma^2)}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2} \left(\frac{1}{2\sigma^2} \sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i} \right]^2 \right) \\
&= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i} \right]^2 \stackrel{!}{=} 0
\end{aligned}$$

The optimal estimate for σ^2 becomes

$$\widehat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \left[X_t - \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i} \right]^2 \tag{30.22}$$

For example, the maximum likelihood estimator for ϕ_1 in the AR(1) process in Equation 30.11, obtained by minimizing the expression in 30.21, can be determined through the following computation:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \phi_1} \sum_{t=1}^{T} \left[X_t - \phi_1 X_{t-1} \right]^2 = 2 \sum_{t=1}^{T} (X_t - \phi_1 X_{t-1})(-X_{t-1}) \\
&= 2\phi_1 \sum_{j=1}^{T} X_{j-1}^2 - 2 \sum_{t=1}^{T} X_{t-1} X_t \Longrightarrow \\
\widehat{\phi}_1 &= \frac{\sum_{t=1}^{T} X_{t-1} X_t}{\sum_{j=1}^{T} X_{j-1}^2}
\end{aligned}$$

Substituting this into Equation 30.22 yields the maximum likelihood estimator for σ^2

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{T} \sum_{t=1}^{T} \left[X_t - \widehat{\phi}_1 X_{t-1} \right]^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} \left[X_t - X_{t-1} \frac{\sum_{i=1}^{T} X_{i-1} X_i}{\sum_{j=1}^{T} X_{j-1}^2} \right]^2
\end{aligned}$$

30.2.2 Parameter Estimation for GARCH(p, q) Processes

The likelihood for the GARCH(p, q) process is obtained as follows: from Equations 30.13 and 30.16 we see that

$$X_t|\{X_{t-1}, \ldots, X_{t-q}, H_{t-1}, \ldots, H_{t-p}\} \sim \mathrm{N}(0, H_t)$$

This implies that, given the information $\{X_{t-1}, \ldots, X_{t-q}, H_{t-1}, \ldots, H_{t-p}\}$, X_t is normally distributed according to $\mathrm{N}(0, H_t)$. The conditional likelihood for one single observation X_t is then

$$L_\theta(X_t|\{X_{t-1}, \ldots, X_{t-q}, H_{t-1}, \ldots, H_{t-p}\}) = \frac{1}{\sqrt{2\pi H_t}} e^{-X_t^2/2H_t}$$

where

$$H_t = \alpha_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \sum_{j=1}^{q} \alpha_j X_{t-j}^2$$

and with a parameter vector

$$\theta = (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)$$

The overall likelihood of all observations together is, in consequence of the independence of $\{\varepsilon_t\}$, merely the product

$$L_\theta = \prod_{t=1}^{T} L_\theta(X_t|\{X_{t-q}, \ldots, X_{t-1}, H_{t-p}, \ldots, H_{t-1}\}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi H_t}} e^{-X_t^2/2H_t}$$

Observe that for the likelihood of the first data point X_1 further data points

$$\{X_0, X_{-1}, \ldots, X_{-q+1}, H_0, H_{-1}, \ldots, H_{-p+1}\}$$

are required in advance. The total required data sequence $\{X_t\}$ thus encompasses $T + q$ data points. If $T + q$ observations of X_t are available, the first are required as information in advance, the remaining T are included in the likelihood function as observed data. In addition the values $\{H_0, H_{-1}, \ldots, H_{-p+1}\}$ are required as information in advance. In choosing the size of T it is necessary to make a compromise between the exactness of the estimator (T is chosen to be as large as possible) and the time scale with which the market mechanisms change (T is chosen to be as small as possible).

Maximizing this likelihood function by allowing the parameter values in θ to vary, we obtain the parameters which, under the model assumption (a GARCH(p, q) process), maximize the probability of a realization of the market values $\{X_t\}$ observed. It is again easier to work with the log-likelihood function in determining this maximum. Since the log function is strictly monotone increasing, the maximum of the likelihood and the log-likelihood function is attained at the same parameter point.

The log-likelihood for the GARCH(p, q) process is given by

$$\begin{aligned}\mathcal{L}_\theta &= \sum_{t=1}^{T} \ln L_\theta(X_t | \{X_{t-q}, \ldots, X_{t-1}, H_{t-p}, \ldots, H_{t-1}\}) \\ &= \sum_{t=1}^{T} \ln\left(\frac{1}{\sqrt{2\pi H_t}} e^{-X_t^2/2H_t}\right) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(H_t) - \frac{1}{2}\sum_{t=1}^{T}\frac{X_t^2}{H_t} \end{aligned} \tag{30.23}$$

where

$$H_t = H_t(\theta) = \alpha_0 + \sum_{j=1}^{p} \beta_j H_{t-j} + \sum_{k=1}^{q} \alpha_k X_{t-k}^2$$

This is the function which must now be maximized through the variation of the parameter vector θ. The space of valid parameters θ is limited by the constraints stated in Equations 30.14 and 30.15. This represents an additional difficulty for the optimization. The optimization is quite difficult because (as opposed to the AR(1) process) maximizing the likelihood function cannot be computed analytically but must be accomplished by means of a *numerical optimization* procedure. As the function to be maximized has multiple local maxima, a complex "likelihood surface" further complicates the optimization process since *local* optimization methods, such as gradient methods, are unsuitable if the initial value is not well chosen, i.e. if it does not lie close to the global maximum. A suitable algorithm for finding a *global* maximum in such a situation is *simulated annealing*.

30.2.3 Simulated Annealing

Simulated annealing is a numerical algorithm used to find a *global* minimum or maximum of a given function. Its construction is motivated by an effect observed in physics, namely cooling. The cooling of a physical body results in its moving through decreasing energy states traveling a path ending in a state of *minimum energy*. The simulated annealing algorithm attempts to imitate this process. The function whose minimum is to be found thus corresponds to the energy of the physical body.

As a physical body cools, the temperature T declines resulting in a steady loss of energy. The body is composed of billions of atoms which all make a contribution to its total energy. This being the case, there are a multitude of possible energy states with a multitude of *local* energy minima. If the temperature declines very *slowly*, the body surprisingly finds its *global* minimum (for example, the atoms in the body may assume a characteristic lattice configuration). A simple approach to this can be taken from *thermodynamics*: the probability of a body being in a state with energy E when the temperature of the body is T is proportional to the *Boltzmann factor,* $\exp(-E/kT)$:

$$P(E) \sim \exp\left(-\frac{E}{kT}\right)$$

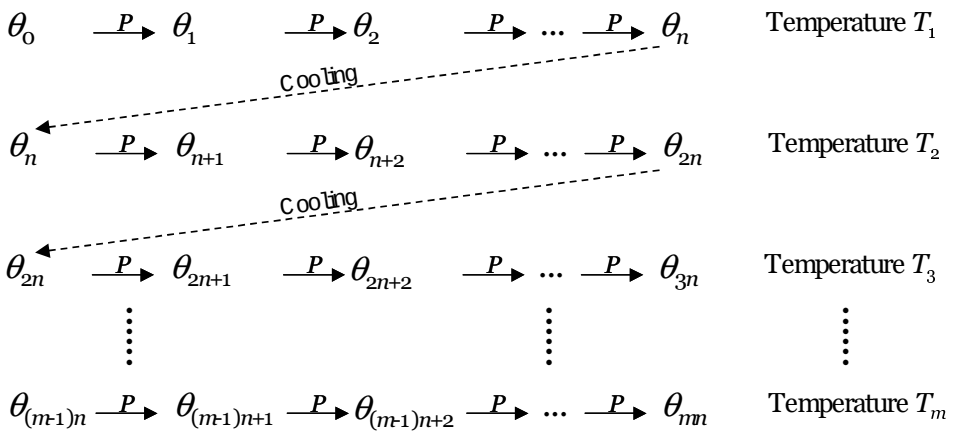


Figure 30.3: Simulated annealing using m Markov chains with n steps in each chain. If the cooling is slow enough and m and n are large enough, then θ_{mn} is a good approximation of the parameter vector necessary to achieve the global minimum of the function f.

where k is a thermodynamic constant, the *Boltzmann constant*. It follows that a higher energy state can be attained at a certain temperature though the probability of such an event declines with a decline in temperature. In this way, "unfavorable" energy states *can* be attained and thus the system *can* escape from local energy minima. However, if the temperature drops too quickly, the body remains in a so-called meta-stable state and cannot reach its global energy minimum[7]. It is therefore of utmost importance to cool the body *slowly*.

This strategy observed in nature is now to be simulated on a computer. In order to replicate the natural scenario, a configuration space (the domain of possible values of the pertinent parameters θ) must be defined. This might be a connected set but could also consist of discrete values (*combinatorial optimization*). In addition, a mechanism is required governing the transition from one configuration to another. And finally, we need a scheme for the cooling process controlling the decline in "temperature" T ($T_0 \to T_1 \to \cdots \to T_n \to \cdots$). The last two points mentioned are of particular importance; the change-of-configuration mechanism determine how efficient the configuration space is sampled while the second of the above requirements serves to realize the "slow cooling".

For each temperature the parameter sequence forms a *Markov chain*. Each new test point θ is accepted with the probability[8]

$$P = \min\left\{e^{-[f(\theta_p)-f(\theta_{p-1})]/T}, 1\right\}$$

[7]Physicists speak in such cases of "frustrated" systems. An example of such a frustrated system is glass.

[8]The minimum function is only required since a probability can be at most equal to one.

where θ_{p-1} represents the previously accepted parameter configuration. The function f is the function to be minimized for each specific problem and is, for example, the (negative) log-likelihood function from Equation 30.23. This function corresponds to the energy function in physics.

After having traveled a certain number of steps in the Markov chain, the temperature declines according to some mechanism which could for instance be as simple as

$$T_n = \alpha T_{n-1} \quad (0 < \alpha < 1)$$

A new Markov chain is then started. The starting point for the new chain is the end point of the previous chain. In a concrete optimization, the temperature is naturally not to be understood in the physical sense; it is merely an abstract parameter directing the course of the optimization by controlling the transition probability in the Markov chain. However, we choose to retain the designations temperature or cooling scheme as a reminder of the procedure's origin. Figure 30.3 shows a schematic representation of the algorithm.

Simulated annealing is demonstrated in the Excel workbook GARCH.XLS by means of a VBA program. The algorithm in the workbook is used to fit the parameters of a GARCH(1,1) process making use of the first 400 points of a given (simulated) data set. No emphasis is placed on the speed of computation since our object is to demonstrate the fundamental principles as clearly as possible.

31

Forecasting with Time Series Models

Having selected a model and fitted its parameters to a given times series, the model can then be used to estimate new data of the time series. If such data are estimated for a time period following the final data value X_T of the given time series, we speak of a *prediction* or *forecast*. The estimation of data lying between given data points is called *interpolation*. The question now arises as to how a model such as those given in Equations 30.6 or 30.13 could be used to obtain an "optimal" estimate. To answer this question the *forecasting error*

$$X_{T+k} - \widehat{X}_{T+k} \ , \ \ k \in \mathbb{N}$$

between the estimated values $\widehat{X}_{T+k}$ and the actual observed time series values X_{T+k} can be used if the last value used in the calibration of the model was X_T. The best forecast is that which minimizes the *mean square error* (*MSE* for short). The MSE is defined as the expectation of the squared forecasting error

$$\text{MSE} := \text{E}[(X_{T+k} - \widehat{X}_{T+k})^2] \tag{31.1}$$

This expression is the mathematical formulation of the intuitive concept of the "distance" between the estimated and the actual values which is to be minimized "on average" (more cannot be expected when dealing with random variables). Minimizing this mean square error yields the result that the best forecasting estimate (called the *optimal forecast*) is given by the conditional expectation

$$\widehat{X}_{T+k} = \text{E}[X_{T+k}|X_T, \ldots, X_2, X_1] \tag{31.2}$$

This is the expectation of X_{T+k}, conditional on all available information about the time series up to and including T.

In practice, however, the concrete computation of this conditional expectation is generally very difficult since the joint distribution of the random variables must be known. Therefore, we often limit our consideration to the *linear forecast*

$$\widehat{X}_{T+k} = u_1 X_1 + u_2 X_2 + \cdots + u_T X_T \tag{31.3}$$

with appropriate coefficients u_i. This linear forecast is, in contrast to the optimal forecast, often more easily interpreted. For the special case that the $\{X_t\}$ are normally distributed, the linear forecast and the optimal forecast agree. The *best* linear forecast can be characterized by the fact that the forecasting error $X_{T+k} - \widehat{X}_{T+k}$ and the $X_1, X_2, \ldots, X_T$ are uncorrelated. The intuitive interpretation is that the $X_1, X_2, \ldots, X_T$ cannot provide any additional information for the forecast and the error is thus purely random.

31.1 Forecasting with Autoregressive Models

This forecasting procedure will now be applied to an AR(p) process, Equation 30.6. The optimal one-step forecast is, according to Equation 31.2, given directly by the conditional expectation in Equation 30.7

$$\widehat{X}_{T+1}^{\text{optimal}} = \mathrm{E}[X_{T+1}|X_T, \ldots, X_1] = \sum_{i=1}^{p} \phi_i X_{T+1-i} \tag{31.4}$$

This has the form indicated in Equation 31.3. The optimal one-step forecast is thus the *best linear* one-step forecast. Equation 30.6 shows that the forecasting error $X_{T+1} - \widehat{X}_{T+1}$ is precisely ε_{T+1} and independent of $X_1, X_2, \ldots, X_T$. The MSE for the one-step forecast is given by Equation 31.1 with $k = 1$. Thus,

$$\mathrm{MSE} = \mathrm{E}[(X_{T+1} - \widehat{X}_{T+1})^2] = \mathrm{E}[\varepsilon_{T+1}^2] = \sigma^2$$

The optimal two-step forecast is the conditional expectation of X_{T+2} on the basis of knowledge of $X_T, \ldots, X_1$, as can be seen from Equation 31.2:

$$\widehat{X}_{T+2}^{\text{optimal}} = \mathrm{E}[X_{T+2}|X_T, \ldots, X_1]$$

This is *not* the equivalent to the conditional expectation of X_{T+2} on the basis of knowledge of $X_{T+1}, \ldots, X_1$. Hence, Equation 30.7 cannot be applied directly. An additional difficulty arises due to the fact that X_{T+1} is unknown. The *optimal* two-step forecast cannot be calculated. We therefore proceed by computing the *linear* two-step forecast. The best linear two-step forecast is obtained by actually calculating the conditional expectation of X_{T+2} as if all the $X_{T+1}, \ldots, X_1$ were known and replacing the (unknown) value X_{T+1} by its best estimate $\widehat{X}_{T+1}$ (which was calculated in the previous step):

$$\widehat{X}_{T+2}^{\text{linear}} = \mathrm{E}[X_{T+2}|\widehat{X}_{T+1}, X_T, \ldots, X_1]$$

Now Equation 30.7 can be applied to this conditional expectation and utilizing Equation 31.4 we obtain

$$\begin{aligned}
\widehat{X}_{T+2}^{\text{linear}} &= \phi_1 \widehat{X}_{T+1} + \sum_{j=2}^{p} \phi_j X_{T+2-j} \\
&= \phi_1 \sum_{i=1}^{p} \phi_i X_{T+1-i} + \sum_{j=2}^{p} \phi_j X_{T+1-(j-1)} \\
&= \phi_1 \sum_{i=1}^{p} \phi_i X_{T+1-i} + \sum_{i=1}^{p-1} \phi_{i+1} X_{T+1-i} \\
&= \phi_1 \phi_p X_{T+1-p} + \sum_{i=1}^{p-1} \left[\phi_1 \phi_i + \phi_{i+1}\right] X_{T+1-i}
\end{aligned}$$

The linear two-step forecast then has the form indicated in Equation 31.3. The forecasting error is found to be

$$\begin{aligned}
X_{T+2} - \widehat{X}_{T+2} &= \underbrace{\sum_{i=1}^{p} \phi_i X_{T+2-i} + \varepsilon_{T+2}}_{X_{T+2}\text{ , see Eqn. 30.6}} - \phi_1 \widehat{X}_{T+1} - \sum_{j=2}^{p} \phi_j X_{T+2-j} \\
&= \varepsilon_{T+2} + \phi_1 (X_{T+1} - \widehat{X}_{T+1}) \\
&= \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}
\end{aligned}$$

Thus, the forecasting error is a sum of two normally distributed random variables and therefore itself normally distributed. The MSE can now be computed as follows:

$$\begin{aligned}
\text{MSE} &= \text{E}[(X_{T+2} - \widehat{X}_{T+2})^2] = \text{E}[(\varepsilon_{T+2} + \phi_1 \varepsilon_{T+1})^2] \\
&= \text{E}[\varepsilon_{T+2}^2] + \phi_1^2 \text{E}[\varepsilon_{T+1}^2] + 2\phi_1 \text{E}[\varepsilon_{T+2} \varepsilon_{T+1}] \\
&= \text{var}[\varepsilon_{T+2}^2] + \phi_1^2 \text{var}[\varepsilon_{T+1}^2] + 2\phi_1 0 \\
&= \sigma^2 (1 + \phi_1^2)
\end{aligned}$$

This implies that the forecasting error $X_{T+2} - \widehat{X}_{T+2}$ of the two-step forecast is normally distributed with variance $\sigma^2(1 + \phi_1^2)$, i.e. $\text{N}(0, \sigma^2(1 + \phi_1^2))$.

Proceeding analogously, the best linear h-step forecast is obtained by taking the conditional expectation of X_{T+h} as if all X_t were known up to X_{T+h-1} and then replacing the yet unknown values of X_t for $T < t < h$ with their best estimators calculated inductively in previous steps as described above:

$$\widehat{X}_{T+h}^{\text{linear}} = \text{E}[X_{T+h} | \widehat{X}_{T+h-1}, \widehat{X}_{T+h-2}, \ldots, \widehat{X}_{T+1}, X_T, \ldots, X_1]$$

Equation 30.7 is then applied to these conditional expectations resulting in

$$\widehat{X}_{T+h} = \sum_{i=1}^{\min(h-1,\ p)} \phi_i \widehat{X}_{T+h-i} + \sum_{j=h}^{p} \phi_j X_{T+h-j}$$

The forecasting error of the h-step forecast is

$$\begin{aligned}
&X_{T+h} - \widehat{X}_{T+h} \\
&= \underbrace{\sum_{i=1}^{p} \phi_i X_{T+h-i} + \varepsilon_{T+h}}_{X_{T+h},\ \text{see Eqn. 30.6}} - \sum_{i=1}^{\min(h-1,\ p)} \phi_i \widehat{X}_{T+h-i} - \sum_{j=h}^{p} \phi_j X_{T+h-j} \\
&= \varepsilon_{T+h} + \sum_{i=1}^{\min(h-1,\ p)} \phi_i X_{T+h-i} + \sum_{j=h}^{p} \phi_i X_{T+h-j} \\
&\quad - \sum_{i=1}^{\min(h-1,\ p)} \phi_i \widehat{X}_{T+h-i} - \sum_{j=h}^{p} \phi_j X_{T+h-j} \\
&= \varepsilon_{T+h} + \sum_{i=1}^{\min(h-1,\ p)} \phi_i \left(X_{T+h-i} - \widehat{X}_{T+h-i} \right)
\end{aligned}$$

This is a recursion expressing the h-step forecasting error in terms of the forecasting errors for fewer than h steps. From this it can be shown that the h-step forecasting error is distributed as $\mathrm{N}(0, \sigma^2(1 + \phi_1^2 + \cdots + \phi_{h-1}^2))$.

The unknown coefficients $\phi_1, \phi_2, \ldots, \phi_p$ are estimated from the time series as shown in Section 30.2.1. The ϕ_i in the forecast equation are simply replaced with $\widehat{\phi}_i$.

31.2 Volatility Forecasts with GARCH(p, q) Processes

GARCH models of the form indicated in Equation 30.13 are not suitable for the prediction of the actual values X_i of a time series since the random variable in Equation 30.13 appears as a *product* (rather than a sum as in Equation 30.6). In consequence, the conditional expectations of the X_i are identically zero. However, GARCH models are well adapted for forecasting the (conditional) *variance* of time series values. According to Equation 31.2, the conditional expectation is in general the optimal forecast. We are therefore looking for the *conditional expectation of the conditional variance.*

31.2.1 Forecast Over Several Time Steps

The One-Step Forecast

Equation 30.16 shows that the conditional variance of X_T is equal to H_T if all X_t for $t \leq T-1$ are known. Its conditional expectation is then the conditional expectation of H_T. Based on Equation 30.17, the conditional expectation of H_T is simply H_T itself if the X values are known up to the time $T-1$. Hence, the *optimal* one-step forecast

for the conditional variance is[1]

$$\begin{aligned}\widehat{\text{var}}_{T+1}^{\text{optimal}} &= \text{E}[\text{var}_{T+1}|X_T, \ldots, X_1] \\ &= \text{E}[H_{T+1}|X_T, \ldots, X_1] \\ &= H_{T+1} \\ &= \alpha_0 + \sum_{j=1}^{q} \alpha_j X_{T+1-j}^2 + \sum_{i=1}^{p} \beta_i H_{T+1-i} \end{aligned} \tag{31.8}$$

To clarify the argument used in the derivation of this result, each of the equalities in Equation 31.8 will receive somewhat more scrutiny. The first equation is obtained from the general forecast equation, Equation 31.2. The second follows from Equation 30.16. The third holds as a result of Equation 30.17 while the fourth equation is derived from Equation 30.13 used in the construction of the GARCH process.

The Two-Step Forecast

The two-step forecast is somewhat more complicated. The optimal two-step forecast is, according to Equation 31.2, the conditional expectation of var_{T+2} under the condition that $X_T, \ldots, X_1$ are known:

$$\widehat{\text{var}}_{T+2}^{\text{optimal}} = \text{E}[\text{var}_{T+2}\,|X_T, \ldots, X_1]$$

Again, this is *not* equal to the conditional expectation of var_{T+2} under the condition that $X_{T+1}, \ldots, X_1$ are known. Equation 30.16 cannot be applied directly. Indeed, the *optimal* two-step forecast cannot be computed. We calculate instead, analogously to the linear forecast of the AR(p) process illustrated in Section 31.1, the best possible two-step forecast by replacing the expectation of var_{T+2} conditional upon $X_T, \ldots, X_1$ with the conditional expectation of var_{T+2} as if the $X_{T+1}, \ldots, X_1$ were all known:

$$\widehat{\text{var}}_{T+2} = \text{E}[\text{var}_{T+2}\,|X_{T+1}, \ldots, X_1]$$

Now Equation 30.16 can be applied to obtain

$$\widehat{\text{var}}_{T+2} = \text{E}[\text{var}_{T+2}\,|X_{T+1}, \ldots, X_1] = H_{T+2}$$

Remember however, that X_{T+1} is *not* known and therefore H_{T+2} appearing here is not known at time T. The best we can do is to replace H_{T+2} by its optimal estimator which, according to Equation 31.2, is given by its conditional expectation

$$\widehat{\text{var}}_{T+2} = \text{E}[H_{T+2}|X_T, \ldots, X_1] \tag{31.11}$$

[1] In order to keep the notation as simple as possible, we will adopt the convention of denoting the conditional variance by $\text{var}_{T+h} := \text{var}[X_{T+h}|X_T, \ldots, X_1]$, likewise for its estimators $\widehat{\text{var}}_{T+h} := \widehat{\text{var}}[X_{T+h}|X_T, \ldots, X_1]$ for any $h > 0$. When these abbreviations for the conditional variances are used, it is always to be understood that the values $X_T, \ldots, X_1$ are known.

Inside *this* expectation, we now replace H_{T+2} in accordance with the construction in 30.13:

$$\begin{aligned}\widehat{\text{var}}_{T+2} &= \text{E}[H_{T+2}|X_T,\ldots,X_1] \\ &= \alpha_0 + \sum_{j=1}^{q}\alpha_j \text{E}[X_{T+2-j}^2|X_T,\ldots,X_1] + \sum_{i=1}^{p}\beta_i \text{E}[H_{T+2-i}|X_T,\ldots,X_1] \\ &= \alpha_0 + \alpha_1 \text{E}[X_{T+1}^2|X_T,\ldots,X_1] + \sum_{j=2}^{q}\alpha_j X_{T+2-j}^2 + \sum_{i=1}^{p}\beta_i H_{T+2-i} \qquad (31.12)\end{aligned}$$

In the last step,we have exploited the fact that all X_t are known for all times $t \leq T$ and, according to Equation 30.17, all of the H_t for the times $t \leq T+1$. The expectation of the known quantities can be replaced by the quantities themselves. Only one unknown quantity remains, namely X_{T+1}^2. Because of Equation 30.16, the conditional expectation of all of the X_{T+h} for all $h > 0$ is always equal to zero. The expectation of X_{T+h}^2 can therefore be replaced by the variance of X_{T+h}:

$$\begin{aligned}\text{E}[X_{T+h}^2|X_T,\ldots,X_1] &= \text{E}[X_{T+h}^2|X_T,\ldots,X_1] - (\underbrace{\text{E}[X_{T+h}|X_T,\ldots,X_1]}_{0})^2 \\ &= \text{var}[X_{T+h}|X_T,\ldots,X_1] \quad \text{for every } h > 0 \qquad (31.13)\end{aligned}$$

For $h = 1$ this implies

$$\text{E}[X_{T+1}^2|X_T,\ldots,X_1] = \text{var}[X_{T+1}|X_T,\ldots,X_1] = H_{T+1}$$

where Equation 30.16 has again been used in the last step. The two-step forecast then becomes

$$\widehat{\text{var}}_{T+2} = \alpha_0 + \alpha_1 H_{T+1} + \sum_{j=2}^{q}\alpha_j X_{T+2-j}^2 + \sum_{i=1}^{p}\beta_i H_{T+2-i} \qquad (31.14)$$

The Three-Step Forecast

For the two-step forecast, only the value of X_{T+1} in Equation 31.12 was unknown, the necessary H values were known up to time $T+1$. This it no longer the case in the three-step forecast. In this case, some of the H values are also unknown. Because of this additional difficulty, it is advisable to demonstrate the computation of a three-step forecast before generalizing to arbitrarily many steps.

The three-step forecast now proceeds analogous to the two-step forecast: the optimal forecast is, as indicated in Equation 31.2, the conditional expectation of X_{T+3} under the condition that $X_T,\ldots,X_1$ are known.

$$\widehat{\text{var}}_{T+3}^{\text{optimal}} = \text{E}[\text{var}_{T+3}\,|X_T,\ldots,X_1]$$

Again, Equation 30.16 cannot be directly applied since the X are only known up to X_T and not up to X_{T+2}. The best possible three-step forecast is thus, analogous to Equation 31.11

$$\widehat{\text{var}}_{T+3} = \text{E}[H_{T+3}|X_T,\ldots,X_1] \qquad (31.16)$$

In this expectation we now replace H_{T+3} with its expression constructed in Equation 30.13 to obtain

$$\begin{aligned}\widehat{\text{var}}_{T+3} &= \text{E}[H_{T+3}|X_T, \ldots, X_1] \\ &= \alpha_0 + \sum_{i=1}^{p} \beta_i \text{E}[H_{T+3-i}|X_T, \ldots, X_1] + \sum_{j=1}^{q} \alpha_j \text{E}[X^2_{T+3-j}|X_T, \ldots, X_1] \\ &= \alpha_0 + \sum_{i=2}^{p} \beta_i H_{T+3-i} + \beta_1 \text{E}[H_{T+2}|X_T, \ldots, X_1] \\ &\quad + \alpha_1 \text{E}[X^2_{T+2}|X_T, \ldots, X_1] + \alpha_2 \text{E}[X^2_{T+1}|X_T, \ldots, X_1] + \sum_{j=3}^{q} \alpha_j X^2_{T+3-j} \quad (31.17)\end{aligned}$$

In the last step, the expectations of the known values were again replaced by the values themselves (all X_t for times $t \leq T$ and all H_t for times $t \leq T + 1$). Only three unknown values remain, namely X^2_{T+1}, X^2_{T+2} and H_{T+2}. For the conditional expectation of X^2_{T+1} and X^2_{T+2} we can use Equation 31.13 to write

$$\begin{aligned}\text{E}[X^2_{T+1}|X_T, \ldots, X_1] &= \text{var}[X_{T+1}|X_T, \ldots, X_1] = H_{T+1} \\ \text{E}[X^2_{T+2}|X_T, \ldots, X_1] &= \text{var}[X_{T+2}|X_T, \ldots, X_1] = \widehat{\text{var}}_{T+2}\end{aligned}$$

Equation 30.16 has been used in the first of the above two equations. In the second equation, this is not possible since taking the conditional variance at time $T + 2$ under the condition that $X_T, \ldots, X_1$ are known is not the same as taking it conditional upon knowing the values of $X_{T+1}, \ldots, X_1$. We have no other choice than to replace the unknown $\text{var}[X_{T+2}|X_T, \ldots, X_1]$ with the (previously calculated) estimator $\widehat{\text{var}}_{T+2}$.

For the expectation $\text{E}[H_{T+2}|X_T, \ldots, X_1]$ we make use of the fact that, according to Equation 31.11, it is equal to the two-step forecast for the variance

$$\text{E}[H_{T+2}|X_T, \ldots, X_1] = \widehat{\text{var}}_{T+2}$$

Substituting all this into Equation 31.17 finally yields

$$\widehat{\text{var}}_{T+3} = \alpha_0 + (\alpha_1 + \beta_1)\, \widehat{\text{var}}_{T+2} + \alpha_2 H_{T+1} + \sum_{j=3}^{q} \alpha_j X^2_{T+3-j} + \sum_{i=2}^{p} \beta_i H_{T+3-i} \quad (31.18)$$

The Forecast for h Steps

The generalization to the forecast for an arbitrary number of steps h is now quite simple. Analogous to Equations 31.11 and 31.16 the best possible estimate is

$$\widehat{\text{var}}_{T+h} = \text{E}[H_{T+h}|X_T, \ldots, X_1] \quad \text{for every } h > 0 \quad (31.19)$$

Within this expectation, we now replace H_{T+h} as in the construction Equation 30.13 and obtain an equation analogous to Equation 31.17

$$\begin{aligned}\widehat{\text{var}}_{T+h} &= \text{E}[H_{T+h}|X_T,\ldots,X_1] \\ &= \alpha_0 + \sum_{j=1}^{q}\alpha_j \text{E}[X^2_{T+h-j}|X_T,\ldots,X_1] + \sum_{i=1}^{p}\beta_i \text{E}[H_{T+h-i}|X_T,\ldots,X_1] \\ &= \sum_{j=1}^{\min(h-1,\ q)}\alpha_j \text{E}[X^2_{T+h-j}|X_T,\ldots,X_1] + \sum_{j=h}^{q}\alpha_j X^2_{T+h-j} + \alpha_0 \\ &+ \sum_{i=1}^{\min(h-2,\ p)}\beta_i \text{E}[H_{T+h-i}|X_T,\ldots,X_1] + \sum_{i=h-1}^{p}\beta_i H_{T+h-i}\end{aligned}$$

In the last step, the expectation of the known values have again been replaced by the values themselves (all X_t for times $t \leq T$ and all H_t for times $t \leq T+1$). The remaining expectations of the Hs are replaced according to Equation 31.19 through the respective variance estimators. We again use Equation 31.13 for the conditional expectation of X^2 and write

$$\begin{aligned}\text{E}[X^2_{T+1}|X_T,\ldots,X_1] &= \text{var}[X_{T+1}|X_T,\ldots,X_1] = H_{T+1} \\ \text{E}[X^2_{T+k}|X_T,\ldots,X_1] &= \text{var}[X_{T+k}|X_T,\ldots,X_1] = \widehat{\text{var}}_{T+k} \quad \text{for } k>1\end{aligned}$$

Substituting all of these relations for the conditional expectations finally yields the general h-step forecast of the conditional volatility in the GARCH(p,q) model:

$$\begin{aligned}\widehat{\text{var}}_{T+h} = &\sum_{j=1}^{\min(h-2,\ q)}\alpha_j \widehat{\text{var}}_{T+h-j} + \sum_{j=h}^{q}\alpha_j X^2_{T+h-j} + \alpha_{h-1}H_{T+1} + \alpha_0 \\ &+ \sum_{i=1}^{\min(h-2,\ p)}\beta_i \widehat{\text{var}}_{T+h-i} + \sum_{i=h-1}^{p}\beta_i H_{T+h-i}\end{aligned} \tag{31.20}$$

Together with the start value, Equation 31.8, in the form of $\widehat{\text{var}}_{T+1} = H_{T+1}$ the h-step forecast can be computed recursively for all h.

From Equations 31.19 and 31.2, the estimator for the variance is simultaneously the estimator for H, and thus

$$\widehat{H}_{T+h} = \widehat{\text{var}}_{T+h} \quad \text{for every } h>0$$

31.2.2 Forecast for the Total Variance

In the financial world, the time series $\{X_T\}$ is usually taken to represent a relative price change (yield), see for example Figure 30.1 of the FTSE data. We are therefore also interested in the variance of the return over an *entire* time period of length h, i.e. in the variance of the sum $\sum_{j=1}^{h} X_{T+j}$. In forecasts such as in Equation 31.20, we are only dealing with a prediction of the conditional variance *after* h steps and not with

a prediction of the variance *over* the entire term of h steps (from X_T to X_{T+h}). In other words, Equation 31.20 is forecasting the conditional variance of X_{T+h} alone, and not predicting the variance of the sum $\sum_{j=1}^{h} X_{T+j}$. The variance of the total return $\sum_{j=1}^{h} X_{T+j}$ for independent (in particular uncorrelated) returns is simply the sum of the variances as can be seen in Equation A.17. Since the X_t of the process in Equation 30.13 are uncorrelated (because the ε_t are *iid*), the estimator for the total variance of the GARCH process over h steps is simply

$$\widehat{\text{var}}[\sum_{j=1}^{h} X_{T+j}|X_T,\ldots,X_1] = \sum_{j=1}^{h} \widehat{\text{var}}[X_{T+j}|X_T,\ldots,X_1] = \sum_{j=1}^{h} \widehat{\text{var}}_{T+j} \tag{31.21}$$

Even in the case of weak autocorrelations between the returns in a given time series, this result holds in good approximation.

31.2.3 Volatility Term Structure

The variance of the total return over a term from T until $T + h$ is a function of this term. The square root of the (annualized) variance of the total return as a function of the term is called the *volatility term structure*. This plays an important role in pricing options since for an option with a lifetime of h, the volatility associated with this term is the relevant parameter value. From the estimator for the variance of the total return over the pertinent term, we obtain the estimator of the volatility structure as

$$\sigma(T, T+h) = \sqrt{\frac{1}{h}\widehat{\text{var}}[\sum_{i=1}^{h} X_{T+i}|X_T,\ldots,X_1]} = \sqrt{\frac{1}{h}\sum_{j=1}^{h} \widehat{\text{var}}_{T+j}} \tag{31.22}$$

31.3 Volatility Forecasts with GARCH$(1, 1)$ Processes

For the GARCH(1,1) process ($q = 1, p = 1$), all of the above estimators can be computed explicitly and the recursion equation 31.20 can be carried out. The start value of the recursion is simply

$$\widehat{\text{var}}_{T+1} = H_{T+1} = \alpha_0 + \alpha_1 X_T^2 + \beta_1 H_T \tag{31.23}$$

as can be seen from Equation 31.8. The two-step forecast as given by Equation 31.14 simplifies to

$$\widehat{\text{var}}_{T+2} = \alpha_0 + \kappa H_{T+1}$$

where we defined the abbreviation

$$\kappa := \alpha_1 + \beta_1$$

For $h > 2$ and $p = q = 1$, the upper limits in the sums over the variance estimators in the general recursion equation 31.20 are simply

$$\min(h-2,\ q) = \min(h-2,\ p) = 1 \quad \text{for } h > 2$$

Neither of the other sums makes any contribution since the lower limit in these sums is greater than the upper limit. The term $\alpha_{h-1}H_{T+1}$ likewise does not exist since $q = 1$ implies that only α_0 and α_1 exist. However, $h - 1$ is greater than 1 for $h > 2$. All things considered, the h-step forecast in Equation 31.20 reduces to

$$\widehat{\text{var}}_{T+h} = \alpha_0 + \kappa\, \widehat{\text{var}}_{T+h-1}$$

where $\kappa = \alpha_1 + \beta_1$. This recursion relation has a closed form expression in the form of a geometric series:

$$\begin{aligned}
\widehat{\text{var}}_{T+h} &= \alpha_0 + \kappa\widehat{\text{var}}_{T+h-1} \\
&= \alpha_0 + \kappa(\alpha_0 + \kappa\widehat{\text{var}}_{T+h-2}) = \alpha_0(1+\kappa) + \kappa^2\widehat{\text{var}}_{T+h-2} \\
&= \alpha_0(1+\kappa) + \kappa^2(\alpha_0 + \kappa\widehat{\text{var}}_{T+h-3}) = \alpha_0(1+\kappa+\kappa^2) + \kappa^3\widehat{\text{var}}_{T+h-3}) \\
&\dots \\
&= \alpha_0 \underbrace{\sum_{i=1}^{h-1} \kappa^{i-1}}_{\text{Geometric Series}} + \kappa^{h-1}\underbrace{\widehat{\text{var}}_{T+1}}_{H_{T+1}} \\
&= \alpha_0 \left(\frac{1-\kappa^{h-1}}{1-\kappa}\right) + \kappa^{h-1} H_{T+1} \\
&= \widetilde{\alpha}_0 + \kappa^{h-1}\left(H_{T+1} - \widetilde{\alpha}_0\right) \\
&= \widetilde{\alpha}_0 + \kappa^{h-1}\left(\alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \widetilde{\alpha}_0\right) \quad , \qquad h > 1
\end{aligned} \tag{31.25}$$

where for $\widehat{\text{var}}_{T+1}$ the start-value of the recursion H_{T+1} was used and the geometric series was calculated according to Equation 16.8. Here

$$\widetilde{\alpha}_0 := \frac{\alpha_0}{1-\alpha_1-\beta_1}$$

again denotes the *un*conditional variance from Equation 30.19. The GARCH(1,1) prediction for the conditional variance after h steps is therefore equal to the unconditional variance plus the difference between the one-step forecast and the unconditional variance dampened by the factor κ^{h-1}. The stationarity condition requiring that $\alpha_1 + \beta_1 < 1$ implies that for $h \to \infty$ (a long prediction period) the GARCH prediction converges towards the unconditional variance.

The variance of the total return $\sum_{i=1}^{h} X_{T+i}$ over a term of length h as the sum of the conditional forecasts is obtained for the GARCH(1,1) process as indicated in Equation 31.21:

$$\begin{aligned}
\widehat{\text{var}}[\sum_{i=1}^{h} X_{T+i} | X_T, \dots, X_1] &= \sum_{i=1}^{h} \widehat{\text{var}}_{T+i} \\
&= \sum_{i=1}^{h} \left[\widetilde{\alpha}_0 + \kappa^{i-1}\left(\alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \widetilde{\alpha}_0\right)\right] \\
&= h\widetilde{\alpha}_0 + \left(\frac{1-\kappa^h}{1-\kappa}\right)\left(\alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \widetilde{\alpha}_0\right)
\end{aligned} \tag{31.26}$$

where Equation 16.8 for the geometric series is used again in the last step. The volatility term structure resulting from the 31.22GARCH(1,1) process is thus

$$\begin{aligned}\sigma(T,T+h) &= \sqrt{\frac{1}{h}\widehat{\mathrm{var}}[\sum_{i=1}^{h} X_{T+i}|X_T,\ldots,X_1]} \\ &= \sqrt{\widetilde{\alpha}_0 + \frac{1}{h}\left(\frac{1-\kappa^h}{1-\kappa}\right)(\alpha_0+\beta_1 H_T+\alpha_1 X_T^2-\widetilde{\alpha}_0)}\end{aligned}$$

which approaches the unconditional variance $\widetilde{\alpha}_0$ like $1/\sqrt{h}$ for large h.

31.4 Volatility Forecasts with Moving Averages

In addition to the relatively modern GARCH models, older methods such as *moving averages* exist in the market, which, despite their obvious shortcomings, are still widely used, thanks to their simplicity. Before entering into a discussion of volatility forecasts via moving averages and comparing them with those of the GARCH models, we will first introduce the two most important varieties, the *simple* moving average, abbreviated here as *MA* and the *exponentially weighted moving average*, abbreviated as *EWMA*.

The (simple) moving average measures the conditional variance (of a time series with zero mean) simply as the sum of evenly weighted squared time series values over a time window of width b. The form for the MA corresponding to Equation 30.16 is simply

$$\mathrm{var}[X_t|X_{t-1},\ldots,X_1] = \frac{1}{b}\sum_{k=1}^{b} X_{t-k}^2$$

The well-known *phantom structures* arise from this equation because every swing in the X_t^2 is felt fully for b periods and then suddenly disappears completely when the term causing the perturbation no longer contributes to the average. An improvement would be to consider *weighted* sums where time series values further in the past are weighted less than values closer to the present. This can be realized, for example, by the exponentially weighted moving average EWMA. The conditional variance in the EWMA is

$$\begin{aligned}\mathrm{var}[X_t|X_{t-1},\ldots,X_1] &= \frac{1}{\sum_{j=1}^{b}\lambda^{j-1}}\sum_{k=1}^{b}\lambda^{k-1}X_{t-k}^2 \\ &= \frac{X_{t-1}^2+\lambda^1 X_{t-2}^2+\cdots+\lambda^{b-1}X_{t-b}^2}{1+\lambda^1+\lambda^2+\cdots+\lambda^{b-1}}\end{aligned}$$

For $\lambda < 1$ the values lying further back contribute less. The values commonly assigned to the parameter λ lie between 0.8 and 0.98. Naturally, the simple MA can be interpreted as a special case of the EWMA with $\lambda = 1$. The conditional variance of the

EWMA is very similar to that of a GARCH(1,1) process since the recursion for H_T in Equation 30.16 can be performed explicitly for $p = q = 1$ and the conditional variance of the GARCH(1,1) process becomes

$$\begin{aligned}
&\operatorname{var}[X_t|X_{t-1}, \ldots, X_1] = H_t \\
&= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 H_{t-1} \\
&= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \left[\alpha_0 + \beta_1 H_{t-2} + \alpha_1 X_{t-2}^2\right] \\
&= \alpha_0(1+\beta_1) + \alpha_1(X_{t-1}^2 + \beta_1 X_{t-2}^2) + \beta_1^2 \left[\alpha_0 + \beta_1 H_{t-3} + \alpha_1 X_{t-3}^2\right] \\
&= \alpha_0(1+\beta_1+\beta_1^2) + \alpha_1(X_{T-1}^2 + \beta_1 X_{T-2}^2 + \beta_1^2 X_{T-3}^2) + \beta_1^3 H_{t-3} \\
&\ldots \\
&= \alpha_0 \sum_{k=1}^{b} \beta_1^{k-1} + \alpha_1 \sum_{k}^{b} \beta_1^{k-1} X_{T-k}^2 + \beta_1^b H_{T-b}
\end{aligned}$$

If we now choose the parameters $\alpha_0 = 0$, $\beta_1 = \lambda$ and $\alpha_1 = \lambda(\sum_{j=1}^{b} \lambda^j)^{-1}$ then this conditional variance after b steps (apart from remainder term $\beta_1^b H_{T-b}$ which contains the influence of factors lying still further in the past) is exactly the same expression as for the EWMA. The difference between the GARCH(1,1) and EWMA models first appears clearly in variance forecasts over more than one time step.

The conditional variances presented above can be interpreted as a one-step forecast for the conditional variance. The forecast over h steps delivers nothing new for the moving averages since both the MA and EWMA are static and fail to take the time structure into consideration. They start with the basic assumption that prices are lognormally distributed with a constant volatility. This implies that

$$\widehat{\operatorname{var}}[X_{T+h}|X_T, \ldots, X_1] = \widehat{\operatorname{var}}[X_{T+1}|X_T, \ldots, X_1] = \operatorname{var}[X_{T+1}|X_T, \ldots, X_1]$$

holds for the $h-$step forecast of the MA as well as for the EWMA. As mentioned after Equation 31.20, the conditional variance *after* h steps is being forecasted and *not* the variance of the total return *over* a term of h steps. The prediction for the variance of the total return as the sum over the conditional one-step forecasts is for moving averages (MA and EWMA) simply

$$\widehat{\operatorname{var}}[\sum_{i=1}^{h} X_{T+i}|X_T, \ldots, X_1] = \sum_{i=1}^{h} \underbrace{\widehat{\operatorname{var}}[X_{T+i}|X_T, \ldots, X_1]}_{\widehat{\operatorname{var}}[X_{T+1}|X_T, \ldots, X_1]} = h \operatorname{var}[X_{T+1}|X_T, \ldots, X_1]$$

This is again the famous *square root law* for the growth of the standard deviation over time. The variance simply increases linearly over time and the standard deviation is therefore proportional to the square root of time. This leads to a static prediction of the volatility, and extrapolating, for example, daily to yearly volatilities in this way can easily result in an overestimation of the volatilities. The *volatility term structure* for the moving average is then, as expected, a constant:

$$\sigma(T, T+h) = \sqrt{\frac{1}{h}\widehat{\operatorname{var}}[\sum_{i=1}^{h} X_{T+i}|X_T, \ldots, X_1]} = \sqrt{\operatorname{var}[X_{T+1}|X_T, \ldots, X_1]}$$

In the Excel workbook GARCH.XLS , the one-step forecast of a GARCH(1,1) process, an MA with $b = 80$ and an EWMA with $b = 80$ and $\lambda = 0,95$ are presented. Furthermore, the ten-step forecast of the GARCH(1,1) process is shown. Since the time series we are dealing with is a simulated GARCH(1,1) process, the "true" volatility is known (it is the H_t from the simulated series) and direct comparison can be made with each of the various estimates. As is clearly illustrated in GARCH.XLS, the one-step GARCH(1,1) forecast (H_t with the parameters $\widehat{\alpha}_0$, $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ fitted by *simulated annealing*) produces estimates which are quite close to the true volatility. The computation of the GARCH (1,1) volatility term structure is presented in GARCH.XLS as well.

32

Principle Component Analysis

32.1 The General Procedure

In addition to the autoregressive models described above, which are used for instance in the form of GARCH models when modeling volatility, a further technique of time series analysis, called *principle component analysis* (abbreviated as *PCA*), is widely applied in the financial world. This technique is employed in the analysis of term structure evolutions, for instance. As mentioned in the introduction of Chapter 15 on term structure models, the approach described in Section 27 in which the term structure was constructed by interpolating between vertices is usually not used to model the (stochastic) *dynamics* of the term structure. Rather than modeling the interest rate at the vertices as risk factors, the stochastic evolution of the term structure is reduced to a small number of stochastic variables (one, two and sometimes three) which act as the driving factors of the entire term structure. This approach has its justification in principle component analysis. Principle component analysis is a statistical technique which extracts the statistical components from the time series which are most relevant for the dynamics of the process in order of their importance. By means of this method applied to interest rates, it can be shown that usually more than 90% of the term structure's dynamics can be ascribed to the one or two most important components.

Two other well-known statistical methods used in time series analysis are *factor analysis* [26] and *cointegration* [72]. Principle component analysis sets itself apart from these other methods by the ease with which the results can be interpreted. An extensive description of principle component analysis can be found in [101], for example.

Since the time dependence of the data is not analyzed in PCA, the remarks made in Section 30 hold for this type of time series analysis as well: principle component analysis is a suitable approach for *stationary* time series only. Time series exhibiting non-stationary behavior like trends, etc. should be first freed of their non-stationary components by, for example, estimating trends and subsequently subtracting them from the time series or by taking differences of the data with respect to time, in other words, by performing a pre-treatment as described in Section 33.1. A reasonable principle component analysis of non-stationary time series is not to be achieved without

first removing these potential non-stationary components.

In principle component analysis, the time series of not just one but *several* stochastic process are considered which may be strongly correlated. Principle component analysis is therefore an example of *multivariate* time series analysis. In financial applications, each process represents a risk factor. The situation is thus similar to that in Equation 21.20 in Section 21.4 where n risk factors were considered as well. In the case of PCA, it is not a matter of the stochastic differential equations of the risk factors, but of their historic evolution. For each risk factor, there exists an associated time series of data for times t_k, $k = 1, \dots T$. For example, the data of the relative changes (as in Equation 29.1)

$$X_i(t_k) = \ln\left(\frac{S_i(t_k + \delta t)}{S_i(t_k)}\right) \approx \frac{S_i(t_k + \delta t) - S(t_k)}{S_i(t_k)}$$
$$i = 1, \dots n \quad , \quad k = 1, \dots T$$

These time series are assumed to be stationary, i.e. pre-treatment of the data has been performed already and the non-stationary components have been removed, if this has proved to be necessary. As a typical example of such a group of n risk factors, we can consider the relative changes of the interest rates (generated from the market prices of bonds by means of bootstrapping) at the vertices of an interest rate curve. The variances and pairwise correlations of these n risk factors can be arranged in a covariance matrix $\delta\mathbf{\Sigma}$ as in Equation 21.21. The entries of the covariance matrix can be determined from historical data as described in Section 29.1.

The dimension of this matrix is obviously equal to the number of risk factors (equivalently, the number of time series) and is thus equal to n. The central idea of principle component analysis is now to reduce the dimension of the problem on the basis of its statistical structure. The reduction is achieved by transforming the variables X_i into new uncorrelated variables Y_i appearing in order of the size of their variance. This results, circumstances permitting, in a significant simplification of the subsequent analysis. It is in many cases possible to neglect statistical dependences of higher order in these new coordinates and to consider the time series of the new (transformed) variables as independent. As will be shown in the construction of the transformations, the new variables generally have different variances. In this case, it is possible to neglect those variables with relatively small variance.

This transformation of the random variables $X_1, \dots, X_n$ into the new variables $Y_1, \dots, Y_n$ is linear and thus we can write $Y_k = \sum_{i=1}^{n} \alpha_{ki} X_i$, or equivalently, in the vector notation used in Section 21.4

$$\mathbf{Y} = \alpha\mathbf{X} \quad \text{with} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \tag{32.1}$$

$$\alpha = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} = \begin{pmatrix} \left(\alpha^1\right)^T \\ \vdots \\ \left(\alpha^n\right)^T \end{pmatrix}$$

In the last step, the rows of the transformation matrix have been written in terms of the vectors α defined as follows:

$$\left(\alpha^k\right)^T := \begin{pmatrix} \alpha_{k1} & \cdots & \alpha_{kn} \end{pmatrix} \Longrightarrow \alpha^k = \begin{pmatrix} \alpha_{k1} \\ \vdots \\ \alpha_{kn} \end{pmatrix}$$

This implies that the components α_i^k of these vectors and the components α_{ki} of the matrix are related as follows:

$$\alpha_i^k = \alpha_{ki} = \left(\alpha^k\right)_i^T \quad \forall\, k, i = 1, \ldots n$$

The transformation Equation 32.1 is thus given by

$$Y_k = \left(\alpha^k\right)^T \mathbf{X} = \sum_{i=1}^{n} \alpha_{ki} X_i \quad , \quad k = 1, \ldots n \tag{32.2}$$

The coefficients of the matrix α are now chosen such that the transformed variables Y_k possess the following properties:

We require every "transformation vector" α_k (each row of the matrix α) to have a norm of 1, i.e.

$$\left(\alpha^k\right)^T \alpha^k = \sum_{i=1}^{n} \alpha_i^k \alpha_i^k = \sum_{i=1}^{n} \left(\alpha_{ki}\right)^2 \stackrel{!}{=} 1 \quad \forall\, k = 1, \ldots n \tag{32.3}$$

We now select the components α_i^1 of α^1 in such a way that the variance of the first transformed variable $Y_1 = \left(\alpha^1\right)^T \mathbf{X}$ is as large as possible (maximal) while satisfying Equation 32.3. This is an optimization problem subject to the constraint $\left(\alpha^1\right)^T \alpha_1 = 1$. It will be shown explicitly in the material below how problems of this type are treated.

Having determined α^1 (and consequently Y_1), we proceed by determining the components α_i^2 of α^2 such that the variance of the second transformed variable $Y_2 = \left(\alpha^2\right)^T \mathbf{X}$ is as large as possible (maximal), again subject to the condition that Equation 32.3 holds. *In addition*, Y_2 must be uncorrelated with the vector Y_1 already determined. This additional condition can be expressed as

$$\operatorname{cov}(Y_2, Y_1) \stackrel{!}{=} 0$$

Using Equation 32.2, we can write this covariance as

$$\begin{aligned} \operatorname{cov}(Y_2, Y_1) &= \operatorname{cov}(\left(\alpha^2\right)^T \mathbf{X}, \left(\alpha^1\right)^T \mathbf{X}) \\ &= \operatorname{cov}(\sum_{i=1}^{n} \alpha_i^2 X_i, \sum_{j=1}^{n} \alpha_j^1 X_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{2i} \alpha_{1j} \underbrace{\operatorname{cov}(X_i, X_j)}_{\delta\Sigma_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\alpha^2\right)_i^T \, \delta\Sigma_{ij} \, \alpha_j^1 \\ &= \left(\alpha^2\right)^T \delta\mathbf{\Sigma} \, \alpha^1 \end{aligned}$$

This implies that, in addition to the constraint expressed in Equation 32.3, α^2 must satisfy the condition

$$\left(\alpha^2\right)^T \delta\mathbf{\Sigma}\, \alpha^1 \stackrel{!}{=} 0 \tag{32.4}$$

Thus, the determination of α^2 (and thus of Y_2) involves an optimization problem with *two* constraints.

The remaining variables $Y_3, \ldots, Y_n$ are determined successively subject to analogous conditions: in the k^{th} step, α^k is chosen such that the variance of the k^{th} transformed variable $Y_k = \left(\alpha^k\right)^T \mathbf{X}$ is as large as possible (maximal) subject to the condition that Equation 32.3 as well as the *additional* $k-1$ conditions hold, namely that Y_k be uncorrelated with *all* of the previously determined Y_i, $i = 1, \ldots k-1$, i.e.

$$\operatorname{cov}(Y_k, Y_i) \stackrel{!}{=} 0 \quad \text{for all } i < k$$

The formulation of these $k-1$ conditions on α^i is analogous to Equation 32.4

$$\left(\alpha^k\right)^T \delta\mathbf{\Sigma}\, \alpha^i \stackrel{!}{=} 0 \quad \text{for all } i = 1, \ldots k-1 \tag{32.5}$$

Overall, k conditions must be satisfied in the k^{th} step. This decreases the maximal possible variance of Y_k from step to step, since with each step, the maximum is taken over a smaller class of vectors. It is thus not surprising that Y_1 has the greatest variance among the Y_k and that the variances of the Y_k decrease rapidly with increasing k. The largest part of the variance of the original system of n factors under consideration is "collected" in the first few Y_k. These transformed variables Y_k are therefore referred to as the *principle components* and the vectors α^k defined in the above construction as the *principle axes* of the system.

To provide the reader with a concrete example, we compute the first principle component Y_1 here. In order to do so, the variance of Y_1 conditional upon the satisfaction of Equation 32.3 is maximized. According to Equation A.12, the variance of Y_1 is

$$\begin{aligned}
\operatorname{var}(Y_1) &= \operatorname{cov}(\left(\alpha^1\right)^T \mathbf{X}, \left(\alpha^1\right)^T \mathbf{X}) = \operatorname{cov}(\sum_{i=1}^{n} \alpha_i^1 X_i, \sum_{j=1}^{n} \alpha_j^1 X_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{1i}\alpha_{1j} \underbrace{\operatorname{cov}(X_i, X_j)}_{\delta\Sigma_{ij}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\alpha^1\right)_i^T \delta\Sigma_{ij}\, \alpha_j^1 \\
&= \left(\alpha^1\right)^T \delta\mathbf{\Sigma}\, \alpha^1
\end{aligned} \tag{32.6}$$

This is now maximized subject to the constraint $\left(\alpha^1\right)^T \alpha_1 = 1$. We can take constraints into account in an optimization problem using the method of *Lagrange multipliers*[1], for example. The method of Lagrange multipliers requires the construction of the *Lagrange function* $\mathcal{L}$. This function is equal to the function to be maximized less a "zero" multiplied by a *Lagrange multiplier* λ. This zero is written in the form of the

[1] A introduction to the method of Lagrange multipliers for solving optimization problems subject to constraints can be found in [27], for instance.

constraint to be satisfied, which in our case here is $0 = \left(\alpha^1\right)^T \alpha_1 - 1$. The Lagrange function is thus given by

$$\begin{aligned}\mathcal{L}(\lambda_1) &= \underbrace{\left(\alpha^1\right)^T \delta\mathbf{\Sigma}\, \alpha^1}_{\text{To be Maximized}} - \lambda_1 \underbrace{\left[\left(\alpha^1\right)^T \alpha^1 - 1\right]}_{\text{Constraint}} \\ &= \sum_{i,j=1}^{n} \alpha_{1i} \delta\Sigma_{ij} \alpha_{1j} - \lambda_1 \sum_{i=1}^{n} \left(\alpha_{1i}\right)^2 + \lambda_1\end{aligned} \tag{32.7}$$

In order to find the optimal values α_{1i} subject to this constraint, we differentiate $\mathcal{L}$ with respect to the parameter α_{1i} and set the resulting expression equal to zero; in other words, we locate the maximum of the Lagrange function[2].

$$0 \stackrel{!}{=} \frac{\partial \mathcal{L}}{\partial \alpha_{1i}} = 2 \sum_{j}^{n} \delta\Sigma_{ij} \alpha_{1j} - 2\lambda_1 \alpha_{1i} \quad \forall\, i = 1, \ldots n$$

from which it follows

$$\sum_{j}^{n} \delta\Sigma_{ij} \alpha_{1j} - \lambda_1 \alpha_{1i} = 0 \quad \forall\, i = 1, \ldots n$$

or in the compact, matrix notation:

$$(\delta\mathbf{\Sigma} - \lambda_1 \mathbf{1})\alpha^1 = 0 \tag{32.8}$$

where, as in Equation 21.31, $\mathbf{1}$ denotes the n-dimensional identity matrix. This condition implies, however, that the covariance matrix applied to the vector α^1 has no other effect than to multiply α^1 by a *number*: $\delta\mathbf{\Sigma}\alpha^1 = \lambda_1 \alpha^1$. Equation 32.8 is therefore the *eigenvalue* equation of the covariance matrix (compare this to Equation 22.16 in Section 22.3.2). The Lagrange multiplier λ_1 must in consequence be an eigenvalue of the covariance matrix and α^1 is the associated *eigenvector*. As discussed in detail in Section 22.3.2, the eigenvalue equation in 32.8 has a non-trivial solution $\alpha^1 \neq \mathbf{0}$ if and only if the matrix $(\delta\mathbf{\Sigma} - \lambda_1 \mathbf{1})$ is singular. For this to be the case, its determinant must be equal to zero

$$\det\left(\delta\mathbf{\Sigma} - \lambda_1 \mathbf{1}\right) = 0$$

The eigenvalue λ_1 is the solution to this determinant equation. Having determined λ_1, it can be substituted into Equation 32.8 to calculate the eigenvector α^1.

The eigenvalue λ_1 has another important intuitive interpretation which can be seen immediately if we multiply both sides of Equation 32.8 on the left by $\left(\alpha^1\right)^T$

[2]Since the Lagrange function differs from the function which we actually wish to maximize only by a zero, the maximum of the Lagrange function coincides with the desired maximum. However, this is only the case if the difference between the two functions is indeed zero, i.e. only when the side condition is satisfied. This is the essential idea of the method of Lagrange multipliers.

$$\begin{aligned}
\lambda_1 \alpha^1 &= \delta\mathbf{\Sigma}\alpha^1 \\
\left(\alpha^1\right)^T \lambda_1 \alpha^1 &= \left(\alpha^1\right)^T \delta\mathbf{\Sigma}\alpha^1 \\
\lambda_1 \underbrace{\left(\alpha^1\right)^T \alpha^1}_{1} &= \underbrace{\left(\alpha^1\right)^T \delta\mathbf{\Sigma}\alpha^1}_{\mathrm{var}(Y_1)}
\end{aligned}$$

Comparison of this with Equation 32.6 reveals that λ_1 is the variance of the principle component Y_1.

Analogously (i.e. maximization of the variances of the new variables Y_k), we proceed with the remaining principle components. As already mentioned above, one more constraint (see Equation 32.5) must be taken into account with each further step. Analogous to Equation 32.7, for each constraint a "zero" (in the form of the constraint multiplied by a Lagrange multipliers $\lambda_i, i = 1, \ldots k$) is subtracted from the variance of Y_k and the resulting Lagrange function $\mathcal{L}$ is maximized. In the computation of the k^{th} principle component, k Lagrange multipliers appear in the expression for the Lagrange function. The complexity of the corresponding computation increases substantially with each new constraint. We thus refrain from explicitly performing the computation here. We wish to remark, however, on the following properties holding for all principle components as was shown for Y_1:

The Lagrange multipliers λ_k can be shown to be the eigenvalues of the covariance matrix $\delta\mathbf{\Sigma} = \mathrm{cov}(X, X)$ appearing in decreasing order. The row vectors of the transformation matrix α, i.e. the α^k, are the associated eigenvectors.

This result paves the way for the implementation of the principle component analysis: we do not take the optimization problem as the starting point for our analysis, but rather we compute the eigenvalues and eigenvectors of the covariance matrix directly. In a second step, the eigenvalues λ_k and their associated eigenvectors α^k are ordered according to the size of the eigenvalues, thereby constructing the transformation matrix from the ordered row vectors α^k as defined in Equation 32.1:

$$\alpha = \begin{pmatrix} \left(\alpha^1\right)^T \\ \vdots \\ \left(\alpha^n\right) \end{pmatrix}$$

Now, the original data X_i in the time series of the n risk factors are transformed into the new variables Y_i by applying the matrix α; this is done for each observation in the time series

$$Y_j(t_k) = \sum_{i=1}^{n} \alpha_{ji} X_i(t_k) \quad \text{for all } j = 1, \ldots n \quad \text{and all} \quad k = 1, \ldots T$$

Finally, those components Y_k of the transformed data are neglected if the associated eigenvalues (and thus the variances[3] of the new random variables Y_k) are small. It

[3]The k^{th} eigenvalue is in fact equal to the variance of the new random variable Y_k.

is often the case that the number of time series in the new variables Y_k necessary for further investigation can in this way be reduced to just two or three.

In many software packages, for example SAS, Matlab, or IDL, principle component analysis is provided as one of the few multivariate time series analysis procedures or at least modules exist allowing for its construction. The steps listed above are automated in these packages; only the original time series are needed as input. The covariance matrix is estimated from this data. The output is then obtained in the form of the eigenvalues of the covariance matrix, the associated eigenvectors, i.e. the transformation matrix, and the time series of the principle components. The reduction in dimension is achieved in that the data is reduced by $n-l-1$ variables $Y_{l+1}, \ldots, Y_n$. Here, l is chosen conditional upon the relative size of the eigenvalues.

32.2 Principle Component Analysis of the German Term Structure

In this section, we conduct a principle component analysis on real interest rate data. The aim of our investigation is to determine the typical drivers of the interest rate term structure and the relative contribution of each individual driver to the total dynamics of the interest rate curve. The results of this section are relevant in several different respects. Firstly, the investigation introduced here will provide a plausible explanation for modeling an entire interest rate curve in very low dimensional spaces as is done in modern one, two, and three-factor term structure models although the term structure is clearly constructed with a larger number of vertices. Secondly, *stress scenarios* for typical and statistically independent movements in the interest rate curve can be identified.

The time series investigated are the yields for ten vertices with terms between one and ten years for the German term structure in monthly time steps spanning over a period of ten years. The data were subjected to a principle component analysis as described in the above section *without* having first been pre-treated by taking time-differences[4]. We find that the variances λ_i of the principle components Y_i, arranged in decreasing order, decrease *very* quickly. The proportion of the variances λ_1 and λ_2 of the first and second principle components with respect to the total variance $\sum_{i=1}^{10} \lambda_i$ is

$$\frac{\lambda_1}{\sum_{i=1}^{10} \lambda_i} \approx 96\% \quad , \quad \frac{\lambda_2}{\sum_{i=1}^{10} \lambda_i} \approx 3\%$$

respectively. The proportion of the third and the fourth principle components is only

$$\frac{\lambda_3}{\sum_{i=1}^{10} \lambda_i} \approx 1\% \quad , \quad \frac{\lambda_4}{\sum_{i=1}^{10} \lambda_i} \ll 1\%$$

[4]It has already been pointed out that principle component analysis assumes that the data in the time series are stationary. For the following investigation, this assumption is made keeping in mind that the results of the investigation should convey only a qualitative impression of term structure dynamics. Using the data directly (i.e. without a pre-treatment such as taking time differences, etc.) will simplify the *interpretation* of the results substantially.

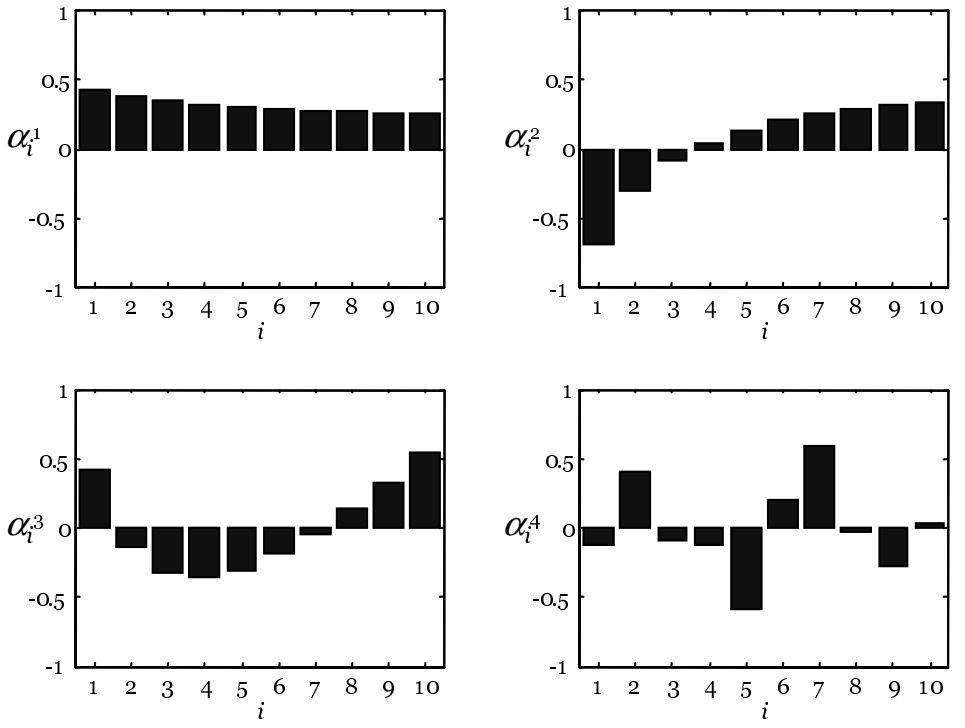


Figure 32.1: The components of the first four eigenvectors $\alpha^1, \ldots, \alpha^4$. For each of these eigenvectors the components $\alpha_1^i, \ldots, \alpha_{10}^i$ are shown. The simple structure facilitates an intuitive interpretation of the associated principal components $Y_1, \ldots, Y_4$.

respectively. The variances of the further principle components disappear almost completely, lying well below one tenth of one percent. Obviously, the dynamics of the term structure can be described by just a very *few* variables. Indeed, this result provides an excellent motivation for modeling the term structure in spaces of small dimension (or even in *one*-dimensional spaces as in Section 15). In Figure 32.1, the first four of ten eigenvectors are presented. These eigenvectors can be interpreted quite easily.

The first eigenvector α^1 weights the interest rates of all terms approximately equally. The eigenvector normalization allows the first principle component Y_1 to be interpreted as the mean interest rate level. The fluctuations of this principle component Y_1 thus represent *parallel shifts of the entire term structure.*

The second eigenvector α^2 weights the short term interest rates negatively and the long term rates positively. Considering the inverse transformation of the principle components to the original interest rate vectors, we can draw conclusions as to the interpretation of the second principle component Y_2: adding the second component to the first has the effect of adjusting the mean interest rate level by a mean slope. A change in this principle component Y_2 thus changes the mean *slope* of the term

structure, or in other words, effects a *rotation of the term structure.*

The third eigenvector α^3 can be interpreted analogously. This vector weights the short and long-term rates positively, interest rates for the terms of intermediate length negatively. The addition of the associated principle component Y_3 to the first two thus effects a change in the mean *curvature of the term structure.* The fourth principle axis α^4 shows a periodic change in sign. With the associated principle component, periodic structures in the term structure can be represented such as those described in [136], for example.

Several practical conclusions for the analysis of scenarios commonly used in risk management can be drawn from the principle component analysis. Many of the common scenarios used to model a change in the term structure can be described in terms of the above decomposition. The most frequently used scenario is the *parallel shift,* which involves an increase or decrease in the entire term structure by a constant number of basis points. A further scenario, called the *twist*, involves a change in the slope of the term structure. This scenario is commonly realized through the addition (or subtraction) of, for example, m basis points to the interest rate corresponding to a term of m years; this is done for all terms in the term structure. Yet another scenario found in risk management is called *hump.* This scenario describes an increase in the short and long term rates and decrease in those for terms of intermediate length or vice versa. These three scenarios, the parallel shift, the twist and the hump, very often deduced on the basis of subjective experience, in fact correspond exactly to the first three principle components of the term structure. The scenarios mentioned here thus represent, from the statistical point of view, the most significant movements in the term structure. From the construction of the principle components, we can assume that these movements are approximately independent of one another and thus a simple representation of the dynamics of the term structure exists.

Finally, the analysis leads to the following results: first, practically all typical fluctuations in the term structure can be described by a combination of the above three scenarios. Second, making use of the eigenvalues, i.e. the variances of the principle components, confidence levels can be determined for the associated random variables. For example, information on the probability of the actual occurrence of one of these scenarios could be computed. We could proceed one step further: since the covariance matrix for the principle components is diagonal, the value at risk can be differentiated with respect to movements in the individual principle components. The threat of VaR losses can be traced back to different types of movements in the term structure and interpreted accordingly. In addition, the values at risk from the three named stress scenarios can be taken to be uncorrelated and the value at risk can be computed by simply taking the sum of their squares.

33

Pre-Treatment of Time Series and Assessment of Models

33.1 Pre-Treatment of Time Series

The *pre-treatment* for the transformation of a given data set into a *stationary* time series has been mentioned several times in the preceding sections and will receive detailed treatment in this section. The basis for pre-treating a time series is its decomposition into a *trend* component g_t, a *seasonal* component s_t, and a random component[1] Z_t:

$$X_t = g_t + s_t + Z_t \tag{33.1}$$

Z_t then represents the stationary time series with $\mathrm{E}[Z_t] = 0$. The *trend* g_t is a deterministic function of the time variable t, which represents a long-term development, for example a polynomial or an exponential function[2]. A weaker trend can sometimes be more readily recognized after a compression of the time axis. The *season* s_t represents a periodic component with a period p:

$$s_t = s_{t+p} \tag{33.2}$$

It follows that the sum $\sum_{i=1}^{p} s_{t+i}$ of p successive values is a constant. This constant can be incorporated into the trend g_t so that, without loss of generality, the sum can be assumed to be equal to zero:

$$\sum_{i=1}^{p} s_{t+i} = 0$$

33.1.1 Differencing

If the trend and the seasonal component are not of interest, they can be eliminated by taking differences of certain values in the time series. Equation 33.2 implies that a

[1] Notation: capital letters here denote random variables, small letters, deterministic functions.

[2] Or, for example, a so-called *logistic curve* $c_0/(1+c_1 \exp(-c_2 t))$. Such curves address the possibility of saturation.

seasonal component with period p can be eliminated by taking the p^{th} difference Δ_p:

$$\begin{aligned}\widetilde{X}_t &:= \Delta_p X_t := X_t - X_{t-p} \\ &= g_t - g_{t-p} + Z_t - Z_{t-p} \\ &=: \widetilde{g}_t + \widetilde{Z}_t\end{aligned}$$

with the new trend $\widetilde{g}_t = g_t - g_{t-p}$ and a new stationary time series $\widetilde{Z}_t = Z_t - Z_{t-p}$. Under the assumption that the remaining trend can be represented by a polynomial of degree k, this trend can be eliminated by taking first order differences in the above function k times. If, for example, the trend is a second degree polynomial, i.e. if

$$\widetilde{g}_t = a + bt + ct^2$$

then differencing twice eliminates this trend: taking a first order difference once yields

$$\Delta\widetilde{X}_t := \widetilde{X}_t - \widetilde{X}_{t-1} = b - c + 2ct + \widetilde{Z}_t - \widetilde{Z}_{t-1}$$

Differencing a second time results in

$$\Delta^2\widetilde{X}_t = \Delta\widetilde{X}_t - \Delta\widetilde{X}_{t-1} = \widetilde{X}_t - 2\widetilde{X}_{t-1} + \widetilde{X}_{t-2} = 2c + \Delta^2\widetilde{Z}_t$$

This time series now has neither a trend nor a seasonal component. After a subsequent elimination of the expectation $2c$, the new time series $\widetilde{\widetilde{Z}}_t := \Delta^2\widetilde{X}_t - 2c$ is stationary and a time series model can be fitted.

In this way, the seasonal and trend components can be eliminated. The original data can, of course, be reconstructed through the inverse transformation. If we are interested in estimating the trend or the seasonal component themselves, we should begin by estimating the trend. This occurs either through one of the above mentioned parametric functions, (the parameters are then determined using a least squares estimator) or through moving averages as described in the next subsection. This estimated trend $\widehat{g}_t$ is then subtracted from the time series and the resulting series $X_t - \widehat{g}_t$ is used to obtain an estimator $\widehat{s}_t$ for the seasonal component. Because of the periodicity, the sum $\sum_{i=1}^{p} \widehat{s}_{t+i}$ is constant in time. This constant is then ascribed to the expectation of the time series so that $\sum_{i=1}^{p} \widehat{s}_{t+i} = 0$ holds. The difference $Z_t = X_t - \widehat{g}_t - \widehat{s}_t$ can then be viewed as a stationary time series and a time series model can be fitted.

33.1.2 Filters

One possibility of estimating the trend and eliminating the season is the application of *moving averages*. Starting from the time series $\{X_t\}$ a new process Y_t is constructed as

$$Y_t = \sum_{i=-m_1}^{m_2} q_i X_{t+i}$$

Frequently $m_1 = m_2$ (*symmetric average*) is chosen. Moving averages are an example of *filters*. Filters admit some information from the old time series into the new one

while removing other information; the data sequence $\{X_t\}$ is filtered. Moving averages are an example of *linear* filters.

The weights q_i employed in the averaging are selected so that the trend is conserved, the seasonal component eliminated and the variance of the rest of the data sequence is minimized. An example taken from empirical economic research is the moving average

$$Y_t = \frac{1}{8}X_{t-2} + \frac{1}{4}X_{t-1} + \frac{1}{4}X_t + \frac{1}{4}X_{t+1} + \frac{1}{8}X_{t+2}$$

which allows a linear trend to "flow through", filters out a period-four seasonal component, (quarterly data) and reduces the variance; this can be seen in the following computation: let X_t be of the form in Equation 33.1 which includes a linear trend with slope a (i.e. $g_{t+i} = g_t + a\,i$) and a seasonal component of period four, i.e. $\sum_{i=1}^{4} s_{t+i} = 0$. Then

$$Y_t = \sum_{i=-2}^{2} q_i X_{t+i} = \sum_{i=-2}^{2} q_i g_{t+i} + \sum_{i=-2}^{2} q_i s_{t+i} + \sum_{i=-2}^{2} q_i Z_{t+i}$$

The trend of Y_t remains the same as that of X_t:

$$\begin{aligned}\sum_{i=-2}^{2} q_i g_{t+i} &= \sum_{i=-2}^{2} q_i (g_t + ai) \\ &= \frac{1}{8}(g_t - 2i) + \frac{1}{4}(g_t - i) + \frac{1}{4}g_t + \frac{1}{4}(g_t + i) + \frac{1}{8}(g_t + 2i) \\ &= g_t - \frac{i}{4} - \frac{i}{4} + \frac{i}{4} + \frac{i}{4} = g_t\end{aligned}$$

The seasonal component is eliminated:

$$\begin{aligned}\sum_{i=-2}^{2} q_i s_{t+i} &= \frac{1}{8}s_{t-2} + \frac{1}{4}s_{t-1} + \frac{1}{4}s_t + \frac{1}{4}s_{t+1} + \frac{1}{8}\underbrace{s_{t+2}}_{=s_{t-2}} \quad \text{since} \quad s_{t+4} = s_t \\ &= \frac{1}{4}\underbrace{(s_{t-2} + s_{t-1} + s_t + s_{t+1})}_{=0} \quad \text{since} \quad \sum_{i=1}^{p} s_{t+i} = 0 \\ &= 0\end{aligned}$$

And the variance is (under the assumption that the stationary time series Z_t has weak autocorrelations or none at all) substantially reduced:

$$\begin{aligned}\mathrm{var}[\sum_{i=-2}^{2} q_i Z_{t+i}] &= \sum_{i,j=-2}^{2} q_i q_j \mathrm{cov}\,[Z_{t+i}, Z_{t+j}] \\ &= (\frac{1}{64} + 3\frac{1}{16} + \frac{1}{64})\mathrm{var}[Z_t] + \sum_{i \neq j} q_i q_j \underbrace{\mathrm{cov}\,[Z_{t+i}, Z_{t+j}]}_{\approx 0} \\ &\approx \frac{7}{32}\mathrm{var}[Z_t]\end{aligned}$$

A further example of a filter is the Spencer 15–point moving average with coefficients

$$\frac{1}{320}[-3,-6,-5,3,21,46,67,74,67,46,21,3,-5,-6,-3]$$

This filter even allows all polynomials up to and including polynomials of order three to flow through.

A complete filter can only be accomplished "within" the time series, namely for those data for which $t = m_1 + 1, \ldots, T - m_2$, where the length of the data sequence, i.e. the number of data points, is denoted by T. One possibility of applying the boundary points of the data sequence is to simply extend the moving average past the end of the known values, replacing the unknown values with the last value X_T (at the end of the sequence) and with the first value X_1 (at the beginning of the sequence). Alternatively, we could use smaller supports for the filters at the ends than in the middle.

33.1.3 Scaling

If the *variance* of the data sequence shows an increasing or decreasing behavior as a function of time, a variance stabilizing transformation, such as the *Box-Cox scaling* defined by

$$\begin{aligned} T_\lambda(X_t) &= \frac{X_t^\lambda - 1}{\lambda}, \quad \text{for} \quad X_t \geq 0, \lambda > 0 \\ T_0(X_t) &= \ln(X_t), \qquad \text{for} \quad X_t > 0, \lambda = 0 \end{aligned}$$

can be implemented. The second expression applies to the limiting case $\lambda \to 0$. If the data sequence contains negative values, we may add a positive constant to each of the data points. The parameter values $\lambda = 0$ or $\lambda = 1/2$ are most commonly used. Intuitively, the transformation implies that, depending on the choice of λ, small values can be magnified (or reduced) or vice versa.

As an example, we consider the transformation $T(X) = \ln(X)$. This transformation yields an approximately constant variance, when the standard deviation of the time series changes linearly with the expectation, i.e. when $\mathrm{std}(X_t) = a\mu_t$. This can be shown as follows: the Taylor series expansion up to first order of the transformed time series about the expectation μ_t yields for the standard deviation

$$\begin{aligned} \mathrm{std}(T(X_t)) &= \sqrt{\mathrm{var}(T(X_t))} \approx \sqrt{\mathrm{var}(T(\mu_t) + T'(\mu_t)(X_t - \mu_t))} \\ &= T'(\mu_t)\,\mathrm{std}(X_t) = T'(\mu_t)\,a\mu_t. \end{aligned}$$

This expression is in fact constant when the transformation $T(X) = \ln(X)$ is chosen since $T'(X) = X^{-1}$.

33.2 Measuring the Goodness of Time Series Models

Having found a model for a stationary time series and fitted its parameters to the time series data, we immediately ask "how well does the model describe the time series?",

in other words, how good is the fit? This question is answered by considering the deviations (also called the *residues)* of the values in the time series actually observed in comparison with those produced by the model. If the model is good, all of the information contained in the time series should also be contained in the fitted model; the residues should be of a purely random nature. For an AR(p) process as in Equation 30.6, the residues are

$$\widehat{\varepsilon}_t = X_t - \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i} \tag{33.3}$$

If the fit is good, the residues should be normally distributed *iid* random variables with variance σ^2.

For a GARCH(p, q) process of the form 30.13, the situation is somewhat different since the random variables are *multiplicative* and not additive. We can, however, define the residues as the *quotient* of the observed data and the values obtained from the fitted time series model.

$$\widehat{\varepsilon}_t = X_t / \sqrt{\widehat{H}_t} \tag{33.4}$$

Here, if the fit is good, the residues should be *iid* standard normal random variables.

In order to test whether the residues form a sequence of *iid* random variables $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$, all statistical procedures for testing whether a sample of observations are the realizations of normally distributed *iid* random variables may be applied. Some of these tests will be introduced below. These tests are applied to the GARCH (1,1) process in the Excel workbook GARCH.XLS.

33.2.1 Hypothesis Tests

Sign Test

We begin with a very simple test. For a sample of size T of *iid* random variables, the number V of points for which $\hat{\varepsilon}_t > \hat{\varepsilon}_{t-1}$ holds, has the following expectation and variance

$$\mathrm{E}[V] = \frac{1}{2}(T-1) \quad \text{and} \quad \mathrm{var}[V] = \frac{1}{12}(T+1)$$

For large values of T, this number V is approximately $\mathrm{N}(\mathrm{E}[V], \mathrm{var}[V])$ distributed. The quotient $\frac{V-\mathrm{E}[V]}{\sqrt{\mathrm{var}[V]}}$ is thus N(0,1) distributed. The hypothesis of a normal distribution is *rejected* with a confidence of $1-\alpha$ (with α typically 5% or 10%) if

$$\frac{|V-\mathrm{E}[V]|}{\sqrt{\mathrm{var}[V]}} = \frac{|V-(T-1)/2|}{\sqrt{(T+1)/12}} > \mathrm{N}^{-1}(1-\alpha/2)$$

holds, where here $\mathrm{N}^{-1}(1-\alpha/2)$ denotes the percentile for the probability $1-\alpha/2$ of the N(0,1) distribution[3]. Conversely, from the measurements of V, the confidence with

[3]The percentile associated with $1-\alpha/2$ is sought since the test under consideration is two-sided (we consider the absolute value of the normalized random variable). In total, however, the confidence level of the test is $1-\alpha$ since $\mathrm{P}\left(\frac{|V-\mathrm{E}[V]|}{\sqrt{\mathrm{Var}[V]}} \leq \mathrm{N}^{-1}(1-\alpha/2)\right) = 1-\alpha$

which this test will reject the hypothesis can be computed. From the above equation, we obtain for α

$$\mathrm{N}\left(\frac{|V-(T-1)/2|}{\sqrt{(T+1)/12}}\right) = 1-\alpha/2$$

where N stands for the cumulative standard normal distribution. The confidence with which the hypothesis will be rejected thus becomes

$$\text{Confidence} = 1-\alpha = 2\mathrm{N}\left(\frac{|V-(T-1)/2|}{\sqrt{(T+1)/12}}\right) - 1$$

However, this test does not reject the hypothesis of an independent and identically distributed random sample in the case of cyclical data because approximately half of the residues are then increasing. In practice, this does not present a problem since the periodicity can be easily spotted by considering a plot of the residues.

Inflection Points Test

This test has a similar structure as in the sign test. For *iid* random variables, the number of inflection points U for which $\hat{\varepsilon}_{t-1} < \hat{\varepsilon}_t > \hat{\varepsilon}_{t+1}$ or else $\hat{\varepsilon}_{t-1} > \hat{\varepsilon}_t < \hat{\varepsilon}_{t+1}$ holds, must have the following expectation and variance

$$\mathrm{E}[U] = \frac{2}{3}(T-2) \quad \text{and} \quad \mathrm{var}[U] = \frac{16T-29}{90}$$

For large T, this number U is approximately $\mathrm{N}(\mathrm{E}[U], \mathrm{var}[U])$ distributed. Thus, analogous to the sign test, the hypothesis of *iid* random variables will be *rejected* with a confidence $1-\alpha$ if

$$\frac{|U-\mathrm{E}[U]|}{\sqrt{\mathrm{var}[U]}} = \frac{|U-2(T-2)/3|}{\sqrt{\mathrm{var}[(16T-29)/90]}} > \mathrm{N}^{-1}(1-\alpha/2)$$

Analogous to the sign test, the confidence with which the hypothesis will be rejected can be calculated from U as

$$\text{Confidence} = 1-\alpha = 2\mathrm{N}\left(\frac{|U-2(T-2)/3|}{\sqrt{\mathrm{var}[(16T-29)/90]}}\right) - 1$$

Percentile Test and Kuiper Statistic

We start by selecting a two-sided confidence level α. Two-sided means that $\alpha/2$ (percent) of the residues may fall below the lower boundary of the confidence interval and $\alpha/2$ above the upper boundary. The confidence interval itself is determined from the conditional variance σ^2 of the residues $\hat{\varepsilon}_t$ (from Equation 33.4 $\sigma^2 = 1$ for GARCH processes, for example) under the assumption that residues are normally distributed, i.e. $\hat{\varepsilon}_t/\sigma$ are assumed to have a standard normal distribution. We now count the residues $\hat{\varepsilon}_t$ lying outside this confidence interval. For these residues we either have $\mathrm{N}(\hat{\varepsilon}_t/\sigma) > 1-\alpha/2$ or $\mathrm{N}(\hat{\varepsilon}_t/\sigma) < \alpha/2$. If the goodness of the fit is high, then the

residues should indeed be $N(0, \sigma^2)$ distributed and therefore the proportion (relative to the total number of residues) of the residues lying outside the confidence interval should be approximately equal to the confidence α. This should hold for every arbitrary α between 0 and 1. We therefore consider the quotients

$$F(\alpha) = \frac{\text{Number of Residues Where } \mathrm{N}(\hat{\varepsilon_t}/\sigma) > 1 - \alpha/2 \text{ or } \mathrm{N}(\hat{\varepsilon_t}/\sigma) < \alpha/2}{\text{Total Number of all Residues}} \text{ for } 0 \leq \alpha \leq 1 \tag{33.5}$$

For a good fit, this should be close to α, i.e. the variable $F(\alpha) - \alpha$ should be close to zero for all $0 \leq \alpha \leq 1$. The deviations $F(\alpha) - \alpha$ should be uniformly distributed over the entire spectrum of α (this implies that the assumption of a normal distribution is correct) and they should be independent (this implies in particular no autocorrelations). In risk management, the performance at the "tails" of the distribution, i.e. for $0 \leq \alpha \leq 0.05$ and $0.95 \leq \alpha \leq 1.0$, is of particular importance.

The *Kuiper statistic* defined as

$$K(F(\alpha), \alpha) = \max_{0 \leq \alpha \leq 1} \{F(\alpha) - \alpha\} + \max_{0 \leq \alpha \leq 1} \{\alpha - F(\alpha)\} \tag{33.6}$$

adds the largest deviation of $F(\alpha)$ from a from above and from below. The Kuiper statistic selects the "worst" deviations over the entire range of α and aggregates them into a single number which can then be compared with the corresponding number of other models for the purpose of evaluating the goodness of each of the models.

Estimation of the Autocorrelation Function

The estimators $\hat{\varrho}(h)$ for the autocorrelations defined in Equation 29.6 are approximately normally distributed for an independent identically distributed sequence of T random variables with finite variance. For large values of T, this distribution is approximately $\mathrm{N}(0, 1/T)$. Thus, for approximately $1 - \alpha$ percent of the autocorrelations between the residues,

$$-\frac{\mathrm{N}^{-1}(1 - \alpha/2)}{\sqrt{T}} < \hat{\varrho}(h) < \frac{\mathrm{N}^{-1}(1 - \alpha/2)}{\sqrt{T}}$$

should hold if the fit is good. Here again, N denotes the cumulative standard normal distribution. More "outliers" would indicate that the assumption of independent, identically distributed random variables is violated by the residues.

QQ Plot

In a *QQ plot*, also referred to as a *percentile plot* or a *quantile plot*, the percentiles of two distributions are plotted against one another. We will denote the two cumulative distribution in question by Φ_1 and Φ_2. For the QQ plot, we consider the question: given the probability $P = \Phi_1(X)$, which argument Y must be substituted into the distribution Φ_2 in order to generate the same probability P. In other words: which Y has to be chosen so that $\Phi_2(Y) = \Phi_1(X)$. Both of the arguments are then the percentiles of the two distributions Φ_1 and Φ_2 for the same probability value P. These

two percentiles are plotted against one another. We plot Y against X explicitly by defining the function $Y = f(X)$ as

$$f(X) = \Phi_2^{-1}\left(\Phi_1(X)\right)$$

If, for example, Φ_1 and Φ_2 are two *normal* distributions, the QQ plot is also referred to as a *normal plot*. For two normal distribution, i.e. for $\Phi_1 = \mathrm{N}(\mu_1, \sigma_1^2)$ and $\Phi_2 = \mathrm{N}(\mu_2, \sigma_2^2)$, the condition for the percentiles X and Y is

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{X} e^{-\frac{(u-\mu_1)^2}{2\sigma_1^2}}\, du = \Phi_1(X) = \Phi_2(Y) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \int_{-\infty}^{Y} e^{-\frac{(u-\mu_2)^2}{2\sigma_2^2}}\, du$$

With the substitution $w = \frac{u-\mu_i}{\sigma_i} \Longrightarrow du = \sigma_i dw$ on both sides, i.e. for $i = 1, 2$, this equation becomes

$$\int_{-\infty}^{\frac{X-\mu_1}{\sigma_1}} e^{-w^2/2}\, dw = \int_{-\infty}^{\frac{Y-\mu_2}{\sigma_2}} e^{-w^2/2}\, dw$$

These two integrals differ only in their upper limit. If they are to be equal, it suffices for their upper limits to be equal, yielding the function for the percentile plot

$$\begin{aligned} &\frac{X-\mu_1}{\sigma_1} = \frac{Y-\mu_2}{\sigma_2} \Rightarrow \\ &f(X) = Y = \frac{\sigma_2}{\sigma_1}\left(X-\mu_1\right) + \mu_2 \end{aligned} \tag{33.7}$$

Thus, if both random variables have a normal distribution, the QQ plot is simply a straight line whose slope is given by the ratio of the standard deviations, $\frac{\sigma_2}{\sigma_1}$, and whose offset, $\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1$, is determined by the expectation of the distributions as well as the ratio of the standard deviations.

Usually, one of the distributions, Φ_2 say, is obtained from empirical sample values (*empirical probability distribution*) and it is tested whether this distribution agrees with another, theoretical distribution Φ_1. To do this, the T sample values $\hat{\varepsilon}_i$ are arranged in increasing order resulting in the sequence $\hat{\varepsilon}_{(1)}, \hat{\varepsilon}_{(2)}, \ldots, \hat{\varepsilon}_{(T)}$ with $\hat{\varepsilon}_{(1)} < \hat{\varepsilon}_{(2)} < \ldots < \hat{\varepsilon}_{(T)}$. To distinguish the ordered from the original sample sequence, the indices of the *ordered* sequence appear in parentheses. The empirical distribution is obtained from the following deliberations: the proportion of all sample values which are less than or equal to $\hat{\varepsilon}_{(i)}$ is i/T. For large T, this ratio is a good approximation of the cumulative empirical distribution, i.e. i/T approximates the *empirical probability* that a random number is less than or equal to $\hat{\varepsilon}_{(i)}$

$$\Phi_2(\hat{\varepsilon}_{(i)}) = P_i \approx i/T$$

This implies that the percentile of the empirical distribution associated with the probability $P_i \approx i/T$ is

$$Y_i = \Phi_2^{-1}(P_i) = \hat{\varepsilon}_{(i)}$$

The associated percentile of the *theoretical* distribution Φ_1 is then[4]

$$X_i = \Phi_1^{-1}(P_i) \approx \Phi_1^{-1}(i/T) \quad \text{for all } i < T \tag{33.8}$$

The QQ plot is now obtained by plotting the ordered pairs (X_i, Y_i), i.e. by graphing $\hat{\varepsilon}_{(i)}$ against $\Phi_1^{-1}(i/T)$ for all $i < T$. In this way it can be checked, for example, whether the sample values are normally distributed by testing how well a QQ plot of the sample values against those obtained from a normal distribution delivers a straight line[5]. It is convenient to chose the *standard* normal distribution, $\Phi_1 = \mathrm{N}(0,1)$, as the theoretical distribution in this case. Equation 33.7 then reduces to $Y_i = \mu_2 + \sigma_2 X_i$ for $i = 1, \ldots, T-1$ and the standard deviation and the expectation of the empirical distribution can be read off directly as the slope and offset of the QQ plot.

A normal plot such as this reveals much more than a common two-sample test since the entire distribution can be tested and deviations from the normal distribution such as non-linearities, skewness, or dissimilar distribution boundaries can be identified. In addition to this visual examination, the *correlation coefficient* between $Y_i = \hat{\varepsilon}_{(i)}$ and $X_i = \Phi_1^{-1}(P_i)$

$$R^2 = \frac{\left(\sum_{i=1}^{T}(Y_i - \overline{Y})\, X_i\right)^2}{\sum_{i=1}^{T}(Y_i - \overline{Y})^2 \sum_{i=1}^{T} X_i^2} \quad \text{with} \quad \overline{Y} = \sum_{i=1}^{T} Y_i/T$$

serves as an analytic variable for testing the linearity[6]. For a straight line, $R^2 = 1$ must hold[7].

33.2.2 Goodness of Fit vs. Goodness of Forecast

All of the above criteria can be applied to not only test the goodness of a *fit* but also the goodness of a *forecast*. The forecast must occur *out of sample*, i.e. a portion of the given data set is used to obtain an estimator (i.e. for fitting the model), while the remaining data is predicted using the fitted model. The goodness is then evaluated on

[4]Note that as long as we have only finitely many observations, i.e. for $T < \infty$, the *empirical* probability Φ_2 for any $\hat{\varepsilon}_i$ being $\leq \hat{\varepsilon}_{(T)}$ is of course 1 since $\hat{\varepsilon}_{(T)}$ is *by definition* the largest value ever observed. The *theoretical* cumulative probability $\Phi_1(x)$, on the other hand, approaches 1 only for x approaching the upper limit of the range of the possible arguments of Φ_1, usually for $x \to \infty$. Or in the language of Equation 33.8:

$$\text{"}X_T \approx \Phi_1^{-1}(T/T) = \Phi_1^{-1}(1) = \infty\text{"}$$

Therefore we have to exclude the largest value $\hat{\varepsilon}_{(T)}$ from the QQ plot when testing an empirical distribution.

[5]It can be shown that using $P_i = \frac{i-3/8}{T-3/4+1}$ for the probabilities in a normal plot is slightly better than using $P_i = i/T$. This, however, is only a marginal improvement over the much more intuitive value $P_i = i/T$.

[6]Here, the correlation coefficient for the special case that the X_i have mean zero, $\overline{X} = 0$, is presented, which is the case for a QQ plot against the standard normal distribution. In general, the correlation coefficient also has the form given in the equation, but with X_i replaced by $(X_i - \overline{X})$.

[7]For a formal test, we need confidence intervals under the hypothesis of normally distributed residues. For example, for $T = 200$, $P(R^2 < 0.987) = 0.05$ (5% interval) and $P(R^2 < 0.989) = 0.1$ (10% interval).

the basis of the "accuracy" of the prediction. This "accuracy" must be quantified for each respective problem. Testing the goodness of a forecast by using the entire sample to fit the model and then "predicting" values *in* the sample is obviously optimistically biased and does not yield a realistic assessment of the actual forecasting ability of the model. Time series models are often constructed for the primary purpose of making forecasts. Thus, the goodness of a model should be judged according to its performance in making predictions. In the Excel workbook GARCH.XLS, the GARCH(1,1) model was fitted to the first 400 time series values via maximum likelihood estimation using simulated annealing. The goodness tests presented in GARCH.XLS refer to the data points 401 to 1001 and are therefore *out of sample forecasting tests*.

However, there are by all means tests available which only make sense *within* that part of the sample which was used for fitting the model (*in sample* tests). In such cases, we are only interested in the goodness of fit achieved by the model, not in its forecasting power. If, for example, the best model is to be identified from among a whole class of different models it is of course natural to ask which model we prefer. The goodness of fit, measured in terms of the smallest possible residues or maximum possible likelihood, can not be the only deciding factor because a model with many parameters might be well fitted but can also be easily overfitted. Moreover, models with few parameters are generally preferable as they are easier to interpret.

The *Akaike information criterion* is a compromise between larger likelihood and fewer parameters. It is given by

$$\mathrm{AIC} = -2\mathcal{L} + \frac{2(Q+1)T}{T-Q-2} \tag{33.9}$$

where Q denotes the number of parameters of the model to be evaluated and $\mathcal{L}$ the log-likelihood of the fitted process. For the GARCH(p,q) process the order is $Q = p+q+1$, for an AR(p)-process we have $Q = p$. The model with the smallest AIC is preferable. The Akaike Criterion represents a compromise between a large likelihood and a small number of parameters; models with many parameters are penalized.

33.2.3 Examples: Goodness of GARCH Models

The comparison of a GARCH$(1,1)$ process with moving averages by means of measures of goodness is explicitly demonstrated in the Excel workbook GARCH.XLS on the CD-ROM accompanying this book. In addition, we will demonstrate these methods using the above mentioned FTSE data set presented in Figure 30.1. The first 2,000 values of the data set are used in the parameter estimation. Simulated annealing algorithms were employed to obtain the optimized parameter values $\alpha_0 = 0.000$, $\alpha_1 = 0.089$ and $\beta_1 = 0.849$. Figure 33.1 shows the conditional standard deviations $h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 X_{t-1}^2$ of the GARCH process, where the corresponding values for X_{t-1} were taken from the FTSE data set.

The Akaike criterion for the goodness-of-fit of the GARCH process is computed as -6519.87. The last 934 values serve to test the one-step forecast for the variance performed by employing Equation 31.23. Here, we use a procedure (which has not been yet been discussed in previous sections) to directly test the estimated conditional

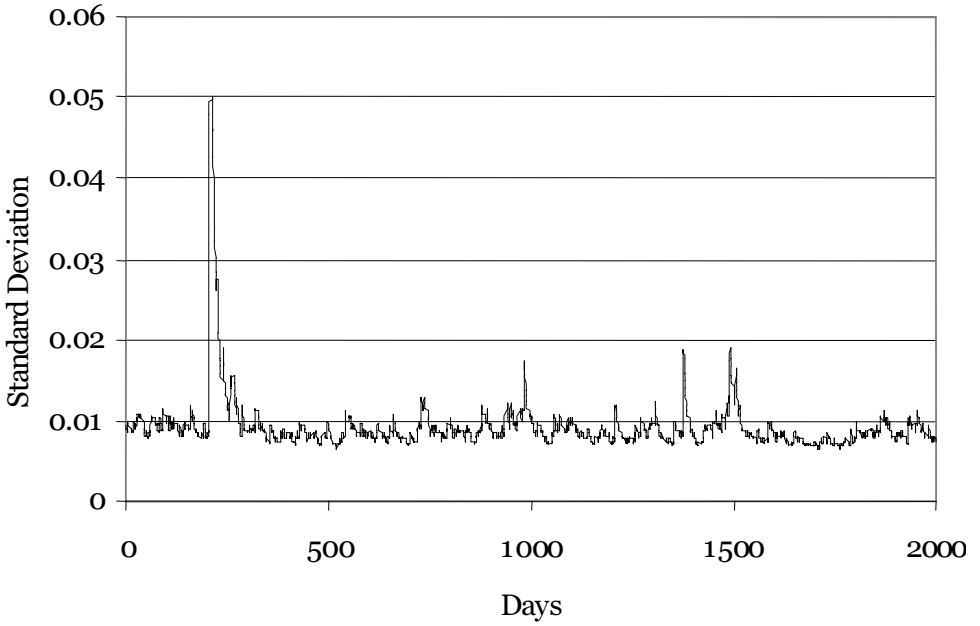


Figure 33.1: Conditional standard deviation of the GARCH$(1,1)$ process fitted to the FTSE time series. The crash in October 1987 is clear to see.

variance. We proceed as follows: beginning with a value X_{t-1} of the FTSE time series (for example with $t-1=2500$), we calculate $h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 X_{t-1}^2$ for the next day ($t = 2501$) with the parameters estimated using the first 2,000 values. This h_t is then used to compute the confidence intervals of the normal distribution $\mathrm{N}(0, h_t^2)$ corresponding to a confidence level of 99% and 95%, respectively, to establish whether the next value $X_t = X_{2501}$ of the FTSE time series lies within this confidence interval. Subsequently, the next h_{2502} is computed from h_{2501} and the value X_{2501} from the FTSE time series data. The corresponding confidence interval is then calculated and it is established whether the actual value X_{2502} lies within this confidence interval, and so on for all 934 out-of-sample values. The number of outliers is counted and compared with the expected number of outliers. For a test of the ten day forecast, Equation 31.25 is used to predict the conditional variance for the daily yield of the FTSE ten days into the future. For example, the conditional variance h_{2510} is estimated using the value X_{2500} from the FTSE time series and the associated value h_{2500}. Using h_{2510}, the confidence interval for the 99% and 95% confidence levels of the normal distribution $\mathrm{N}(0, h_{2510}^2)$ are computed and it is subsequently established whether the FTSE time series value X_{2510} lies within this confidence interval, and so on for all out-of-sample values. Again, the number of outliers is counted and compared with the expected number of outliers[8]. The results of these tests are presented in Tables 33.1 and 33.2.

[8]The forecast of the variance of the *total* return (for example, the 10-day total return) can be

1-Day Forecast	10-Day Forecast
GARCH: 8 (9)	GARCH: 77 (67)
GLD: 21 (9)	GLD: 145 (67)
EGLD: 18 (9)	EGLD: 117 (67)

Table 33.1: Confidence level $\alpha = 99\%$.

1-Day Forecast	10-Day Forecast
GARCH: 21 (46)	GARCH: 158 (337)
GLD: 69 (46)	GLD: 517 (337)
EGLD: 45 (46)	EGLD: 391 (337)

Table 33.2: Confidence level $\alpha = 95\%$.

The same tests were also performed for moving averages. The time window consists of 250 days (one year). The number of time series values lying outside of the confidence interval (outliers) is given for the uniformly weighted moving average (MA) and for the exponentially weighted moving average (EWMA). The weighting parameter for the EWMA is $\lambda = 0.94$. The values in parentheses indicate the theoretically expected number of outliers. For the 1-day predictions, the GARCH forecast is always superior. For the 10-day forecast, we observe mixed results: at 5% confidence, the GARCH process underestimates the theoretical number of outliers by almost one half, at 1% confidence, the number of outliers is very close to the theoretical value. The converse holds for the EWMA. The MA shows the worst performance. Fitting a GARCH(1,2) process leads to almost identical results.

Of all the investigated time series, the GARCH process is superior to all others in making a 1-day forecast. For the 10-day forecasts, the evaluation is not as uniform. For a similar fit to the S&P500 index over the same time span as for the FTSE, the 10 day prediction is better at 5% confidence than at 1%; the opposite to what we observed for the FTSE. For longer prediction periods, it is therefore important to conduct a separate analysis for each time series and repeat this regularly at certain times in order to asses the goodness of the methods and identify changes as early as possible.

Figure 33.2 shows the number of outliers for the EWMA as a function of λ. In this way, it can be established which λ is the best, based on the goodness criterion of interest.

A further variant of analysis is the Kuiper statistic K. This statistic will be illustrated using data taken from the S&P500 index over the same time window as for the FTSE index in the above example (see Figures 33.3 and 33.4). Again, a GARCH(1,1) process was fitted to the first 2,000 time series values of the S&P500 returns, the last

tested analogously. The estimator $\mathrm{Var}[\sum_{i=1}^{10} X_{2500+i} | X_{2500}, \ldots, X_1]$ is computed according to 31.26. The result is used to determine the confidence interval and then it is checked whether $\sum_{i=1}^{10} X_{2500+i}$ lies within this confidence interval. The X_{2500+i} are taken from the FTSE time series. Again, the number of outliers is compared with the expected number of outliers.

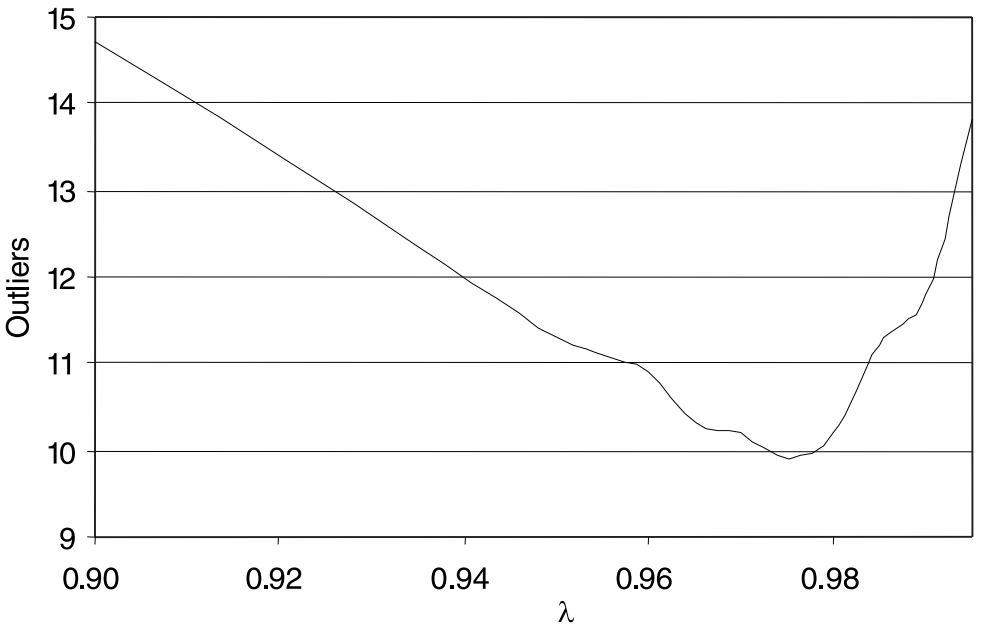


Figure 33.2: Number of outliers outside a 99% confidence interval as a function of λ for a 10 day EWMA volatility forecast. The theoretical value is 67 outliers.

934 were reserved for out-of-sample testing of the forecasting.

For MA and EWMA, the time window is again 250 days. The Kuiper statistic K is least favorable for the GARCH process. However, in the areas so important in risk management, namely to the far left and far right of the distribution corresponding to small and large percentiles, respectively, the GARCH process shows the best performance. The best results for the 1-day forecasts are those from the GARCH model. The deviation in the middle of the distribution indicates that the assumed conditional normal distribution does not properly reflect the skewness in the actual distribution of the returns. For $\lambda = 0.8$, the value of K for the EWMA is the smallest but we observe a "reverse" in the performance of the EWMA at the tails of the distribution: here its performance is poorest. For $\lambda = 0.94$, the value of K is higher but the fit at the tails is better, for example between 0 and 0.05, it is better than MA.

Finally, Figure 33.5 shows daily standard deviations for the S&P500 data set after the 400^{th} day in the out-of-sample region (i.e. the region not used for fitting the parameters of the GARCH model). The increasing smoothness in the sequence GARCH, EWMA and MA is clear to see. In addition, we can observe the "phantom patterns" for the EWMA and more pronouncedly for the MA after the peak shortly before the 600^{th} day.

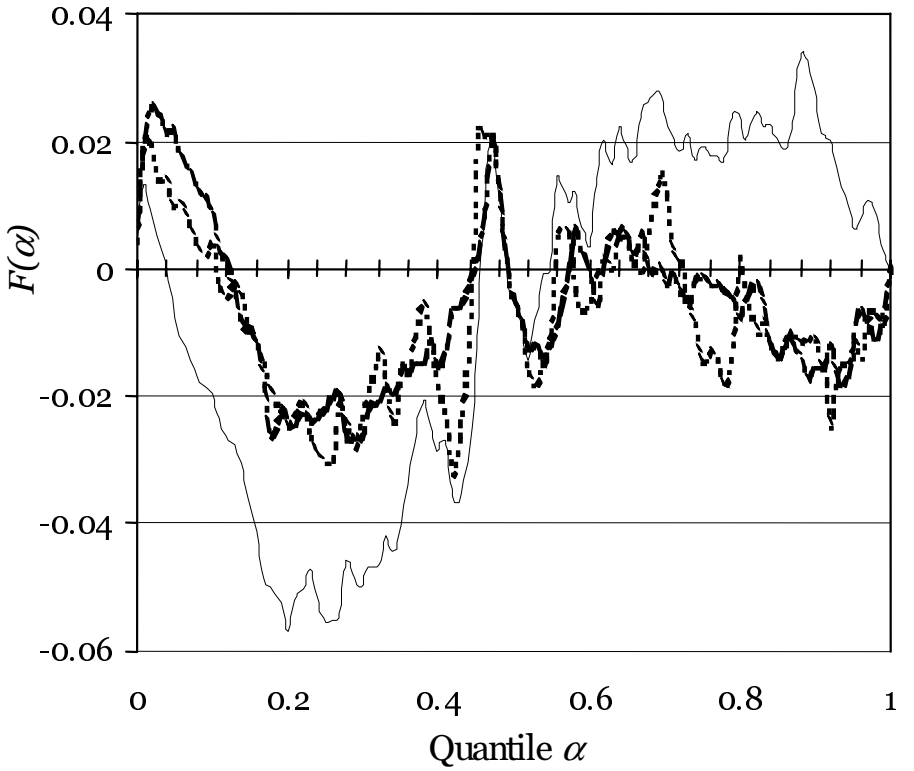


Figure 33.3: $F(\alpha)$ as defined in Equation 33.5 for GARCH (solid line), MA (dotted line) and EWMA (dash-dotted line). The Kuiper statistics, Equation 33.6, are K(GARCH)=0.093, K(MA)=0.058, K(EWMA)=0.055 for $\lambda = 0.8$.

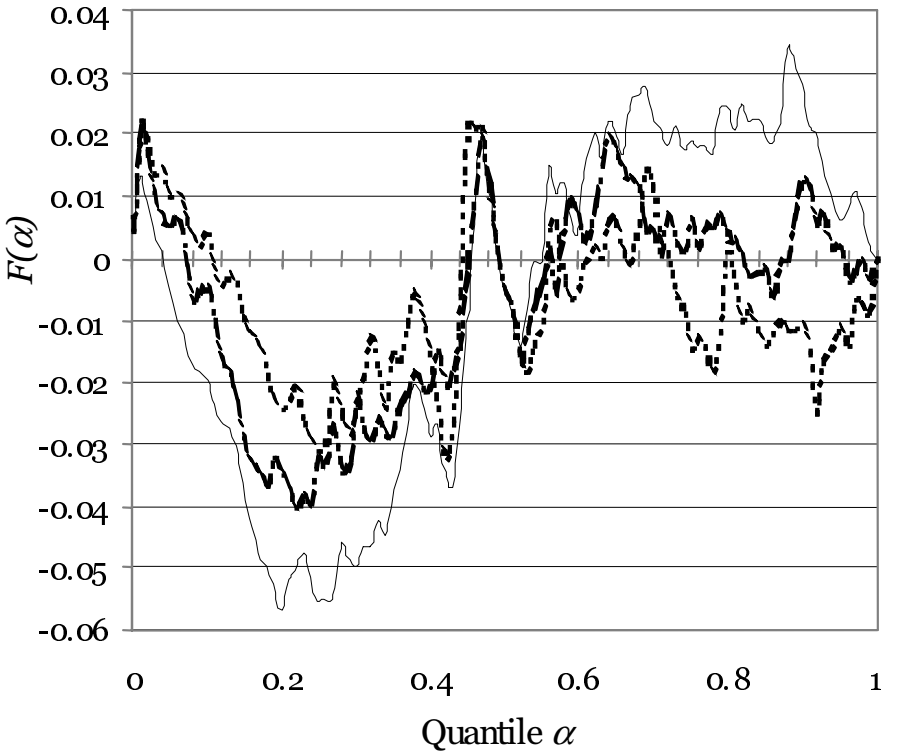


Figure 33.4: $F(\alpha)$ as in Figure 33.3. The only thing changed is that λ is now 0.94 for the EWMA. The Kuiper statistic for this new EWMA is K(EWMA)=0.064 for $\lambda = 0.94$.

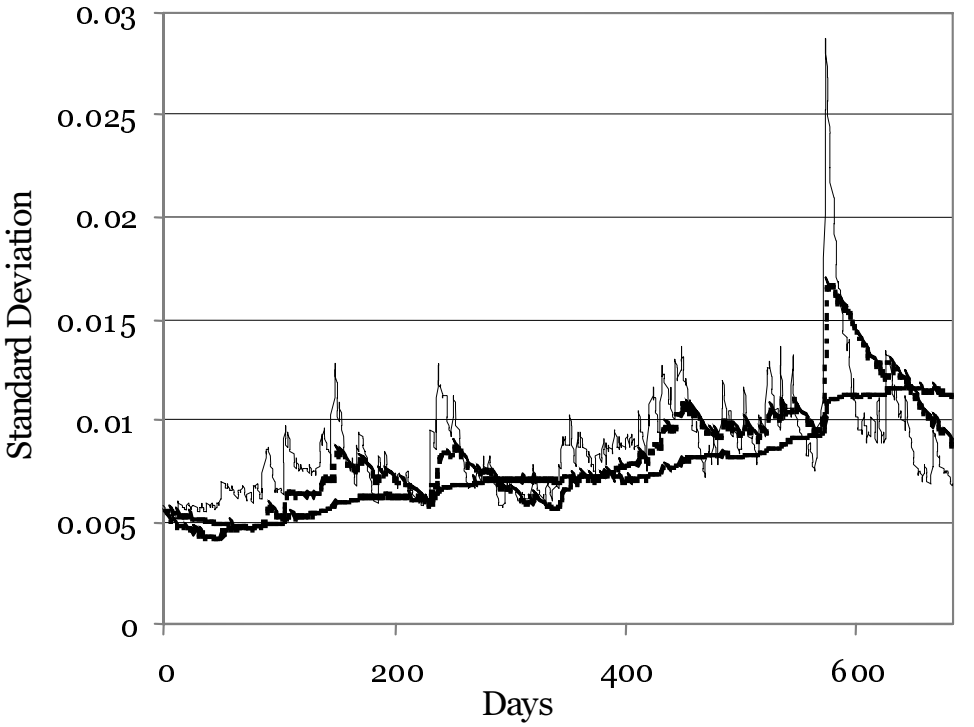


Figure 33.5: Daily standard deviations of the S&P500 returns from GARCH (solid line), MA (dotted line) and EWMA (dash-dotted line). The window for the moving averages spans 250 time series values. The EWMA damping factor is $\lambda = 0.975$.

Appendix A

Probability and Statistics

The basics of probability and statistics are presented in this appendix, providing the reader with a reference source for the main body of this book without interrupting the flow of argumentation there with unnecessary excursions. Only the basic definitions and results have been included in this appendix. More complicated concepts (such as stochastic processes, Ito Calculus, Girsanov Theorem, etc.) are discussed directly in the main body of the book as the need arises.

A.1 Probability, Expectation and Variance

A variable whose values is dependent on *random events*, is referred to as a *random variable* or a *random number*. An example of a random variable is the number observed when throwing a dice. The distribution of this random number is an example of a *discrete* distribution. A random number is discretely distributed if it can take on only discrete values such as whole numbers, for instance. A random variable has a *continuous* distribution if it can assume arbitrary values in some interval (which can by all means be the infinite interval from $-\infty$ to ∞). Intuitively, the values which can be assumed by the random variable lie together *densely*, i.e. arbitrarily close to one another.

An example of a discrete distribution on the interval from 0 to ∞ (infinity) might be a random variable which could take on the values 0.00, 0.01, 0.02, 0.03, ..., 99.98, 99.99, 100.00, 100.01, 100.02, ..., etc., like a Bund future with a tick-size of 0.01% corresponding to a value change of 10 euros per tick on a nominal of 100,000 euros. In contrast, a continuously distributed random variable can take on arbitrary values in the pertinent interval, for example, $\sqrt{5}$, π or $7/3$.

The *probability* $P(x < a)$ that a random variable x will be less than some arbitrary number a is the sum of the probabilities of all events in which x takes on values less than a. For continuous distributions, for which the possible values of x lie "infinitely dense" in some interval, the sum takes on the form of an integral:

$$P(x < a) = \sum_{i \text{ where } x_i < a} p(x_i) \longrightarrow \int_{-\infty}^{a} p(x)dx \tag{A.1}$$

The function p is called the *probability density* of the random variable x. It is often referred to as the *probability distribution*, *distribution density* or simply *the distribution*. In this text, the abbreviation pdf for *probability density function* will frequently be used in reference to the function p.

The function P is called the *cumulative probability*. It is often referred to as the *cumulative probability distribution* or simply *cumulative distribution* as well. We will frequently use the abbreviation cpf for *cumulative probability function* in reference to P.

It is certain that a random number will take on *some* value. The probability of something at all happening is thus equal to one. This property is called the *normalization to one* and holds for all probability distributions:

$$1 = \sum_i p(x_i) \longrightarrow \int_{-\infty}^{\infty} p(x)dx = 1 \tag{A.2}$$

The *expectation* of x is computed by taking the weighted sum of all possible values taken on by the random variable where the weights are the corresponding probabilities belonging to those values

$$\mathrm{E}[x] = \sum_i x_i p(x_i) \longrightarrow \int_{-\infty}^{\infty} xp(x)dx \tag{A.3}$$

A function f of a random variable x is again a random variable. The expectation of this function is calculated analogously by taking the weighted sum of the values of the function $f(x)$ evaluated at all possible values of x. The weights are again the probabilities belonging to the values of x:

$$\mathrm{E}[f(x)] = \sum_i f(x_i)p(x_i) \longrightarrow \int_{-\infty}^{\infty} f(x)p(x)dx \tag{A.4}$$

A special case is particularly interesting: setting $f(x)$ equal to the square of the deviation of x from its expected value, i.e. $f(x) = (x - \mathrm{E}[x])^2$, measures how strongly x fluctuates around its expected value. This measure is called the *variance*.

$$\mathrm{var}[x] = \mathrm{E}\left[(x - \mathrm{E}[x])^2\right] = \sum_i (x_i - \mathrm{E}[x])^2 p(x_i) \longrightarrow \int_{-\infty}^{\infty} (x - \mathrm{E}[x])^2 p(x)dx \tag{A.5}$$

The square root of the variance is called the *standard deviation*, abbreviated as std.

$$\mathrm{std}[x] := \sqrt{\mathrm{var}[x]} = \sqrt{\mathrm{E}\left[(x - \mathrm{E}[x])^2\right]} \tag{A.6}$$

For both discrete and continuous distributions, there exists a simple connection between the variance and the expectation: the variance is equal to the difference between

the *expectation of the square* of the random variable and the *square of the expectation* of the same random variable:

$$\operatorname{var}[x] = \mathrm{E}\left[x^2\right] - \mathrm{E}[x]^2 \tag{A.7}$$

The derivation is presented here for discrete distributions; the proof for the continuous case is completely analogous:

$$\begin{aligned}
\mathrm{E}\left[(x - \mathrm{E}[x])^2\right] &= \sum_i (x_i - \mathrm{E}[x])^2 \, p(x_i) \\
&= \underbrace{\sum_i x_i^2 p(x_i)}_{\mathrm{E}[x^2]} - 2\mathrm{E}[x] \underbrace{\sum_i x_i p(x_i)}_{\mathrm{E}[x]} + \mathrm{E}[x]^2 \underbrace{\sum_i p(x_i)}_{1} \\
&= \mathrm{E}\left[x^2\right] - 2\mathrm{E}[x]^2 + \mathrm{E}[x]^2 = \mathrm{E}\left[x^2\right] - \mathrm{E}[x]^2
\end{aligned}$$

A.2 Multivariate Distributions, Covariance, Correlation and Beta

Two random variables x and y which are not statistically independent (for example, the price of a Siemens share and the DAX) are said to be *correlated*. The probability $P(x < a, y < b)$ that a random variable x will be less than some value a and *simultaneously* that the second random variable y will be less than a value b equals the sum of the probabilities of all events in which $x < a$ and $y < b$. For continuous distributions, i.e. for infinitely dense values of the random variables, the sums converge to integrals:

$$P(x < a, y < b) = \sum_{x_i < a} \sum_{y_j < b} p(x_i, y_j) \longrightarrow \int_{-\infty}^{a} dx \int_{-\infty}^{b} dy \; p(x, y) \tag{A.8}$$

The function p is in this case the *joint probability density* for a pair of random variables x and y. Since such density functions refer to more than one (in this case two) random variables, they are referred to as *multivariate probability densities* or *multivariate probability distributions*, *multivariate distribution densities* or simply *multivariate distributions*.

Just as for a single random variable, the expectation of an arbitrary function $f(x, y)$ is calculated by taking the weighted sum of the values of the function $f(x, y)$ evaluated at all possible values of x and y. The weights are the joint probabilities belonging to the value pairs of (x, y):

$$\mathrm{E}[f(x, y)] = \sum_i \sum_j p(x_i, y_j) f(x_i, y_j) \longrightarrow \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \; p(x, y) f(x, y) \tag{A.9}$$

A special case is particularly interesting: setting the arbitrary function $f(x, y)$ equal to the product of the deviation of the random variables from their respective expectations,

i.e. $f(x,y) = (x - \mathrm{E}[x])(y - \mathrm{E}[y])$, yields a measure for the fluctuation of each random variable from its expectation as well as the degree of statistical dependence between the two random variables; or in other words, a "simultaneous" measure for the variance and the correlation. This expectation is called the *covariance* between x and y.

$$\begin{aligned} \mathrm{cov}\,[x,y] &= \mathrm{E}\,[(x - \mathrm{E}\,[x])\,(y - \mathrm{E}\,[y])] \\ &= \sum_{i,j} (x_i - \mathrm{E}\,[x])\,(y_j - \mathrm{E}\,[y])\,p(x_i, y_j) \\ &\longrightarrow \int\int dx\ dy\ p(x,y)\,(x - \mathrm{E}\,[x])\,(y - \mathrm{E}\,[y]) \end{aligned} \tag{A.10}$$

For both discrete and continuous distributions, there exists a natural extension (which is just as easily shown) of the relation in Equation A.7 between the variance and the expectations: the covariance is the difference between the expectation of the product and the product of the expectations of the random variables under consideration:

$$\mathrm{cov}\,[x,y] = \mathrm{E}\,[xy] - \mathrm{E}\,[x]\,\mathrm{E}\,[y] \tag{A.11}$$

The symmetry property of the covariance can be seen immediately from this equation; the covariance of x with y is the same as the covariance of y with x. The covariance of a random variable with itself is its variance.

$$\mathrm{cov}\,[x,y] = \mathrm{cov}\,[y,x]\ \ ,\ \ \mathrm{cov}[x,x] = \mathrm{var}\,[x] \tag{A.12}$$

The covariance has an additional useful property, it is bilinear: let a, b, c, d be constants and x, y, u, z random variables. Then

$$\begin{aligned} \mathrm{cov}[ax + by, cu + dz] = ac\mathrm{cov}[x,u] + ad\mathrm{cov}[x,z] \\ + bc\mathrm{cov}[y,u] + bd\mathrm{cov}[y,z] \end{aligned} \tag{A.13}$$

As often mention in the main body of this book, the covariances between n random variables can be represented in the form of an n by n matrix, called the *covariance matrix*. Because of the symmetry in Equation A.12, the information in the matrix entries appearing above the diagonal is the same as that below the diagonal; the entries in the diagonal itself are the variances of each of the n respective variables.

Dividing the covariance by the standard deviations of the two respective random variables yields the *correlation coefficient* ρ, also called the *correlation* between x and y. This value always lies between -1 and $+1$:

$$\rho(x,y) := \frac{\mathrm{cov}\,[x,y]}{\sqrt{\mathrm{var}[x]\mathrm{var}[y]}} = \frac{\mathrm{E}\,[xy] - \mathrm{E}\,[x]\,\mathrm{E}\,[y]}{\sqrt{(\mathrm{E}\,[x^2] - \mathrm{E}\,[x]^2)(\mathrm{E}\,[y^2] - \mathrm{E}\,[y]^2)}} \tag{A.14}$$

The symmetry property of the correlation can be seen immediately from this equation: The correlation of x with y is the same as the correlation of y with x. The correlation of a random variable with itself equals one.

$$\rho(x,y) = \rho(y,x)\ \ ,\quad \rho(x,x) = 1 \tag{A.15}$$

Two random numbers x_1 and x_2 are called *uncorrelated* if the correlation between the two is zero, i.e. if

$$\rho(x_i, x_j) = \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \quad , \quad i, j \in \{1, 2\} \tag{A.16}$$

where δ_{ij} denotes the well known *Kronecker delta*.

The *variance of a sum* of random variables is just the sum of the *covariances* of these random variables (in a similar ways as the expectation of a sum equals the sum of the expectations):

$$\begin{aligned} \mathrm{E}\left[\sum_i x_i\right] &= \sum_i \mathrm{E}\left[x\right] \\ \mathrm{var}\left[\sum_i x_i\right] &= \sum_{i,j} \mathrm{cov}[x_i, x_j] = \sum_{i,j} \rho\left(x_i, x_j\right) \sqrt{\mathrm{var}[x_i]\mathrm{var}[x_j]} \end{aligned} \tag{A.17}$$

The first of these equations, i.e. the linearity of the expectation, follows directly from the definition, Equation A.3. This linearity of the expectation is the only thing needed to show the second equation, i.e. the equation for the variance of a sum:

$$\begin{aligned} &\mathrm{var}\left[\sum_i x_i\right] \\ &= \mathrm{E}\left[\left(\sum_i x_i - \mathrm{E}\left[\sum_j x_j\right]\right)^2\right] = \mathrm{E}\left[\left(\sum_i x_i - \sum_j \mathrm{E}\left[x_j\right]\right)^2\right] \\ &= \mathrm{E}\left[\left(\sum_i \left(x_i - \mathrm{E}\left[x_i\right]\right)\right)^2\right] = \mathrm{E}\left[\sum_{i,j} \left(x_i - \mathrm{E}\left[x_i\right]\right)\left(x_j - \mathrm{E}\left[x_j\right]\right)\right] \\ &= \sum_{i,j} \mathrm{E}\left[\left(x_i - \mathrm{E}\left[x_i\right]\right)\left(x_j - \mathrm{E}\left[x_j\right]\right)\right] = \sum_{i,j} \mathrm{cov}[x_i, x_j] \end{aligned}$$

The first step is just the definition of the variance, Equation A.5. The linearity of the expectation is used in the second step for the inner expectation and in the fifth step for the outer expectation.

A direct application of Equation A.17 is the fact that the variance of a sum of *uncorrelated* random numbers (see Equation A.16) is just the sum of the individual variances:

$$\mathrm{var}\left[\sum_i x_i\right] = \sum_{i,j} \underbrace{\rho\left(x_i, x_j\right)}_{\delta_{ij}} \sqrt{\mathrm{var}[x_i]\mathrm{var}[x_j]} = \sum_i \mathrm{var}[x_i] \tag{A.18}$$

Thus, the standard deviation, Equation A.6, of such a sum of *uncorrelated* random numbers is obtained by adding the squares of each individual standard deviation and then taking the square root of this sum:

$$\mathrm{std}\left[\sum_i x_i\right] = \sqrt{\sum_i \mathrm{var}[x_i]} = \sqrt{\sum_i \mathrm{std}[x_i]^2} \tag{A.19}$$

Besides the above two symmetric quantities (covariance and correlation), a further *a*symmetric quantity is quite useful as well. This is the *beta* of y with respect to x defined as the covariance of x and y divided by the variance of x:

$$\beta(x,y) = \frac{\mathrm{cov}[x,y]}{\mathrm{var}[x]} = \sqrt{\frac{\mathrm{var}[y]}{\mathrm{var}[x]}}\rho(x,y) = \frac{\mathrm{E}[xy] - \mathrm{E}[x]\,\mathrm{E}[y]}{\mathrm{E}[x^2] - \mathrm{E}[x]^2} \tag{A.20}$$

A symmetry property can be seen immediately in this definition: the beta of a random variable with itself is indeed equal to one, the beta of x with respect to y is however not the same as the beta of y with respect to x. The conversion can be accomplished as follows:

$$\beta(y,x) = \frac{\mathrm{var}[x]}{\mathrm{var}[y]}\beta(x,y) \quad , \quad \beta(x,x) = 1 \tag{A.21}$$

Note that all of the above equations (in particular the very useful Equation A.17 and the properties A.18 and A.19 of uncorrelated variables) have been derived directly from first principles and are therefore valid no matter what probability distribution the random variables may have.

A.3 Moments and Characteristic Functions

The expectation and the variance are examples of the *moments* of a distribution. In general, the n-th moment of the distribution of the random variable x is defined as the expectation of the n^{th} power of the random variable:

$$\mathrm{E}[x^n] = \sum_i x_i^n p(x_i) \longrightarrow \int_{-\infty}^{\infty} x^n p(x)dx \tag{A.22}$$

The *central moments* μ_i of a distribution are defined as the "expectation of the powers of the difference between a random variable and its expectation":

$$\mu_j := \mathrm{E}[(x - \mathrm{E}[x])^j] \tag{A.23}$$

The first central moment ($j = 1$) is thus by definition equal to zero. The *expectation* of x is the first moment, the *variance* the *second central* moment of the distribution. With the third central moment $\mathrm{E}[(x - \mathrm{E}[x])^3]$ we can calculate the *skewness*, a measure for the asymmetry of the density. With the fourth central moment $\mathrm{E}[(x - \mathrm{E}[x])^4]$ we can calculate the *curtosis*, a measure for the weight of the distribution at the tail ends of its range. The exact definitions of the skewness and curtosis are

$$\begin{aligned} \text{Skewness} &:= \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\mathrm{E}[(x - \mathrm{E}[x])^3]}{\mathrm{E}[(x - \mathrm{E}[x])^2]^{3/2}} \\ \text{Curtosis} &:= \frac{\mu_4}{\mu_2^2} = \frac{\mathrm{E}[(x - \mathrm{E}[x])^4]}{\mathrm{E}[(x - \mathrm{E}[x])^2]^2} \end{aligned} \tag{A.24}$$

A.3.1 Moment Generating Functions

The moment generating function is a very useful tool for the explicit computation of moments. The *moment generating function* (in short *MGF*) of a random variable x with density function $\mathrm{pdf}(x)$ is defined as the expectation of e^{sx} for an arbitrary real value s

$$G_x(s) = \mathrm{E}[e^{sx}] = \int_{-\infty}^{\infty} e^{sx}\,\mathrm{pdf}(x)dx \tag{A.25}$$

if this integral exists. This corresponds to the *Laplace transformation* of the pdf. Expanding the exponential function e^{sx} in its Taylor series, we see that the coefficient of s^n is determined by the n^{th} moment of the distribution:

$$\begin{aligned} G_x(s) &= \int_{-\infty}^{\infty} \underbrace{\sum_{n=0}^{\infty} \frac{1}{n!}\, s^n x^n}_{e^{sx}}\,\mathrm{pdf}(x)dx \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \int_{-\infty}^{\infty} x^n\,\mathrm{pdf}(x)dx \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!}\mathrm{E}[x^n] \end{aligned} \tag{A.26}$$

Differentiating the moment generating function with respect to s at the point $s = 0$ yields all moments of the distribution (and thus the name):

$$\left.\frac{\partial^n G_x(s)}{\partial s^n}\right|_{s=0} = \mathrm{E}[x^n] \tag{A.27}$$

This extraordinarily useful fact can be shown as follows:

$$\begin{aligned} \left.\frac{\partial^n G_x(s)}{\partial s^n}\right|_{s=0} &= \left.\frac{\partial^n}{\partial s^n} \sum_{i=0}^{\infty} \frac{1}{i!} s^i \mathrm{E}[x^i]\right|_{s=0} \\ &= \left.\frac{1}{i!}\frac{\partial^{n-1}}{\partial s^{n-1}} \sum_{i=0}^{\infty} \mathrm{E}[x^i]\frac{\partial s^i}{\partial s}\right|_{s=0} \\ &= \left.\frac{1}{i!}\frac{\partial^{n-1}}{\partial s^{n-1}} \sum_{i=1}^{\infty} \mathrm{E}[x^i] i\, s^{i-1}\right|_{s=0} \\ &= \left.\frac{1}{i!}\frac{\partial^{n-2}}{\partial s^{n-2}} \sum_{i=2}^{\infty} \mathrm{E}[x^i] i(i-1)\, s^{i-2}\right|_{s=0} \\ &= \cdots \\ &= \left.\frac{1}{i!} \sum_{i=n}^{\infty} \mathrm{E}[x^i] i(i-1)\cdots(i-n+1) s^{i-n}\right|_{s=0} \end{aligned}$$

Notice that the lower limit in the sum increases by one each time a derivative is taken. The first term of the sum is always independent of the differentiating variable. For

$i = 0$, for example, $\partial s^i/\partial s = \partial s^0/\partial s = \partial 1/\partial s = 0$, and so on. In the last step, all derivatives have been performed. The expression can now be evaluated at $s = 0$. Naturally, $s^{i-n} = 0$ for $s = 0$ and $i > n$. Thus, *only* the first summand where $i = n$ makes a contribution to the sum since in this term we have $s^{i-n} = s^0 = 1$. Setting $i = n$ in this term immediately yields Equation A.27.

The *central* moments defined in Equation A.23 can likewise be calculated using the moment generating function: If a random number x has a distribution $\mathrm{pdf}(x)$ with expectation $\mu = \mathrm{E}[x]$, then the moments of the random number $\widetilde{x} := x - \mu$ are exactly equal to the *central* moments of x. But for the moments of $\widetilde{x}$ we can use the MGF of the distribution of $\widetilde{x}$.

$$\begin{aligned} G_{\widetilde{x}}(s) &= \int_{-\infty}^{\infty} e^{s\widetilde{x}}\,\mathrm{pdf}(\widetilde{x})d\widetilde{x} = \int_{-\infty}^{\infty} e^{s\widetilde{x}}\,\mathrm{pdf}(x)d\widetilde{x} \\ &= \int_{-\infty}^{\infty} e^{s(x-\mu)}\,\mathrm{pdf}(x)dx = e^{-s\mu}\int_{-\infty}^{\infty} e^{sx}\,\mathrm{pdf}(x)dx \end{aligned}$$

The first step is just Definition A.25. For the second step we made use of the fact that if the difference between two random numbers is just a constant they must have the same distribution[1]. In the third step we used that $\widetilde{x} = x - \mu$ means for the differentials $d\widetilde{x} = dx$. Thus the MGF for the central moments is simply $e^{-s\mu}$ times the MGF for the (ordinary) moments where μ denotes the first (ordinary) moment:

$$G_{x-\mu}(s) = e^{-s\mu}G_x(s) \tag{A.28}$$

With this MGF the central moments defined in Equation A.23 can be calculated completely analogously to Equation A.27:

$$\mathrm{E}[(x - \mathrm{E}[x])^n] = \frac{\partial^n}{\partial s^n}\exp\left(-s\mathrm{E}[x]\right)G_x(s)\bigg|_{s=0} \tag{A.29}$$

The general procedure for calculating central moments is thus: first calculate the expectation using Equation A.27. Then insert the result into Equation A.29 for the central moments.

For many distributions an explicit analytical expression for the MGF can be obtained using the integral representation in Equation A.25. This will be demonstrated below for several important distributions. Having obtained such an expression, the moments can be calculated by simply differentiating this function as indicated in Equation A.27.

The MGF has another very useful property: If two random variables x and y are independent then

$$G_{x+y}(s) = G_x(s)\,G_y(s) \tag{A.30}$$

The *distribution* of a *sum* of independent random variables is generally very difficult to determine, even when the distributions of the individual random variables in the sum are known[2]. The *MGF* of such a sum, in contrast, can be calculated quite easily

[1] This is trivial: if $\widetilde{x} = x - \mu$ then the probability for $\widetilde{x} < a - \mu$ is of course the same as the probability for $x < a$.

[2] Only in a few special cases, for example when each of the random variables is normally distributed, can the distribution of the sum be easily specified.

by taking the product of the MGFs of each of the distributions! This is the most useful property of the moment generating function. In Equation A.30, each of the random variables in the sum can by all means be governed by completely *different* distributions. The only condition which needs to be satisfied for Equation A.30 to hold is the statistical *independence* of the random variables under consideration. Equation A.30 is quite simple to prove:

$$G_{x+y}(s) \equiv \mathrm{E}[e^{s(x+s)}] = \mathrm{E}[e^{sx}e^{sy}] = \mathrm{E}[e^{sx}]\mathrm{E}[e^{sy}] = G_x(s)\,G_Y(s)$$

This is a consequence of the fact that $\mathrm{E}[f(x)g(y)] = \mathrm{E}[f(x)]\mathrm{E}[g(y)]$ holds for arbitrary functions f, g of independent random variables x, y.

A further property of the MGF in connection with Equation A.30 is that for all non-stochastic values a, b and random variables x we have

$$G_{ax+b}(s) = e^{sb}G_x(as) \tag{A.31}$$

The proof of this result is also quite simple:

$$G_{ax+b}(s) = \mathrm{E}[e^{s(ax+b)}] = e^{bs}\mathrm{E}[e^{(as)x}] = e^{sb}G_x(as)$$

We have already encountered a special case of this in Equation A.28.

A.3.2 Characteristic Functions

Similar to the moment generating function, the *characteristic function* of a random variable x with probability density function $\mathrm{pdf}(x)$ is defined as the expectation of e^{isx} for an arbitrary real value s

$$\Phi_x(s) := \mathrm{E}[e^{isx}] = \int_{-\infty}^{\infty} e^{isx}\,\mathrm{pdf}(x)dx \tag{A.32}$$

Here i denotes the imaginary number satisfying $i^2 = -1$. This is just the *Fourier transformation* of the pdf .

The *Fourier transformation* of the cpf, the *cumulative* distribution function, is sometimes used as well. The results derived below hold for these functions as well. Thus, we will formulate the characteristic function more generally by writing

$$\Phi_x(s) = \int_{-\infty}^{\infty} e^{isx}f(x)dx$$

where the function f can be taken to be either $f(x) = \mathrm{pdf}(x)$ or $f(x) = \mathrm{cpf}(x)$.

The advantage of the characteristic function is that its inverse function, the *inverse Fourier transformation* always exists:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\Phi_x(s)ds \tag{A.33}$$

Thus, if Φ_x is known, then the *distribution* (pdf or cpf) can be computed directly (and not only its moments as was the case with the moment generating function). The

validity of Equation A.33 can be shown quite easily:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\Phi_x(s)ds &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\int_{-\infty}^{\infty} e^{isx'}f(x')dx'ds \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-is(x-x')}f(x')dx'ds \\ &= \int_{-\infty}^{\infty}\delta(x-x')f(x')dx' \\ &= f(x)\end{aligned}$$

where the *Dirac delta function* was used in the above derivation:

$$\delta(x-x^{'}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-is(x-x^{'})}ds$$

This is not a function in the strict sense but a so-called *distribution* with the defining property

$$\int_{-\infty}^{\infty}\delta(x-x')f(x')dx' := f(x)$$

It is precisely this property of the delta function that yields the invertibility of the Fourier transformation.

Analogously to the moment generating function, Equation A.30 holds for the characteristic function as well, i.e.

$$\Phi_{x+y}(s) = \Phi_x(s)\,\Phi_y(s) \tag{A.34}$$

for independent random variables x and y. Likewise, for non-stochastic values a, b and a random variable x

$$\Phi_{ax+b}(s) = e^{ibs}\Phi_x(as) \tag{A.35}$$

holds. The proof is completely analogous to that of Equation A.31

A.4 A Collection of Important Distributions

A.4.1 The Uniform Distribution

The simplest of all distributions is the *uniform distribution*. This distribution has both a continuous and a discrete version. A random variable is said to be *uniformly distributed* if its distribution density is constant. The normalizing equation A.2 implies immediately that the distribution density for a discrete random variable with n possible outcomes, respectively a continuously distributed random variable taking on values in

an interval $[a, b]$ is given by:

$$1 = \sum_{i=1}^{n} \underbrace{p(x_i)}_{\text{Constant}} = np \quad \Longrightarrow \quad p = \frac{1}{n}$$

$$1 = \int_{-\infty}^{\infty} \underbrace{p(x)}_{\text{Constant}} \, dx = p \int_{a}^{b} dx = p(b-a) \Longrightarrow \quad p = \frac{1}{b-a} \tag{A.36}$$

The expectation and the variance of the continuous form of the uniform distribution are

$$\mathrm{E}[x] = \int_{-\infty}^{\infty} xp(x)dx = \frac{1}{b-a} \int_{a}^{b} xdx = \frac{1}{b-a} \frac{b^2-a^2}{2} = \frac{a+b}{2}$$

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} (x - \mathrm{E}[x])^2 \, p(x)dx = \frac{1}{b-a} \int_{a}^{b} \left(x - \frac{a+b}{2} \right)^2 dx = \frac{(b-a)^2}{12} \tag{A.37}$$

Most random number generators generate uniformly distributed random numbers between 0 and 1. For $a = 0$ and $b = 1$, the expectation and the variance of the uniform distribution is given by $1/2$ and $1/12$, respectively.

The moment generating function of the uniform distribution is by definition A.25

$$G_x(s) \equiv \int_{-\infty}^{\infty} e^{sx} p(x)dx = \frac{1}{b-a} \int_{a}^{b} e^{sx} dx = \frac{1}{b-a} \left[\frac{1}{s} e^{sx} \right]_{x=a}^{x=b}$$

Thus

$$G_x(s) = \frac{e^{bs} - e^{as}}{s(b-a)} \tag{A.38}$$

Naive differentiation of G_x with respect to s does not lead us directly to the desired result since the factor s appears in the denominator forbidding us from setting $s = 0$. The function can, however, be written in such a way that it is well defined for this value since as s approaches zero the numerator converges towards zero faster than the denominator (since the exponential function converges towards 1 faster than s converges to zero). This can be seen by expanding the exponential function in its Taylor series as in Equation A.38:

$$\frac{1}{s} e^{bs} = \frac{1}{s} \sum_{n=0}^{\infty} \frac{1}{n!} b^n s^n = \sum_{n=0}^{\infty} \frac{1}{n!} b^n s^{n-1} = \frac{1}{s} + \sum_{i=0}^{\infty} \frac{b^{i+1}}{(i+1)} \frac{s^i}{i!}$$

and analogously for e^{as}/s. Taking the difference of these two expressions, the $1/s$ term disappears and we are left with an equivalent form of the MGF in Equation A.38. This is given by

$$G_x(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

Comparison of the coefficients of s^n with Equation A.26 yields *all* of the moments immediately

$$\mathrm{E}[x^n] = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

For example, the first moment ($n = 1$) is simply

$$\mathrm{E}[x] = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

which agrees with the result in Equation A.37.

The characteristic function, A.32, of the continuous uniform distribution function is given by

$$\Phi_x(s) \equiv \int_{-\infty}^{\infty} e^{isx} \frac{1}{b-a} dx = \frac{e^{ibs} - e^{ias}}{is(b-a)}$$

This equation can be found in every table of Fourier-transformed functions since it is just the Fourier transform of a constant. Note that the characteristic function can also be found by replacing s with is in the moment generating function Equation A.38.

A.4.2 The Binomial Distribution and the Bernoulli Experiment

Suppose that an experiment has only two possible outcomes (for example, heads or tails when tossing a coin or up or down in one time step of a binomial model) and the probability of one of the two outcomes (for example, "heads" or "up") of the experiment is p. Then the normalizing equation A.2 implies that the probability of the alternate outcome ("tails" or "down") is $1 - p$. In mathematics, such an experiment is referred to as a *Bernoulli experiment*.

If p is independent of the number of experimental trials then the probability of the observing exactly j outcomes associated with the probability p in n trials is given by the *binomial distribution*.

$$p_n(j) = \left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} \tag{A.39}$$

Obviously, this is a discrete distribution taking on $n + 1$ possible values. The term $p^j(1-p)^{n-j}$ is the probability that the result of the n Bernoulli trials will occur *in a certain order*, for instance, up, up, down, up, down, down, etc., with precisely j "ups". However, since the number of "up" terms does not depend on the order in which they appear, this probability must be multiplied by the number of all *permutations* having j "ups". The number of these permutations is given by the *binomial coefficient*[3]:

$$\left(\begin{array}{c} n \\ j \end{array} \right) := \frac{n!}{j!(n-j)!} = \frac{1 \times 2 \times \cdots \times n}{(1 \times 2 \times \cdots \times j)(1 \times 2 \times \cdots \times n - j)} \tag{A.40}$$

This is a result from the theory of combinations which has long been known.

[3]this reads "n choose j" or "n over j".

The probability that "up" will be observed *at least* k times is naturally the sum from k to n of the density defined in Equation A.39. This yields precisely the binomial probability arising in the recombinant binomial trees for European options:

$$B_{n,p}(j \geq k) = \sum_{j=k}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} \tag{A.41}$$

The moment generating function of the binomial distribution is, by the definition in Equation A.25,

$$\begin{aligned} G_{B_{n,p}}(s) \equiv \mathrm{E}\left[e^{sx}\right] &= \sum_{j=0}^{n} e^{sj} p_n(j) \\ &= \sum_{j=0}^{n} e^{sj} \left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} \\ &= \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) \left[pe^s\right]^j \left[1-p\right]^{n-j} \end{aligned}$$

Note that this is precisely the function appearing in the *binomial formula* for the n^{th} power of a sum $(a+b)^n$ where $a = pe^s$ and $b = 1-p$. Thus the moment generating function of the binomial distribution is simply

$$G_{B_{n,p}}(s) = \left(pe^s + 1 - p\right)^n \tag{A.42}$$

According to Equation A.27 all of the moments can now be calculated directly by differentiating the MGF with respect to s. Doing so yields, for example, the expectation as

$$\mathrm{E}[j] = \left. \frac{\partial G_{B_{n,p}}(s)}{\partial s} \right|_{s=0} = \left. n \left(pe^s + 1 - p\right)^{n-1} pe^s \right|_{s=0} = np$$

The second moment can be computed analogously:

$$\begin{aligned} \mathrm{E}[j^2] = \left. \frac{\partial^2 G_{B_{n,p}}(s)}{\partial s^2} \right|_{s=0} &= np \left. \frac{\partial}{\partial s} \left(pe^s + 1 - p\right)^{n-1} e^s \right|_{s=0} \\ &= np\,(n-1) \left. \left(pe^s + 1 - p\right)^{n-2} pe^s e^s \right|_{s=0} \\ &+ np \left. \left(pe^s + 1 - p\right)^{n-1} e^s \right|_{s=0} \\ &= n(n-1)p^2 + np \end{aligned}$$

Thus, according to Equation A.7, the second *central* moment, the variance of the binomial distribution, becomes

$$\mathrm{var}\left[j\right] = \mathrm{E}[j^2] - (\mathrm{E}[j])^2 = np(1-p)$$

Summarizing, the expectation and variance of the binomial distribution are given by

$$\begin{aligned} \mathrm{E}[j] &\equiv \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} j = np \\ \mathrm{var}[j] &\equiv \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) p^j (1-p)^{n-j} (j-np)^2 = np(1-p) \end{aligned} \tag{A.43}$$

Analogously to Equation A.42, the characteristic function of the binomial distribution, can be found from Equation A.32 to be

$$\Phi_{B_{n,p}}(s) = \left(pe^{is} + 1 - p\right)^n \tag{A.44}$$

A.4.3 The Normal Distribution and the Central Limit Theorem

It is well known that the sum of random variables is itself a random variable. One of the most important theorems of mathematical statistics, the *central limit theorem*, makes a broad statement on the sum of random variables. The intuitive content of this theorem can be expressed as follow:

Theorem 10 (Central Limit) *The sum of a large number of independent random variables is approximately normally distributed, regardless of how the individual random variables are distributed, if the contribution of each random variable to the sum is almost negligible.*

This theorem is the reason for the extraordinary importance of the normal distribution above all others. For instances, the *mean* of some random variables is defined as the *sum* of these random variables divided by the number of variables. Thus, such means are *always* approximately normally distributed according to the central limit theorem, *regardless of the distribution of the random variables*. The fact that such sums are often *not* normally distributed can only mean that the assumptions in the statement of the theorem are not satisfied. Either the measured variables are not (purely) random variables and/or they are not completely independent (uncorrelated). The most common reason, however, is an insufficient number of available trial results. Because only in the *limit*, i.e. for an *infinite* number of random variables, does the central limit theorem hold exactly and not merely as an approximation (hence the name central *limit* theorem).

The *normal distribution*, also called *Gaussian distribution*, is a continuous distribution which is completely determined by two parameters (denoted by μ and σ). The density has the explicit form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.45}$$

The expectation and the variance of the normal distribution according to the definition in Equation A.4 and Equation A.5 are given by

$$\begin{aligned} \mathrm{E}[x] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu \\ \mathrm{var}[x] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2 \end{aligned} \tag{A.46}$$

Hence, the two parameters μ and σ are equal to the expectation and the standard deviation of the distribution, respectively. This is quite practical, since through observing the results of random trials x, the expectation and the variance can be approximated by measuring the mean of x and x^2 and applying Equation A.7, thus obtaining an approximation of the entire distribution. The density of the distribution can be written as

$$p(x) = \frac{1}{\sqrt{2\pi \mathrm{E}\left[(x - \mathrm{E}[x])^2\right]}} \exp\left\{ -\frac{1}{2} \frac{(x - \mathrm{E}[x])^2}{\mathrm{E}\left[(x - \mathrm{E}[x])^2\right]} \right\} \tag{A.47}$$

The ratio of the square of the deviation of the random variable from its expectation to the *expectation* of this same factor appears in the argument of the exponential function.

The probability $P(x \leq a)$ of the event that a random variable x, distributed according to Equation A.45 with parameter values μ and σ, will be less than or equal to a given value a is the cumulative probability cpf of the normal distribution evaluated at a, frequently denoted by $\mathrm{N}_{\mu,\sigma}(a)$. Thus, according to Equation A.1 we have

$$\mathrm{N}_{\mu,\sigma}(a) = \mathrm{cpf}(a) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{A.48}$$

The following notation is also often used to express the fact that a random variable x is normally distributed with expectation μ and variance σ^2

$$x \sim \mathrm{N}(\mu, \sigma^2) \tag{A.49}$$

The reader should be careful not to confuse this expression with that for the cumulative distribution function $\mathrm{N}_{\mu,\sigma}(a)$ in Equation A.48.

The integral in the definition of the *cumulative normal distribution* cannot be calculated as a closed form expression. But through a simple change of variable

$$\begin{aligned} y &:= \frac{x-\mu}{\sigma} \Longrightarrow dx = \sigma dy \\ y(x = -\infty) &= -\infty \quad , \quad y(x = a) = \frac{a-\mu}{\sigma} \end{aligned} \tag{A.50}$$

the normal distribution can be transformed into the *standard* normal distribution. This is the normal distribution with expectation 0 and variance 1, i.e. with the density function

$$p(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \tag{A.51}$$

In the notation introduced in Equation A.49 we write

$$y \sim \mathrm{N}(0,1)$$

Random variables which are standard normally distributed with density function A.51 form the basis of many random walk models applied in this book. The probability $P(x \leq a)$ that a standard normally distributed random variable x will be less than or equal to a given number a is called the *cumulative standard normal distribution* $\mathrm{N}(a)$:

$$\mathrm{N}(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx \tag{A.52}$$

This distribution appears for instance in the Black-Scholes formula.

The density EquationA.51 of the standard normal distribution is symmetric about zero and thus $p(x) = p(-x)$. From this we derive a very useful *symmetry* relation holding for the standard normal distribution, namely

$$\mathrm{N}(-a) = 1 - \mathrm{N}(a) \tag{A.53}$$

This symmetry is so important in practical applications that we will provide the proof here. We first write

$$\mathrm{N}(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx}_{1} - \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{x^2}{2}} dx$$

The first integral equals 1 because of Equation A.2. In the second integral, we make the substitution

$$u := -x \Rightarrow du = -dx$$
$$u(x = a) = -a \ , \quad u(x = \infty) = -\infty$$

thus obtaining

$$\mathrm{N}(a) = 1 + \frac{1}{\sqrt{2\pi}} \int_{-a}^{-\infty} e^{-\frac{x^2}{2}} du = 1 - \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-\frac{x^2}{2}} du}_{\mathrm{N}(-a)}$$

which proves the symmetry relation in Equation A.53.

The integral in Equation A.52 cannot be computed explicitly. However, there exist tables and numerical routines for the computation of the cumulative standard normal distribution. A very good polynomial approximation, exact up to six decimal places, is (see for example [1])

$$\mathrm{N}(x) = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sum_{i=1}^{5} a_i y^i \qquad \text{for } x \geq 0 \tag{A.54}$$

where

$$\begin{aligned} y &= \frac{1}{1 + 0.2316419\,x} \\ a_1 &= 0.319381530\ , \quad a_2 = -0.356563782 \\ a_3 &= 1.781477937\ , \quad a_4 = -1.821255978 \\ a_5 &= 1.330274429 \end{aligned}$$

This approximation holds only for non-negative values $x \geq 0$. But the values for $x < 0$ can simply be obtained by applying the symmetry relation A.53.

Calculations with the *standard* normal distribution are thus quite simple. A frequently used method is therefore to transform normally distributed random variables into *standard* normal random variables via a transformation as in Equation A.50. Then all necessary calculations are performed using tools like Equations A.54 and A.53. Having completed the calculation, an inverse transformation can be performed to determine the original variables. For example, a 99% confidence interval for a standard normal distribution is

$$0,99 = P(y \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{y^2}{2}} dy \Rightarrow a \approx 2,326$$

This is the upper bound for the *standard* normally distributed random variable y. Performing the inverse transformation back to the variable x according to Equation A.50 yields the boundary of the confidence interval for the original random variable: the expectation plus 2.326 standard deviations.

The moment generating function of the standard normal distribution is, by Definition A.25

$$\begin{aligned} G_{\mathrm{N}(0,1)}(s) = \mathrm{E}[e^{sx}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2 - 2sx}{2}\right\} dx \\ &= \exp\left(\frac{1}{2}s^2\right) \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-s)^2}{2}\right\} dx}_{1} \end{aligned}$$

where we have used the method of completing the squares in the last step[4]. The remaining integral is precisely the probability that a normally distributed random variable $x \sim \mathrm{N}(s, 1)$ will take on any arbitrary value. Equation A.2 implies that this probability is 1. The moment generating function of the *standard* normal distribution is thus simply

$$G_{\mathrm{N}(0,1)}(s) = \exp\left(\frac{1}{2}s^2\right) \tag{A.55}$$

[4] $x^2 - 2sx = (x-s)^2 - s^2$

The moment generating function of the normal distribution with expectation μ and variance σ now follows immediately from the transformation A.50 with Equation A.31

$$G_{\mathrm{N}(\mu,\sigma^2)}(s) = e^{\mu s} G_{\mathrm{N}(0,1)}(\sigma s) = \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right) \tag{A.56}$$

From this equation, all *moments* of a normal distribution can be calculated by means of Equation A.27. The *central* moments are given by Equation A.29:

$$\begin{aligned}\mathrm{E}[(x-\mathrm{E}[x])^n] &= \frac{\partial^n}{\partial s^n} \exp(-s\mu) \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right)\Big|_{s=0} \\ &= \frac{\partial^n}{\partial s^n} \exp\left(\frac{1}{2}\sigma^2 s^2\right)\Big|_{s=0}\end{aligned}$$

The second central moment, i.e. the variance is

$$\begin{aligned}\mathrm{E}[(x-\mathrm{E}[x])^2] &= \frac{\partial^2}{\partial s^2} e^{\sigma^2 s^2/2}\Big|_{s=0} = \frac{\partial}{\partial s}\sigma^2 s e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= \sigma^2 e^{\sigma^2 s^2/2} + \left(\sigma^2 s\right)^2 e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= \left(1+\sigma^2 s^2\right)\sigma^2 e^{\sigma^2 s^2/2}\Big|_{s=0} = \sigma^2\end{aligned}$$

Differentiating the result in the penultimate step yields the third central moment

$$\begin{aligned}\mathrm{E}[(x-\mathrm{E}[x])^3] &= \frac{\partial^3}{\partial s^3} e^{\sigma^2 s^2/2}\Big|_{s=0} = \frac{\partial}{\partial s}\left(1+\sigma^2 s^2\right)\sigma^2 e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= 2s\sigma^4 e^{\sigma^2 s^2/2} + \left(1+\sigma^2 s^2\right) s\sigma^4 e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= \left(3s+\sigma^2 s^3\right)\sigma^4 e^{\sigma^2 s^2/2}\Big|_{s=0} = 0\end{aligned}$$

Again, differentiating the result in the penultimate step yields the next (i.e. fourth) central moment

$$\begin{aligned}\mathrm{E}[(x-\mathrm{E}[x])^4] &= \frac{\partial^4}{\partial s^4} e^{\sigma^2 s^2/2}\Big|_{s=0} = \frac{\partial}{\partial s}\left(3s+\sigma^2 s^3\right)\sigma^4 e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= \left(3+3\sigma^2 s^2\right)\sigma^4 e^{\sigma^2 s^2/2} + \left(3s+\sigma^2 s^3\right)\sigma^6 s e^{\sigma^2 s^2/2}\Big|_{s=0} \\ &= \left(3+3\sigma^2 s^2+3\sigma^2 s+\sigma^4 s^3\right)\sigma^4 e^{\sigma^2 s^2/2}\Big|_{s=0} = 3\sigma^4\end{aligned}$$

We summarize these results for reference:

$$\begin{aligned}\mathrm{E}[x] &= \mu \\ \mathrm{E}[(x-\mathrm{E}[x])^2] &= \sigma^2 \\ \mathrm{E}[(x-\mathrm{E}[x])^4] &= 3\sigma^4 \\ \mathrm{E}[(x-\mathrm{E}[x])^n] &= 0 \quad \text{for all odd } n > 2\end{aligned} \tag{A.57}$$

From these moments the *skewness* and the *curtosis* of the normal distribution follow directly from their respective definitions, Equation A.24:

$$\begin{aligned} \text{Skewness} &\equiv \frac{\mathrm{E}[(x-\mathrm{E}[x])^3]}{\mathrm{E}[(x-\mathrm{E}[x])^2]^{3/2}} = 0 \\ \text{Curtosis} &\equiv \frac{\mathrm{E}[(x-\mathrm{E}[x])^4]}{\mathrm{E}[(x-\mathrm{E}[x])^2]^2} = 3 \end{aligned} \tag{A.58}$$

The characteristic function A.32 of the standard normal density function can be found by replacing s with is in Equation A.55

$$\Phi_{\mathrm{N}(0,1)}(s) = \exp\left(-\frac{1}{2}s^2\right) \tag{A.59}$$

The characteristic function of a normal distribution with expectation μ and variance σ follows immediately from Equation A.35

$$\Phi_{\mathrm{N}(\mu,\sigma^2)}(s) = e^{i\mu s}\Phi_{\mathrm{N}(0,1)}(\sigma s) = \exp\left(i\mu s - \frac{1}{2}\sigma^2 s^2\right) \tag{A.60}$$

The Multivariate Normal Distribution

If two random variables $x_i, i = 1, 2$ are both normally distributed with expectation μ_i and covariance[5]

$$\mathrm{cov}\,[x_i, x_j] = \sigma_i \rho_{ij} \sigma_j$$

then their *joint probability distribution*, i.e. the probability that both random numbers simultaneously will have certain values, is given by the *bivariate normal distribution* with the density

$$p(x_1, x_2) = \frac{\exp\left\{-\frac{1}{1-\rho_{1,2}^2}\frac{1}{2}\sum_{i,j=1,2}\frac{x_i-\mu_i}{\sigma_i}\rho_{ij}\frac{x_j-\mu_j}{\sigma_j}\right\}}{\sqrt{1-\rho_{1,2}^2}\prod_{i=1,2}\sqrt{2\pi\sigma_i^2}}$$

With this density, the joint (cumulative) probability, as defined in Equation A.8, for $x_1 < a$ and $x_2 < b$ is

$$P(x_1 < a, x_2 < b) = \frac{\int_{-\infty}^{a} dx_1 \int_{-\infty}^{b} dx_2 \; \exp\left\{-\frac{1}{1-\rho_{1,2}^2}\frac{1}{2}\sum_{i,j=1}^{2}\frac{x_i-\mu_i}{\sigma_i}\rho_{ij}\frac{x_j-\mu_j}{\sigma_j}\right\}}{\sqrt{1-\rho_{12}^2}\prod_{i=1}^{2}\sqrt{2\pi\sigma_i^2}}$$

If the two random variables are *uncorrelated*, i.e. if

$$\rho_{ij} = \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

[5]This means that they have variances σ_1^2 and σ_2^2 and the correlation between the two is ρ_{12} (see Equation A.14).

where δ_{ij} denotes the *Kronecker delta*, then the joint distribution of the variables is equal to the product of the distributions of each individual variable:

$$\begin{aligned} p(x_1, x_2) &= \frac{\exp\left\{-\frac{1}{1-0}\frac{1}{2}\sum_{i,j=1,2}\frac{x_i-\mu_i}{\sigma_i}\delta_{ij}\frac{x_j-\mu_j}{\sigma_j}\right\}}{\sqrt{1-0}\prod_{i=1,2}\sqrt{2\pi\sigma_i^2}} \\ &= \frac{\exp\left\{-\frac{1}{2}\sum_{i=1,2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\}}{\prod_{i=1,2}\sqrt{2\pi\sigma_i^2}} \\ &= \prod_{i=1,2}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\} \\ &= \prod_{i=1,2} p(x_i) \end{aligned} \tag{A.61}$$

If the joint probability density of two random variables x_1 and x_2 fulfills

$$p(x_1, x_2) = p(x_1)p(x_2) \tag{A.62}$$

then the two random variables are said to be *independent*. As we have just seen, a *necessary* condition for independence is that the two variables are uncorrelated. However, this is generally not a *sufficient* condition, i.e. uncorrelated random variables are not always independent. But – as we have just seen – in the special case of normally distributed random variables independence and uncorrelation are equivalent.

A.4.4 The Lognormal Distribution

A random variable is *lognormally distributed* if its logarithm is normally distributed. Since the lognormal distribution is defined through the normal distribution, it is also continuous and completely determined by the two parameters μ and σ. The precise definition of the *lognormal distribution* can be stated as follows: let x be a lognormally distributed random variable with parameters μ and σ. Then the probability $\mathrm{H}_{\mu,\sigma}(a)$ that x will be less than a given value a is equal to the cumulative distribution function of the normal distribution, Equation A.48, with the same parameters, evaluated at $\ln(a)$:

$$\mathrm{H}_{\mu,\sigma}(a) := \mathrm{N}_{\mu,\sigma}(\ln(a)) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\ln(a)} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \tag{A.63}$$

The density function of the lognormal distribution can be derived from the above definition. It has the explicit form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}\exp\left\{-\frac{1}{2}\frac{(\ln(x)-\mu)^2}{\sigma^2}\right\} \tag{A.64}$$

Note the factor $1/x$. The density is thus not merely obtained by substituting x with $\ln(x)$ in Equation A.45. With this density function we can express $\mathrm{H}_{\mu,\sigma}$ explicitly as

$$\mathrm{H}_{\mu,\sigma}(a) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^a \exp\left\{-\frac{1}{2}\frac{(\ln(x)-\mu)^2}{\sigma^2}\right\}\frac{1}{x}dx \tag{A.65}$$

Note that the lower limit in the integral is zero. The range of such a lognormally distributed random number lies only between zero and infinity.

With the simple change in variable given by $u = \ln(x)$ we can immediately verify that the cumulative distribution function $\mathrm{H}_{\mu,\sigma}(a)$ actually satisfies the definition in Equation A.63:

$$\begin{aligned}
\mathrm{H}_{\mu,\sigma}(a) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^a \exp\left\{-\frac{1}{2}\frac{(\ln(x)-\mu)^2}{\sigma^2}\right\}\frac{1}{x}dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\ln(0)=-\infty}^{\ln(a)} \exp\left\{-\frac{1}{2}\frac{(u-\mu)^2}{\sigma^2}\right\} du = \mathrm{N}_{\mu,\sigma}(\ln(a)) \\
\text{where } u := \ln(x) &\Longrightarrow du = \frac{1}{x}dx, \quad x = e^u
\end{aligned}$$

The expectation and variance of the lognormal distribution can be calculated using Equations A.4 and A.5 as

$$\begin{aligned}
\mathrm{E}[x] &= \int_0^\infty xp(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}dx = e^{\mu+\sigma^2/2} \\
\mathrm{var}[x] &= \int_0^\infty (x - \mathrm{E}[x])^2 p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty (x - e^{\mu+\sigma^2/2})^2 \frac{1}{x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}dx \\
&= e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2})
\end{aligned} \tag{A.66}$$

In general, the moments of the lognormal distribution are given by

$$\mathrm{E}[x^n] = \exp\left(n\mu + n^2\sigma^2/2\right) \quad \text{for } n = 1, 2, \ldots \tag{A.67}$$

A.4.5 The Gamma Distribution

The *gamma distribution* is an important distribution because it encompasses a whole class of different distributions (which includes the exponential distributions and the χ^2-distribution, for example). Like the lognormal distribution, the range of the gamma distribution consists solely of the non-negative real numbers. A random variable x has a gamma distribution with parameters λ and t if it is governed by the following distribution density

$$\mathrm{pdf}(x) = \frac{1}{\Gamma(t)}\lambda^t x^{t-1} e^{-\lambda x} \quad \text{where} \quad \lambda > 0,\ t > 0 \quad \text{and} \quad x \in [0, \infty[\tag{A.68}$$

where $\Gamma(t)$ denotes the *gamma function*. This function is defined by

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \tag{A.69}$$

A description of its properties can be found in any mathematical collection of special mathematical functions (see for instance [1]). An important property of the gamma function is the recursion

$$\Gamma(t+1) = t\,\Gamma(t) \tag{A.70}$$

This allows the gamma function to be interpreted as a generalization of the factorial operation. Two special function values which often serve as the initial values in the above recursion relation are given by

$$\Gamma(1) = 1 \quad , \quad \Gamma(\frac{1}{2}) = \sqrt{\pi} \tag{A.71}$$

It follows that the gamma function for whole numbers t is in fact nothing other than the factorial:

$$\Gamma(n) = (n-1)! \quad \text{for} \quad n = 1, 2, \ldots$$

Furthermore, this function has a symmetry property which allows it to be evaluated for negative values if the function evaluated at the corresponding positive value is known:

$$\Gamma(-x) = -\frac{\pi}{x \sin(\pi x)} \frac{1}{\Gamma(x)}$$

From this symmetry relation, we see immediately that the gamma function has a simple pole at each negative integer value of x. It is well defined for all positive values. Fortunately, in our consideration of the gamma distribution, Equation A.68, our needs are restricted to the positive arguments of the gamma function only.

The explicit form of the moment generating function of the gamma distribution can be calculated quite easily. Consider the definition A.25. Then for the gamma distribution, we have

$$G_{\Gamma(t,\lambda)}(s) = \mathrm{E}[e^{sx}] = \frac{\lambda^t}{\Gamma(t)} \int_0^\infty x^{t-1} e^{-(\lambda - s)x} dx$$

The integral appearing in this expression is quite similar to that in the definition of the gamma function, Equation A.69. We make the substitution

$$u := (\lambda - s)x \Longrightarrow du = (\lambda - s)dx,$$
$$u(x = 0) = 0 \ , \quad u(x = \infty) = \infty \text{ for } s < \lambda$$

The condition $s < \lambda$ is required in order for the upper limit of the integral to remain equal to $+\infty$. From Equation A.27 we see that the s of interest are in a small neighborhood of zero. Since λ in the distribution Equation A.68 is strictly greater than zero, the condition $s < \lambda$ is no obstacle for our purposes here. A simple substitution yields

$$G_{\Gamma(t,\lambda)}(s) = \frac{\lambda^t}{\Gamma(t)} \int_0^\infty \frac{1}{(\lambda - s)^{t-1}} u^{t-1} e^{-u} \frac{1}{\lambda - s} du$$
$$= \frac{\lambda^t}{\Gamma(t)} \frac{1}{(\lambda - s)^t} \underbrace{\int_0^\infty u^{t-1} e^{-u} du}_{\Gamma(t)}$$

The moment generating function of the gamma distribution thus has the following simple form:

$$G_{\Gamma(t,\lambda)}(s) = \left(\frac{\lambda}{\lambda - s}\right)^t \text{ with } \lambda > 0,\ t > 0,\ s < \lambda \tag{A.72}$$

Replacing s with is in Equation A.72 yields the characteristic function given in Equation A.32 for the gamma distribution

$$\Phi_{\Gamma(t,\lambda)}(s) = \left(\frac{\lambda}{\lambda - is}\right)^t \tag{A.73}$$

All *moments* can be obtained as indicated in Equation A.27 directly by differentiating the expression in Equation A.72 with respect to s. For example, the expectation becomes

$$\mathrm{E}[x] = \left.\frac{\partial G_{\Gamma(t,\lambda)}(s)}{\partial s}\right|_{s=0} = \lambda^t t \left.\frac{1}{(\lambda - s)^{t+1}}\right|_{s=0} = \frac{t}{\lambda} \tag{A.74}$$

The second moment is

$$\mathrm{E}[x^2] = \left.\frac{\partial^2 G_{\Gamma(t,\lambda)}(s)}{\partial s^2}\right|_{s=0} = \lambda^t t(t+1) \left.\frac{1}{(\lambda - s)^{t+2}}\right|_{s=0} = \frac{t(t+1)}{\lambda^2}$$

Proceeding analogously, *all* moments can be explicitly calculated for the gamma distribution:

$$\mathrm{E}[x^n] = \frac{t(t+1)\cdots(t+n-1)}{\lambda^n} \quad \text{for } n = 1, 2, \ldots \tag{A.75}$$

The variance is then

$$\mathrm{var}\,[x] = \mathrm{E}[x^2] - \mathrm{E}[x]^2 = \frac{t}{\lambda^2} \tag{A.76}$$

A.4.6 The χ^2-Distribution

As was shown in Section A.4.3, a sum of normally distributed random variables is itself normally distributed. A situation frequently encountered (for example, in a value at risk computation or the determination of variances from historical data) involves taking sums of the *squares* of random variables. If the random variables whose squares are added have a standard normal distribution, the distribution of the resulting random

variable is easily determined: the sum of n *squared, independent, standard normally distributed random variables,* $x_i, (i = 1, \ldots, n)$ has a distribution known as the χ^2-*distribution with* n *degrees of freedom*

$$x_i \sim \mathrm{N}(0,1)\ ,\ \ i = 1, \ldots, n\ ,\quad x_i\ \ iid\ \Longrightarrow \sum_{i=1}^{n} x_i^2 =: y \sim \chi^2(n) \tag{A.77}$$

It is essential that the random variables x_i in the above definition be independent. The *degree of freedom*, n, of the χ^2-distribution $\chi^2(n)$ can be intuitively thought of as the number of independent (standard normal) random variables which "make up" the random variable with the $\chi^2(n)$-distribution; thus the name "degree of freedom".

It is often the case that the above sum is taken over only one single element. If a random variable x has a standard normal distribution, then this random variable *squared* is governed by the χ^2*-distribution with one degree of freedom.* We write

$$x \sim \mathrm{N}(0,1) \Longrightarrow x^2 =: y \sim \chi^2(1) \tag{A.78}$$

In fact, $\chi^2(1)$ (or its non-central counterpart, see below) is sufficient for almost all our needs if we are working with moment generating functions since via Equation A.30 and A.31, the MGFs of sums of independent random variables can be written as products of MGFs of $\chi^2(1)$.

The MGF also proves to be a helpful tool in calculating the density function of the χ^2-distribution. We first generalize the statement of the problem somewhat, keeping in mind that our goal is to arrive at the MGF of $\chi^2(1)$.

Suppose a random variable x is distributed according to some known distribution whose distribution function we denote by pdf_x. Then, for any arbitrary function f, the random variable $y = f(x)$ is distributed according to another (in general, unknown) distribution with distribution function pdf_y. The MGF of the unknown distribution pdf_y is the expectation of a function g of y (specifically, $g(y) = e^{sy}$). According to Equation A.4, the expectation of a function g of y is given by

$$\mathrm{E}\left[g(y)\right]_{\mathrm{pdf}_y} = \int g(y)\,\mathrm{pdf}_y(y)dy$$

Here, the expectation is computed with respect to the (unknown) distribution pdf_y (this fact is emphasized by the subscript in the expectation). Since $y = f(x)$, the function

$$g(y) = g(f(x)) = h(x)$$

is also a function of x and therefore the expectation of h is according to Equation A.4

$$\mathrm{E}\left[h(x)\right]_{\mathrm{pdf}_x} = \int h(x)\,\mathrm{pdf}_x(x)dx$$

Here, the expectation is computed by means of the *known* distribution pdf_x. Because $h(x) = g(f(x)) = g(y)$ and $f(x) = y$, these expectations are the same

$$\int g(y)\,\mathrm{pdf}_y(y)dy = \int g(f(x))\,\mathrm{pdf}_x(x)dx$$

With choice of the function $g(y) = e^{sy}$, the left hand side becomes the MGF of the unknown distribution pdf_y

$$\begin{aligned} G_{\mathrm{pdf}_y}(s) = \mathrm{E}[e^{sy}]_{\mathrm{pdf}_y} &= \int e^{sy}\,\mathrm{pdf}_y(y)dy \\ &= \int e^{sf(x)}\,\mathrm{pdf}_x(x)dx = \mathrm{E}[e^{sf(x)}]_{\mathrm{pdf}_x} \end{aligned}$$

Thus, for every function f of a random variable x, the moment generating function of the distribution of $f(x)$ can be expressed through the expectation with respect to the distribution of x:

$$G_{\mathrm{pdf}_{f(x)}}(s) = \mathrm{E}[e^{sf(x)}]_{\mathrm{pdf}_x} \tag{A.79}$$

For the χ^2-distribution with one degree of freedom, we have $f(x) = x^2$ and obtain[6] the moment generating function of $\chi^2(1)$ as

$$G_{\chi^2(1)}(s) = \mathrm{E}[e^{sx^2}]_{\mathrm{N}(0,1)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx^2} e^{-x^2/2} dx \tag{A.80}$$

This integral can be solved explicitly using the substitution

$$\begin{aligned} u = x\sqrt{1-2s} \Rightarrow dx &= \frac{du}{\sqrt{1-2s}} \\ u(x=-\infty) &= -\infty \quad \text{für } s < 1/2 \\ u(x=+\infty) &= +\infty \quad \text{für } s < 1/2 \end{aligned} \tag{A.81}$$

Note that the integration limits remain the same only for $s < 1/2$.

$$\begin{aligned} \mathrm{E}[e^{sx^2}]_{\mathrm{N}(0,1)} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1-2s)x^2/2} dx \\ &= \frac{1}{\sqrt{1-2s}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du}_{1} \end{aligned}$$

The remaining integral is the probability that a standard normal random variable will assume *any* arbitrary value and, according to Equation A.2, equals one. Thus, the MGF is simply

$$G_{\chi^2(1)}(s) = \frac{1}{\sqrt{1-2s}} \tag{A.82}$$

Since the random variables in the sum appearing in Definition A.77 are all *independent* we can use Equation A.30 to immediately obtain the MGF for a χ^2-distribution with n degrees of freedom

$$G_{\chi^2(n)}(s) = \frac{1}{(1-2s)^{n/2}} \text{ for } n = 1, 2, \ldots \tag{A.83}$$

[6] with the notation: $\mathrm{pdf}_{f(x)} = \mathrm{pdf}_{x^2} = \chi^2(1)$, $\mathrm{pdf}_x = \mathrm{N}(0,1)$

However, a clearer understanding and, perhaps more importantly, an explicit expression for the density function of the χ^2-distribution is obtained by taking the following approach: if we succeed in transforming Equation A.80 into the form $\int e^{sx} p(x)dx$, this integral can be interpreted as the expectation of e^{sx} taken with respect to the probability density $p(x)$. This expectation is then by definition the MGF of $p(x)$. On the other hand, Equation A.80 is the MGF of $\chi^2(1)$-distribution. Since two distribution are exactly equal if they have the same MGF[7] we could then conclude that p is the density function of $\chi^2(1)$. In this way, knowledge of p would give us an explicit expression for the density function of $\chi^2(1)$. We wish to take this approach now.

First, we observe that in the integrand in Equation A.80, x appears only in the form x^2, i.e. the integral is symmetric about zero; thus we can write

$$G_{\chi^2(1)}(s) = \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{sx^2} e^{-x^2/2} dx$$

We then perform the following substitution in the integral

$$\begin{aligned} y := x^2 &\Longrightarrow y = +\sqrt{y} \text{ since } x \geq 0 \\ y(x=0) = 0, \;\; & y(x=\infty) = \infty \\ \frac{dy}{dx} = 2x &\Longrightarrow dx = \frac{1}{2\sqrt{y}} dy \end{aligned}$$

and obtain

$$\begin{aligned} G_{\chi^2(1)}(s) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{sy} e^{-y/2} \frac{1}{\sqrt{y}} dy \\ &= \int_0^\infty e^{sy} p(y) dy = \mathrm{E}[e^{sy}]_p \end{aligned}$$

We have thus attained our goal. The MGF of $\chi^2(1)$ is expressed as the expectation of e^{sy} with respect to a density function p. This density is the explicit expression for the density function of $\chi^2(1)$:

$$\mathrm{pdf}_{\chi^2(1)}(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} \text{ with } \;\; x \in [0, \infty[\tag{A.84}$$

The range of an χ^2-distributed random variable is thus restricted to the positive real numbers which makes sense since it is the *square* of a standard normally distributed random variable. Comparing this density to the density of the gamma distribution in Equation A.68 shows that $\chi^2(1)$ is equal to the gamma distribution for the parameters $\lambda = 1/2$ and $t = 1/2$. This can also be seen immediately by comparing their moment generating function in Equations A.82 and A.72; they are equal for $\lambda = 1/2$ and $t = 1/2$ and thus the associated density functions must be the same for these parameter values as well. Likewise, a comparison of the moment generating functions in Equations A.83

[7] Since the MGF defines all (infinitely many) moments of a distribution, the distribution is completely determined by the MGF. This implies that if two MGFs are equal then they are the MGF of the *same* distribution.

and A.72 shows that the χ^2-distribution with n degrees of freedom equals the gamma distribution with parameters $\lambda = 1/2$ and $t = n/2$:

$$\mathrm{pdf}_{\chi^2(n)}(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{\frac{n}{2}-1} e^{-x/2} \text{ with } \; x \in [0, \infty[\; , \;\; n = 1, 2, \ldots \qquad \text{(A.85)}$$

Due to the recursion relation in Equation A.70 with the initial values A.71, the gamma functions appearing here can be given explicitly as :

$$\Gamma(n/2) = \begin{cases} (n/2-1)! & \text{for even values of } n \\ (n/2-1)(n/2-2)(n/2-3)\cdots(1/2)\sqrt{\pi} & \text{for odd values of } n \end{cases}$$

The *moments* can be derived directly from Equation A.27 by differentiating the function in Equation A.83 with respect to s or simply from Equation A.75 with $\lambda = 1/2$ and $t = n/2$:

$$\begin{aligned} \mathrm{E}[x^k]_{\chi^2(n)} &= \frac{t(t+1)\cdots(t+k-1)}{\lambda^k} \\ &= \frac{\frac{n}{2}(\frac{n}{2}+1)\cdots(\frac{n}{2}+k-1)}{\left(\frac{1}{2}\right)^k} \\ &= 2^k \frac{n}{2}(\frac{n}{2}+1)\cdots(\frac{n}{2}+k-1) \\ &= n(n+2)\cdots(n+2(k-1)) \end{aligned}$$

Therefore

$$\mathrm{E}[x^k]_{\chi^2(n)} = \prod_{i=0}^{k-1}(n+2i) \qquad \text{(A.86)}$$

For example, the expectation and variance are given by

$$\mathrm{E}[x]_{\chi^2(n)} = n \; , \quad \mathrm{var}\,[x]_{\chi^2(n)} = 2n$$

From the definition in Equation A.32, the characteristic function of the χ^2-distribution can be obtained by replacing s with is in Equation A.83 or from Equation A.73 with $\lambda = 1/2$ and $t = n/2$.

$$\Phi_{\chi^2(n)}(s) = \frac{1}{(1-2is)^{n/2}} \text{ for } \; n = 1, 2, \ldots \qquad \text{(A.87)}$$

The Non-Central χ^2-Distribution

The χ^2-distribution described above is the distribution of a sum of n squared independent *standard* normal random numbers $x_i, (i = 1, \ldots, n)$. A slight but often needed generalization of this is the situation in which the random numbers x_i have expectations $\mu_i \neq 0$. The distribution of a sum of n squared random numbers of this type is

called the *non-central* χ^2-distribution with n degrees of freedom and with *non-central parameter* θ, where θ denotes the sum of the squared expectations μ_i:

$$\begin{aligned} & x_i \sim \mathrm{N}(\mu_i, 1) \quad , \quad i = 1, \ldots, n \quad , \quad x_i \;\; iid \\ & \qquad \Longrightarrow \\ & \sum_{i=1}^{n} x_i^2 =: y \sim \chi^2(n, \theta) \quad \text{with} \quad \theta = \sum_{i=1}^{n} \mu_i^2 \end{aligned} \tag{A.88}$$

The square of a single random number $x \sim \mathrm{N}(0, \mu)$ thus has the non-central χ^2-distribution with one degree of freedom:

$$x \sim \mathrm{N}(0, \mu) \Longrightarrow x^2 =: y \sim \chi^2(1, \mu^2)$$

To determine the moment generating function of the non-central χ^2-distribution we again start from the general Equation A.79:

$$\begin{aligned} G_{\chi^2(1,\mu^2)}(s) &= \mathrm{E}[e^{sx^2}]_{\mathrm{N}(\mu,1)} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx^2} e^{-(x-\mu)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ sx^2 - \frac{1}{2}(x-\mu)^2 \right\} dx \end{aligned}$$

The decisive step to calculate this integral is to complete the square in the argument of the exp-function:

$$\begin{aligned} sx^2 - \frac{1}{2}(x-\mu)^2 &= -\frac{1}{2}\left[(1-2s)x^2 - 2\mu x + \mu^2\right] \\ &= -\frac{1}{2}\left[\left(\sqrt{1-2s}\, x - \frac{1}{\sqrt{1-2s}}\mu\right)^2 - \frac{1}{1-2s}\mu^2 + \mu^2\right] \\ &= -\frac{1}{2}\left(\sqrt{1-2s}\, x - \frac{1}{\sqrt{1-2s}}\mu\right)^2 + \frac{s\mu^2}{1-2s} \end{aligned}$$

Thus, the expectation becomes

$$\begin{aligned} & \mathrm{E}[e^{sx^2}]_{\mathrm{N}(\mu,1)} \\ & = \exp\left\{\frac{s\mu^2}{1-2s}\right\} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\left(\sqrt{1-2s}\, x - \frac{1}{\sqrt{1-2s}}\mu\right)^2 \right\} dx \\ & = \frac{1}{\sqrt{1-2s}} \exp\left\{\frac{s\mu^2}{1-2s}\right\} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\left(u - \frac{\mu}{\sqrt{1-2s}}\right)^2 \right\} du}_{1} \end{aligned}$$

where in the last step we have used the substitution A.81 (again with the condition $s < 1/2$). The remaining integral is the probability that a normally distributed random

number with expectation $\mu/\sqrt{1-2s}$ will assume *any* arbitrary value and, according to Equation A.2, equals one. Thus, the MGF of the non-central χ^2-distribution with one degree of freedom is

$$G_{\chi^2(1,\mu^2)}(s) = \frac{1}{\sqrt{1-2s}} \exp\left\{\frac{s\mu^2}{1-2s}\right\} \tag{A.89}$$

Since the random variables in the sum appearing in the Definition A.88 are all *independent* we can use Equation A.30 to immediately obtain the MGF for a non-central χ^2-distribution with n degrees of freedom:

$$G_{\chi^2(n,\theta)}(s) = \frac{1}{(1-2s)^{n/2}} \exp\left\{\frac{s\,\theta}{1-2s}\right\} \quad \text{with} \quad \theta = \sum_{j=1}^{n} \mu_j^2 \tag{A.90}$$

The characteristic function, Equation A.32, of the non-central χ^2-distribution follows again by replacing s by is in Equation A.90.

$$\Phi_{\chi^2(n,\theta)}(s) = \frac{1}{(1-2is)^{n/2}} \exp\left\{\frac{is\,\theta}{1-2is}\right\} \quad \text{mit} \quad \theta = \sum_{j=1}^{n} \mu_j^2 \tag{A.91}$$

A.5 Transformations Between Distributions

It is possible to transform random variables with a certain distribution into random variables which have *another* distribution. This is particulary useful, for instance, when simulating random walks with the Monte Carlo method. Most *random number generators* generate uniformly distributed random numbers. These can be transformed into random numbers distributed according to a more suitable (for example, normal) distribution function. Two examples of transformations into normally distributed random variables and one transformation into random variables governed by any desired distribution will be introduced below.

A.5.1 Summations

By directly applying the central limit theorem, a sufficient number of independent, identically distributed, random variables are added to obtain an (approximately) normally distributed random variable:

$$z_i \text{iid, uniformly distributed between 0 and 1} \Rightarrow \sqrt{\frac{12}{n}} \sum_{i=1}^{n} (z_i - \frac{1}{2}) =: x \sim \mathrm{N}(0,1) \tag{A.92}$$

According to Equation A.37, uniformly distributed random variables z_i on [0,1] have an expectation of 1/2 and a variance of 1/12. The random numbers $(Z_i - 1/2)$ are uniformly distributed between -1/2 and +1/2, thus having zero expectation. However, the variance of these variables remains 1/12. According to Equation A.18 the variance of the sum of these variables is

$$\mathrm{var}\left[\sum_{i=1}^{n}(z_i - \frac{1}{2})\right] = \sum_{i=1}^{n} \mathrm{var}[(z_i - \frac{1}{2})] = \sum_{i=1}^{n} \frac{1}{12} = \frac{n}{12}$$

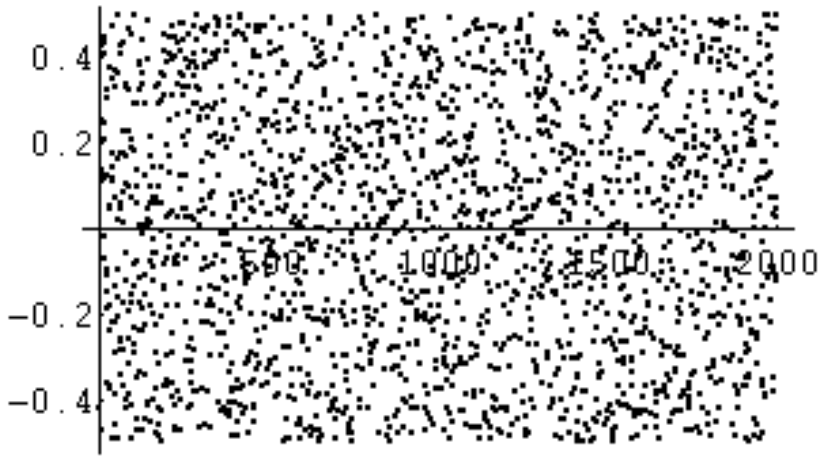


Figure A.1: 2000 *iid* random numbers uniformly distributed between -1/2 and +1/2.

Thus, this sum has zero expectation and a variance of $n/12$, or equivalently a standard deviation of $\sqrt{n/12}$. Dividing the sum by the factor $\sqrt{12/n}$ compensates for this standard deviation. The resulting random variable x in Equation A.92 has a variance of 1 and is thus in consequence of the central limit theorem (approximately) a *standard* normally distributed random variable.

n uniformly distributed *iid* random numbers are required in order to generate a single, approximately normally distributed random variable. For most applications, it is sufficient to take $n = 12$. Figures A.1 and A.2 show the effect of the transformation for $n = 12$.

Note that the transformed random variables can only take on values between $-\sqrt{3n}$ and $+\sqrt{3n}$. Events lying more than $\sqrt{3n}$ standard deviations from the expectation will never occur when using this transformation. The probability of such an event is so small, however, that it does not play a role in most practical applications in finance. For example, for $n = 12$, the probability is approximately 1 to half a billion:

$$1 - \frac{1}{\sqrt{2\pi}} \int_{-6}^{6} e^{-x^2/2} dx \approx 1,973210^{-9}$$

A.5.2 Box-Muller Transformations

The transformation in Equation A.92 requires a significant amount of effort, since n random numbers must be generated in order to obtain one single normally distributed random number. The transformation found by *Box* and *Muller* is much more effec-

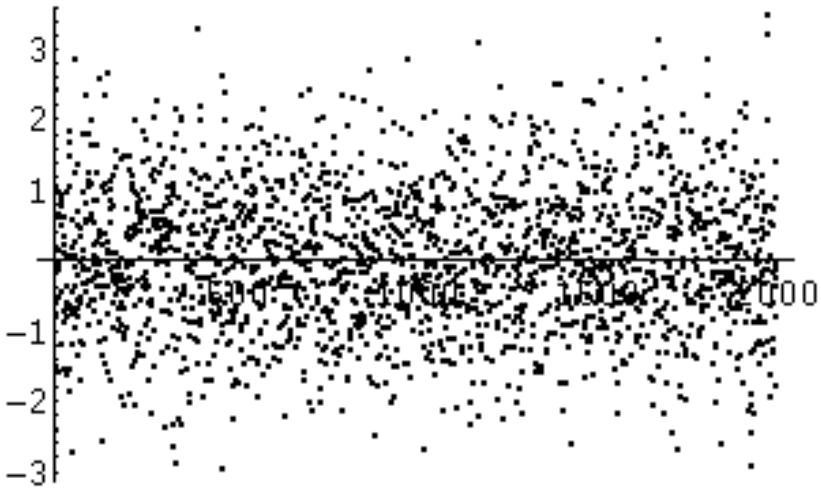


Figure A.2: The random numbers from Figure A.1 after Transformation A.92 with $n = 12$. The random numbers are now approximately N(0,1)-distributed.

tive. From two independent uniformly distributed random numbers, two independent, normally distributed random variables are generated as follows:

$$\begin{aligned} x_1 = \sqrt{-\ln(z_1)}\cos(2\pi\, z_2)\ ,\ \ x_2 = \sqrt{-\ln(z_1)}\ \sin(2\pi\, z_2) \\ \text{with uniformly distributed } z_1, z_2 \text{ between 0 and 1} \Rightarrow \\ x_1 \sim \mathrm{N}(0,1)\ ,\ \ x_2 \sim \mathrm{N}(0,1)\ ,\ \ \mathrm{cov}\,[x_1, x_2] = 0 \end{aligned} \tag{A.93}$$

A.5.3 Inversion of Cumulative Distribution Functions

A very simple method for generating random numbers obeying any desired distribution from a set of random numbers uniformly distributed on the interval between 0 and 1is to evaluate the inverse cumulative probability function of the desired distribution with the uniformly distributed random numbers as its arguments. Or more precisely: let z be a *uniformly* distributed random variables taking on values in the interval between 0 and 1. Let $f(z)$ be an arbitrary distribution density function with $f(z) > 0\ \forall z$. Then

$$x := F^{-1}(z) \quad \text{with} \quad F(z) = \int_{-\infty}^{z} f(u)\, du \tag{A.94}$$

is distributed according to the distribution associated with the density f. Intuitively, since the uniformly distributed random number z lies between 0 and 1, it can be interpreted as a "probability". The inverse of the cumulative distribution function is then nothing other than the *percentile* of this "probability" z with respect to the desired distribution. The random, uniformly distributed numbers are thus interpreted

as probabilities. The percentiles associated with these "probabilities" with respect to the *desired* distribution are then random variables distributed according to the *desired* distribution.

This surprisingly simple method for generating random numbers distributed according to an arbitrary distribution is so important in practice that we will provide the reader with a proof. We begin by proving the existence of an inverse of the cumulative distribution by showing that the cumulative distribution is strictly increasing if the associated density function is strictly positive:

$$\begin{aligned} F(x+dx) \equiv \int_{-\infty}^{x+dx} f(u)\,du &= \int_{-\infty}^{x} f(u)\,du + \int_{x}^{x+dx} f(u)\,du \\ &= F(x) + f(x)dx > F(x) \quad \text{since} \quad f(x) > 0 \end{aligned}$$

Generally $f(x)dx$ is the probability that a random number governed by a distribution density f will take on a value between x and $x + dx$. This is just the definition of the distribution density, i.e.

$$\begin{aligned} f(x)dx &= F(x+dx) - F(x) \\ &= z + dz - z = dz \end{aligned}$$

where in the second step, we have made use of the fact that Equation A.94 implies $F(x) = z$ and, because F is monotone, $F(x + dx) = z + dz$ holds as well. In summary, we may write

$$z = F(x) \Longleftrightarrow dz = f(x)dx \tag{A.95}$$

A random variable, uniformly distributed on the interval $[a, b]$, has a constant density function $p(z) = 1/(b - a)$ for all z in the interval $[a, b]$ as was shown in Equation A.36. In the interval $[0, 1]$ the density of the uniform distribution is thus $p(z) = 1$. The probability that such a uniformly distributed random variable will lie between z and $z+dz$ is therefore simply $p(z)dz = dz$. It follows that Equation A.95 can be interpreted as follows: if the random variable x is governed by a distribution with density function $f(x)$ then $z = F(x)$ is *uniformly* distributed with density function 1, i.e. uniformly distributed on an interval of length 1. Since equality holds in each step of the above derivation, the conclusion holds in both directions: if z is uniformly distributed on an interval of length 1 (and thus has a distribution density of 1) then $x = F^{-1}(z)$ is distributed according to the density $f(x)$. qed.

One application of Equation A.94 is the generation of standard normally distributed random numbers:

$$z \text{ uniformly distributed between 0 and 1} \Rightarrow \mathrm{N}^{-1}(z) =: x \sim \mathrm{N}(0,1) \tag{A.96}$$

The cumulative distribution function of the normal distribution has the disadvantage that it and its inverse function cannot be computed analytically, although they can be computed with applications commonly offered in many widely available software packages. For demonstration purposes, however, it is instructive to apply this method to a distribution which *can* be treated analytically. Consider therefore what is known

as the *Cauchy distribution.* The Cauchy distribution has a parameter $\lambda > 0$ and the density function

$$f_\lambda(x) = \frac{\lambda}{\pi\,(\lambda^2 + x^2)} \tag{A.97}$$

The cumulative distribution function of a Cauchy distributed random variable can be computed analytically with little difficulty. The necessary integral can be found in any collection of mathematical formulas (see for instance [21]):

$$\begin{aligned} F_\lambda(x) &\equiv \int_{-\infty}^{x} f_\lambda(u)\,du = \frac{\lambda}{\pi}\int_{-\infty}^{x} \frac{1}{(\lambda^2 + u^2)}\,du \\ &= \frac{\lambda}{\pi}\left[\frac{1}{\lambda}\arctan(\frac{u}{\lambda})\right]_{u=-\infty}^{u=x} = \frac{1}{\pi}\arctan(\frac{x}{\lambda}) + \frac{1}{2} \end{aligned}$$

where several well-known properties of the *arctangent function* have been employed. We thus have an analytic expression of the cumulative Cauchy distribution. The inverse function can now be obtained by solving the equation $z = F_\lambda(x)$ for x:

$$\begin{aligned} F_\lambda(x) &= z \\ \frac{1}{\pi}\arctan(\frac{x}{\lambda}) + \frac{1}{2} &= z \\ \arctan(\frac{x}{\lambda}) &= \pi(z - \frac{1}{2}) \\ \frac{x}{\lambda} &= \tan\left[\pi(z - \frac{1}{2})\right] \\ x &= \lambda\tan\left[\pi z - \frac{\pi}{2}\right] \\ &= -\lambda\cot(\pi Z) \end{aligned}$$

Thus, the inverse of the cumulative distribution function of the Cauchy distribution is

$$F_\lambda^{-1}(z) = x = -\lambda\cot(\pi z)$$

This implies that if z is a uniformly distributed random variable on $[0, 1]$, then $x = -\lambda\cot(\pi z)$ is a Cauchy distributed random variable taking on values between $-\infty$ and ∞ and having a distribution density function as given in Equation A.97.

Bibliography

[1] Abramowitz M., Stegun I., *Handbook of Mathematical Functions*, Dover Publications, New York, 1972

[2] Alexander, C. (Ed.), *The Handbook of Risk Management and Analysis*, John Wiley & Sons, Chichester, 1996

[3] Anderson, N., Breedon, F., Deacon, M. et al., *Estimating and Interpreting the Yield Curve*, John Wiley & Sons, 1996

[4] Babbel, D. F., Merrill, C. B., *Valuation of Interest-Sensitive Instruments*, Society of Actuaries, Schaumburg, Illinois, 1996

[5] Baschnagel, J., Paul, W., *Stochastic Processes – From Physics to Finance*, Springer Verlag, Heidelberg

[6] Basler Ausschuß für Bankenaufsicht, *Richtlinien für das Risikomanagement im Derivativgeschäft*, Juli 1994

[7] Baxter, M., Rennie, A., *Financial Calculus*, Cambridge University Press, 1997

[8] Beau C. L., Lucas D., *Technical Traders Guide to Computer Analyses of the Futures Market*, Hoppenstedt 1992

[9] Becker, U., *Lexikon Terminhandel*, Gabler Verlag, Wiesbaden, 1994

[10] Best, P., *Implementing Value at Risk*, John Wiley & Sons, 1998

[11] Binder, K., *Monte Carlo Simulations in Statistical Physics*, John Wiley & Sons, New York 1986

[12] Bittner, J., Goddard, R., *Successful Bank Asset-Liability Management*, John Wiley & Sons, 1992

[13] Black F., Karasinski P., *Bond and Option Pricing when Short Rates are Lognormal*, Financial Analysts Journal, July/August 1991, 52-59

[14] Boyle, Phelin et al., *Financial Economics*, The Actuarial Foundation

[15] Brace A., Gatarek D., Musiela M., *The Market Model of Interest Rate Dynamics*, UNSW Preprint, Department of Statistics S95-2

[16] Brigham, E.O., *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974

[17] Broadie, M., Glasserman, P. (Eds.), *Hedging with Trees*, Risk Publications, London, 1998

[18] Brockhaus, O., Farkas, F. et al., *Equity Derivatives and Market Risk Models*, Risk Books, London, 2000

[19] Brockhaus, Oliver, Ferraris, A., Gallus, C. et al., *Modelling and Hedging Equity Derivatives*, Risk Books, London, 1999

[20] Brockwell, P. J., Davis, R. A., *Introduction to Time Series and Forecasting*, Springer Verlag, Heidelberg, 1998

[21] Bronstein, S., *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt 1979

[22] Bundesaufsichtsamt für das Kreditwesen, *Neufassung des Grundsatzes I, Erläuterungen zum Vorentwurf*, Berlin, 10.5.96, Geschäftsnummer 17-A223-2/93

[23] Bundesaufsichtsamt für das Kreditwesen, *Verlautbarung über Mindestanforderungen an das Betreiben von Handelsgeschäften der Kreditinstitute*, Berlin, 23.10.95, Geschäftsnummer I4-42-3/86

[24] Byron, J.T.M., *Elements of Simulation*, Chapman & Hall, 1984

[25] Calamos, J. P., *Convertible Securities*, McGraw-Hill, New York, 1998

[26] Campbell, J. Y., Lo, A. W., Mackinlay A. G., *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey, 1997

[27] Chiang, A. C., *Fundamental Methods of Mathematical Economics*, 2.Edition, McGraw-Hill

[28] Chriss, N. A., *Black Scholes and Beyond*, McGraw-Hill, New York, 1997

[29] Clarke R., *Estimating and Using Volatility: Part 1*, Derivatives Quarterly, Fall 1994, S. 40-44

[30] Clewlow, Les, Strickland, Chris, *Implementing Derivatives Models*, John Wiley & Sons

[31] Collinson, F., Giddings, M., Sullivan, M., *Financial Products: A Survival Guide*, Euromoney Books, 1996

[32] Connolly, K. B., *Pricing Convertible Bonds*, John Wiley & Sons

[33] Cornish E.A., Fisher R.A., *Moments and Cumulants in the Specification of Distributions*, Review of the International Statistical Institute, 5, 307-320

[34] Cox, J.C., Rubinstein, M., *Options Markets*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985

[35] Cox J. C., Ingersoll J. E., Ross S. A., *A Theory of the Term Structure of Interest Rates*, Econometrica, 53, 1985, 385-407

[36] Culp, L., Miller, M.H. (Eds.), *Corporate Hedging in Theory and Practice*, Risk Publications, London, 1999

[37] Das, S., *Credit Derivatives*, John Wiley & Sons, New York, 1998

[38] Das, S., *Swap and Derivative Financing*, revised ed., McGraw-Hill, New York, 1994

[39] Dattatreya R. E., Fabozzi F. J., *Active Total Return Management of Fixed Income Portfolios*, Probus Publishing Company, 1989

[40] Davenport, W.B., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Books Co., New York, 1958, Chap.6

[41] Dixit, A. K., Pindick, R. S., *Investment under Uncertainty*, Princeton University Press, Princeton, New Jersey, 1994

[42] Doi, M., Edwards, S.F., *The Theory of Polymer Dynamics, Clarendon Press*, Oxford University Press, 1986

[43] Dothan, M, *Prices in Financial Markets*, Oxford University Press, New York 1990

[44] Dowd, K., *Beyond Value-At-Risk*, John Wiley & Sons

[45] Duffie D., *Security Markets – Stochastic Models*, Academic Press, San Diego, 1988

[46] Duffie, D., *Dynamic Asset Pricing Theory*, 2. Edition, Princeton University Press, Princeton, New Jersey, 1996

[47] Dunis, C., Zhon, B., *Nonlinear Modelling of High Frequency Financial Time Series*, John Wiley & Sons

[48] Dupire, B., *Monte Carlo – Methodologies and Applications*, Risk Books, London, 1998

[49] Eller R. (Ed.), *Handbuch Derivativer Instrumente*, Schäffer-Poeschel Verlag, Stuttgart, 1996

[50] Eller R. (Ed.), *Handbuch des Risikomanagements*, Schäffer-Poeschel Verlag, Stuttgart, 1997

[51] Eller, R., Gruber, W., Reif, M. (Eds.), *Handbuch Bankenaufsicht und Interne Risikosteuerungsmodelle*, Schäffer-Poeschel Verlag, Stuttgart, 1998

[52] Elton, E., Gruber, M., *Modern Portfolio Theory and Investment Analysis*, 5. Edition, John Wiley & Sons, New York, 1995

[53] Embrechts, P., Klüppelberg, C., Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Heidelberg, 1999

[54] Fabozzi, F. J., *The Handbook of Fixed Income Securities*, 5. Edition, McGraw-Hill

[55] Fabozzi, F. J., *Valuation of Fixed Income Securities and Derivatives*, 3. Edition, Fabozzi Associates, 1998

[56] Feller, *Probability Theory and its Applications*, vols. 1 and 2, John Wiley & Sons, New York, 1950

[57] Ferris, K., Jones, M., *The Reuters Guide to Official Interest Rates*, 2. Edition, Probus Pub., 1995

[58] Fishman, G.S., *Monte Carlo Concepts, Algorithms and Applications*, Springer Verlag, 1996

[59] Flury, B., *A First Course in Multivariate Statistics*, Springer Verlag, Heidelberg

[60] Francis, J., Clark, F., Joyce A., Whittaker, J. Gregg, *The Handbook of Credit Derivatives*, McGraw-Hill, 1998

[61] Gardiner, C. W., *Handbook of Stochastic Methods*, 2. Edition, Springer Verlag, Heidelberg, 1985

[62] Gardner, M. J., Mills, D. L., *Managing Financial Institutions*, 3. Edition, Dryden, Fort Worth, 1994

[63] Gershenfeld, N., *The Nature of Mathematical Modelling*, Cambridge University Press

[64] Gibson, R. (Ed.), *Model Risk – Concepts, Calibration and Pricing*, Risk Publications, London 1999

[65] Greene, W.H., *Econometric Analysis*, 3. Edition, Prentice-Hall, London, 1997

[66] Grill, Gramlich, Eller, *Gabler Bank Lexikon*, 11. Edition, Gabler Verlag, Wiesbaden, 1995

[67] Grimmett, G. R., Stirzaker, D. R., *Probability and Random Processes*, 2. Edition, Oxford University Press, Oxford, 1998

[68] Grinold, R., Kahn, R., *Active Portfolio Management*, Probus, 1995

[69] Group of Thirty, *Derivatives, Practices and Principles*, Washington D.C., 1993

[70] Gruber, W., Overbeck, L., Working Paper, Deutsche Bundesbank, 1996

[71] Gwilym, O., Sutcliffe, C. (Eds.), *High-Frequency Financial Market Data*, Risk Publications, London, 1999

[72] Hamilton, J. D., *Time Series Analysis*, Princeton University Press, Princeton, New Jersey, 1994

[73] Hammer D., *Dynamic Asset Allocation*, John Wiley & Sons

[74] Hanenberg, L., *Zur Verlautbarung über Mindestanforderungen an das Betreiben von Handelsgeschäften der Kreditinstitute des Bundesaufsichtsamtes für das Kreditwesen*, Die Wirtschaftsprüfung, S. 637-648, Heft 18/1996, Jahrgang 49

[75] Harrison, M., Pliska, S., *Martingales and Stochastic Integrals In the Theory of Continuous Trading, Stochastic Processes and their Applications*, 11 (1981), 215-260

[76] Haug, Gaarder, *The Comlete Guide to Option Pricing Formulas*, McGraw Hill, New York, 1997

[77] Haugen, R. A., *Modern Investment Theory*, 4. Edition, Phipe Prentice Hall, Upper Saddle River, New Jersey, 1997

[78] Heath D., Jarrow R., Morton A., *Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation*, Econometrica, 60, 1992, 77-105

[79] Heidorn, T., *Vom Zins zur Option – Finanzmathematik in der Bankpraxis*, Gabler Verlag, Wiesbaden, 1994

[80] Hess, D. E., *Die Dynamik der Zinsstruktur*, Gabler Edition Wissenschaft, Wiesbaden, 1995

[81] Hicks, A., *Foreign Exchange Options*, 2. Edition, Woodhead Publishing Ltd.

[82] Higgins, R. C., *Analysis for Financial Management*, 5. Edition, Irwin/ McGraw-Hill, 1998

[83] Ho S. Y., Lee S.-B., *Term Structure Movements and Pricing Interest Rate Contingent Claims*, Journal of Finance, 41, 1986, 1011-1029

[84] Ho, Thomas, *Fixed Income Solutions*, Irwin, Chicago, 1996

[85] Hughston, L. (Ed.), *Vasicek and Beyond: Approaches to Building and Applying IR-Models*, Risk Publications/Euro Brokers, 1996

[86] Hughston, L. (Ed.), *Options*, Risk Publications, London, 1999

[87] Hughston, L. (Ed.), *The New Interest Rate Models: Recent Developments in the Theory and Application of Yield Curve Dynamics*, Risk Books, 2000

[88] Hull, J. C., *Options, Futures and other Derivatives*, 4. Edition, Prentice Hall, Upper Saddle River, New Jersey, 2000

[89] Hull, J. C., White A., *Hull-White on Derivatives*, Risk Publications, London, 1996

[90] Hull, J. C., White A., *Pricing Interest Rate Derivative Securities*, The Review of Financial Studies, 3, 1990, 573-592

[91] Hunt, P. J., Kennedy, J. E., *Financial Derivatives in Theory and Practice*, John Wiley & Sons 2000

[92] Ingersoll Jr., Jonathan E., *Theory of Financial Decision Making*, Rowman & Littlefield, Bollman Place, Savage, Maryland, 1987

[93] James, J., Webber, N., *Interest Rate Modelling*, John Wiley & Sons, 2000

[94] Jamshidian, F., Mathematical Finance 3, 2 (1993), p 149-159

[95] Jarrow, R., Turnball S., *Derivative Securities*, Southwestern Publishers, Cincinnati 1996

[96] Jarrow, R., *Modelling Fixed Income Securities and Interest Rate Options*, McGraw Hill, 1996

[97] Jarrow, R. (Ed.), *Over the Rainbow*, Risk Publications, London, 1995

[98] Jarrow, R. (Ed.), *Volatility – New Estimation Techniques for Pricing Derivatives*, Risk Publications, London, 1998

[99] Johnson, N.L., Kotz, S., Balakrishan N., *Continuous Univariate Distributions*, Vol. 1, John Wiley & Sons, 1994

[100] Johnson, R. A., Wichern, D. W., *Applied Multivariate Statistical Analysis*, 4. Edition, Prentice Hall, Upper Saddle River, New Jersey, 1999

[101] Jollife, I. T., *Principal Component Analysis*, Springer Verlag, Heidelberg, 1986

[102] Jordan, D. W., Smith, P., *Mathematical Techniques*, 2. Edition, Oxford University Press, Oxford, 1998

[103] Jorion, P., *Value at Risk: The New Benchmark for Controlling Derivatives Risk*, McGraw-Hill, 1997

[104] Kalos, M.H., Whitlock, P.A., *Monte Carlo Methods*, John Wiley & Sons, Inc., New York 1986

[105] Karatzas, I., Shreve, S. E., *Brownian Motion and Stochastic Calculus*, 2. Edition, Springer Verlag, Heidelberg, 1999

[106] Karlin, H.M., *A First Course in Stochastic Processes*, Academic Press, New York 1975

[107] Klein, R.A., Lederman, J., *Derivatives Risk and Responsibility*, Irwin, 1996

[108] Kloeden, P.E., Platen, E., Schurz, H., *Numerical Solutions of SDE Through Computer Experiments*, Springer Verlag, Heidelberg, 1994

[109] Kofler, M., *VBA-Programmierung mit Excel 97*, Addison-Wesley, 1997

[110] Korajzykk, R.A. (Ed.), *Asset Pricing and Portfolio Performance*, Risk Publications, London 1999

[111] Krzanowski, W.J., *Principles of Multivariate Analysis*, Oxford Science Publications, Oxford, reprinted 1998

[112] Kwok, Y.K., *Mathematical Models of Financial Derivatives*, Springer Verlag, Heidelberg, 1998

[113] Lamberton D., Lapeyere B., *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, 1996

[114] Leibowitz M., Kogelman S., Klaffky T., *A Shortfall Approach to Duration Management*, Salomon Brothers 1990

[115] Lewis, A. L. , *Option Valuation under Stochastic Volatility*, Financepress, 2000

[116] Lewis, E., *An Introduction to Credit Scoring*, Reprint Edition, Fair, Isaac & Co., 1994

[117] Loistl, O., *Kapitalmarkttheorie*, Oldenbourg-Verlag, München, 1994

[118] Longerstaey J., *Introduction to RiskMetrics*, Morgan Guaranty Trust Company, 1998

[119] Mantegna, R. N., Stanley, H. E., *An Introduction to Econophysics*, Cambridge University Press

[120] Markowitz, H. M., *Mean Variance*, Blackwell, 1987

[121] Markowitz, H. M., *Portfolio Selection*, 2. Edition, Blackwell, 1998

[122] Matten, C., *Managing Bank Capital*, John Wiley & Sons, 1996

[123] Merton, R. C., *Continuous Time Finance*, revised Edition, Blackwell, Malden MA, 1998

[124] Mills, T. C., *The Econometric Modelling of Financial Time Series*, Cambridge University Press, Cambridge, 1997

[125] Miron, P.; Swannel, P., *Pricing and Hedging of Swaps*, Euromoney Books, London 1991

[126] Musiela, M., Rutkowski, M., *Martingale Methods in Financial Modelling*, Springer Verlag, Heidelberg, 1997

[127] Neftci, Salih N., *An Introduction to the Mathematics of Financial Derivatives*, 2. Edition, Academic Press, San Diego, 2000

[128] Nelken, Israel, *The Handbook of Exotic Options*, Probus Publications, 1995

[129] Niederreiter, H., *Random Number Generation and Quasi Monte Carlo Methods*, Siam, Philadelphia, 1992

[130] Nielsen, L.T., *Pricing and Hedging of Derivative Securities*, Oxford University Press, 1999

[131] Oksendal, B., *Stochastic Differntial Equations*, 5. Edition, Springer Verlag, Heidelberg, 1998

[132] Pindyck, R. S., Rubinfeld, D. L., *Econometric Models and Economic Forecasts*, 4. Edition, McGraw Hill, 1997

[133] Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C – The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1994

[134] Prudenzi, P. S., *VBA mit Excel 97*, Addison-Wesley, 1997

[135] Questa, G.S., *Fixed Income Analysis for the Global Financial Market*, John Wiley & Sons, 1999

[136] Rebonato, R., *Interest-Rate Option Models*, John Wiley & Sons, Chichester, 1996

[137] Rebonato, R., *Volatility & Correlation*, John Wiley & Sons, Chichester, 1999

[138] Rice, J. A., *Mathematical Statistics and Data Analysis*, 2. Edition, Duxbury Press

[139] Ripley, B. D., *Stochastic Simulation*, John Wiley & Sons, New York, 1987

[140] RiskMetrics, *Technical Document*, 4. Edition JP Morgan Global Research, New York, 1996

[141] Rönz B., Förster E., *Regressions- und Korrelationsanalyse, Grundlagen, Methoden, Beispiele*, Gabler Verlag, Wiesbaden, 1992

[142] Rouvinez, C., *Going Greek with VaR*, RISK Magazine, Vol.10 No.2, Februar 1994

[143] Rubinstein, M., *Rubinstein on Derivatives*, Risk Publications, London, 1999

[144] Rubinstein, M., *Return to Oz*, RISK Magazine, Vol.10 No.11, November 1994

[145] Rubinstein, R.Y., *Simulation and the Monte Carlo Method*, John Wiley & Sons, 1981

[146] Schierenbeck H., *Bank- und Versicherungslexikon*, Oldenbourg Verlag 1990

[147] Schierenbeck, H., *Ertragsorientiertes Bankmanagement*, 5. Edition, Gabler Verlag, Wiesbaden, 1997

[148] Sharpe, W. F., Alexander, G.J., Bailey, J. V., *Investments*, 5. Edition, Prentice Hall, 1995

[149] Smith C. W., Smithson C. W., *The Handbook of Financial Engineering, New Financial Product Innovations, Applications, and Analyses*, Harper Business Books 1990

[150] Smith, C. D., *Option Strategies*, John Wiley & Sons, New York, 1996

[151] Smith, G. D., *Numerical Solution of Partial Differential Equations*, 3. Edition, Oxford University Press, 1998

[152] Taleb, Nassim, *Dynamic Hedging*, John Wiley & Sons

[153] Tavakoli, J. M., *Credit Derivatives*, John Wiley & Sons

[154] Thomas, J.W., *Numerical Partial Differential Equations*, Springer Verlag, Heidelberg, 1995

[155] Tavella, D., Randall, C., *Pricing Financial Instruments: The Finite Difference Method*, John Wiley & Sons, 2000

[156] Tiner, J. I., Conneely, J. M., *Accounting for Treasury Products*, 2. Edition, Woodhead-Faulkner, 1989

[157] Tomkins, R., *Options Explained*, MacMillan Business, London 1994

[158] Tong, Howell, *Non-linear Time Series*, Oxford Science Publications, Oxford, reprinted 1999

[159] Uyemura, D. G., van Deventer, D. R., *Financial Risk Management in Banking*, McGraw Hill, New York, 1993

[160] Vasicek O. A., *An Equilibrium Characterization of the Term Structure*, Journal of Finance, 5, 1977, 177-188

[161] Walmsley J., *The New Financial Instruments*, John Wiley & Sons, 1988

[162] Welcker J., Kloy, J.W., Schindler, K., *Professionelles Optionsgeschäft*, MI Verlag, Zürich 1992.

[163] Williams, D., *Probability with Martingales*, Cambridge Mathematical Textbooks, Cambridge, 1998

[164] Wilmott, P. J., *Derivatives – The Theory and Practice of Financial Engineering*, John Wiley & Sons, 1998

[165] Wilmott, P. J., Dewynne, J., Howison, S., *The Mathematics of Financial Derivatives*, Cambridge University Press, Cambridge, 1998

[166] Wilmott P.J., Dewynne J., Howison S., *Option Pricing – Mathematical Models and Computation*, Oxford Financial Press, Oxford, 1997

[167] Wilson, T.C., in C.Alexander, *Risk Management and Analysis*, John Wiley & Sons, 1998

[168] Zangari P., *A VaR Methodology for Portfolios that Include Options*, RiskMetrics Monitor, 1st quarter, 1996

[169] Zhang, P. G., *Exotic Options: a Guide to Second Generation Options*, World Scientific, Singapore, 1997

[170] Zimmermann, H., Jaeger, S., Staub, Z., *Asset- und Liability-Management*, 2. Edition, Verlag neue Zürcher Zeitung, 1996

Index

30/360, 12, 51
30E/360, 12

Accrued Interest, 71, 309
Act/360, 51
Act/365, 51, 319
Act/Act, 12, 51
Actual/365f, 12
Add On, 457
ADP, 253
Aging, 452
Akaike Information Criterion, 558
Annualization, 51
Annualized, 51
Annuity Loan, 297, 429
Approximation, 75
AR(1), 511
AR(p), 508
Arbitrage, 21, 81
Arbitrage Free
    Market, 206
    Short Rate Model, 246
Arctangent, 597
Arrow-Debreu Price, 253
    Instantaneous, 254
Asian Option, 353
    American, 359
    Average Price, 341
    Average Strike, 341
Asset Swap, 366
Asset-or-Nothing Option, 345
At The Money, 58
    Forward, 476
    Point, 476
Atlantic Option, 339, 342
Autocorrelation, 497, 555
    Time, 503
Autocovariance, 497
    Estimator for, 503
Autoregressive Process, 508
    First Order, 511
    p-th Order, 508
Average
    Moving, 550
    Symmetric, 550

Backtesting, 9, 451
Backward Equation, 39
Backward Induction, 246
Bank Account, 91, 232
Bank Holiday, 14
Barrier Option, 126, 346
    Down-And-In, 346
    Down-And-Out, 346
    Monte Carlo Valuation, 356
    Up-And-In, 346
    Up-And-Out, 346
Base Currency, 487, 491
Base Point Value, 77
Basel
    Accord, 5
    Market Risk Paper, 5
Basis, 85, 325
    Risk, 85
    Trading, 85
Basle Committee
    on Banking Supervision, 8
BDC, 14
Bermuda Option, 339, 342
Bernoulli Experiment, 454, 576
Beta, 195, 570
    of an Option, 196
BGM Model, 230
Binary Knockout Range, 350

Binary Option, 344
Binomial
  Coefficient, 151, 576
  Distribution, 151, 153, 321, 576
  Formula, 577
Binomial Model
  for American Options, 321
  for European Options, 321
Binomial Tree, 147, 205, 245
  Recombining, 150
Black-76, 112
  for Caps, Floors, 329
  for Swaptions, 334
  Model, 110, 111, 326
Black-Derman-Toy, 258, 275
Black-Karasinski, 275
Black-Scholes
  Differential Equation, 91, 183
    for Forward Contracts, 183
    for Futures Contracts, 184
  für Digitals, 345
  for American Options, 93
  Option Pricing Formula, 107, 108, 161
Black-Scholes World, 44, 103
BOBL, 54
Bobl Future, 309
Boltzmann Constant, 522
Boltzmann Factor, 521
Bond, 287
  Callable, 58
    Step Up, 366
  Putable, 58
  Structured, 366
  Synthetic, 309
Bond Future, 308
Bond Option
  American, 326
  European, 326
Book Value, 373
Bootstrapping, 462, 463, 469
  Classical, 467
Boundary Condition, 92
  Dirichlet, 123
    for Plain Vanilla Options, 127
  Free, 94, 135
  Neumann, 127
    for Plain Vanilla Options, 131
Box-Cox Scaling, 507, 552
Box-Muller Transformation, 594
Brace-Gatarek-Musiela Model, 230
Brownian Motion, 31, 212
  Geometric, 376
Bull Floating Rate Note, 50
Bund Future, 309
Bundesanleihe, 54, 310
Bundesobligation, 54, 310
Bundesschatzanweisung, 54, 310
Bundesschatzbrief
  Typ A, 54
  Typ B, 52
Bushy Tree, 358, 360
Business Day Convention, 14, 469
Buy and Hold Strategy, 306, 365

CAD, 5
  Standard Method, 428
CAD II, 8, 9, 377, 452, 457
Calibration, 269
Call, 58, 339
Callable Bond, 58
Cap, 59, 237, 253, 327
Capital Adequacy Directive, 5, 8, 452, 457
  Standard Method, 428
Capital Asset Pricing Model, 195
Capital Growth Bond, 52
Capital Market, 48
  Paper, 52
Caplet, 237, 327, 438
CAPM, 195
Caption, 343
Cash and Carry Arbitrage, 81
Cash Flow, 15, 47, 363, 427
Cash Flow Mapping, 441
  Duration-Based, 443
  Risk-Based, 442
Cash-or-Nothing Option, 344
CATS, 364
Cauchy Distribution, 597
  Cumulative, 597
    Inverse, 597

CBOT, 56
CD, 49, 52
CD-Equivalent Yield, 52
Central Limit Theorem, 503, 578, 593
Certificate of Deposit, 49, 52
Characteristic Function, 415, 573
    of the Binomial Distribution, 578
    of the Chi-Squared Distribution, 591
        Non-Central, 593
    of the Gamma Distribution, 587
    of the Normal Distribution, 583
    of the Uniform Distribution, 576
Characteristic Polynomial, 509
Cheapest To Deliver, 311
Chi-Squared Distribution, 406, 588
    Degrees of Freedom, 588
    Non-Central, 407, 592
Chicago Board of Trade, 56
Cholesky Decomposition, 177, 389
Chooser Option, 342
Clean Price, 72
Clearing House, 57
Cliquet Option, 340
Clustering, 513
Coarse Graining, 26
Cointegration, 539
Collar, 59
Commercial Paper, 48
Commodity Future, 56
Complete Market, 218
Compound Option, 342
Compounded Interest, 18
Compounding, 16
    Annual, 11
    Continuous, 11, 19
    Discrete, 11, 17
    Linear, 11, 19
    Simple, 11, 17
Compounding Interest Effect, 51
Compounding Method, 16
Computer Simulation, 167
Confidence, 374
    Interval, 374
        of the Normal Distribution, 581
        One-Sided, 374
        Symmetric, 374
Contract Size, 309
Convenience Yield, 23
Convention, 11, 20, 51
Conversion Factor, 310
Convexity, 76, 77
    of a Bond, 291
    of a Floater, 286
    of a Swap, 296
    of a Zero Bond, 284
    of an Annuity Loan, 305
Convexity Adjustment, 239, 468
Convolution, 417
Cornish-Fisher Expansion, 414
Correlation, 176, 383, 568
    Coefficient, 557, 568
    Historical, 495
Correlation Effect, 448
Cost of Carry, 23
COUGAR, 364
Counterparty Risk, 373
Coupe Option, 340
Coupon, 53, 287
Coupon Bond, 287, 429
Coupon Issue, 364
Coupon Strip, 468
Coupon Stripping, 363, 364
Covariance, 176, 383, 485, 568
Covariance Matrix, 176, 384, 568
Cox-Ingersoll-Ross, 276
CP, 48
cpf, 375, 566
Crank-Nicolson, 116, 121
Crash Scenario, 424
Credit
    Derivative, 297
    Linked Note, 297
    Risk, 306
Credit Worthiness, 306, 461
Cross Rate, 491
CTD, 311
Cumulative Probability Function, 375, 566
Current Level, 385, 484
Curtosis, 414, 570
    of the Normal Distribution, 583
Curvature

of the Term Structure, 547
Cutoff, 504

Daily Range Accrual, 346
DAX, 24, 58, 195, 323
Day Count Convention, 12, 469
DCC, 12
Default
Probability, 306
Risk, 306, 373
Deferred Swap, 311
Delivery Price, 81
Delta, 76, 192
Approximation, 392
Hedge, 91, 193
Neutral, 193
Risk, 428
Delta-Gamma
Approximation, 392
Hedge, 193
Delta-Normal, 445
Method, 395, 398
Deposit
at Notice, 50
Time, 50
Derivative, 55
Synthetic, 181
Derivative Security, 55
Determinant, 401
Deutsche Terminbörse, 56
Diagonalization
of a Matrix, 401
Difference
p-th, 550
Difference Quotient, 116
Difference Scheme
Consisitency, 140
Convergence, 140
Stability, 140
Differential Equation
Parabolic, 98
Second Order, 98
Stochastic, 37
Diffusion Equation, 98
Diffusion Process, 34
Digital Option, 344
American, 344
European, 344
One Touch, 344
DINGO, 364
Dirac Delta Function, 38, 574
Dirty Price, 68, 309
Discount
Factor, 15
Paper, 48
Rate, 52
Discounting, 16
Disk, 169
Distribution, 566, 574
Continuous, 565
Cumulative, 566
Density, 566
Multivariate, 567
Discrete, 565
Multivariate, 567
Diversification Effect, 448
Dividend
Implied, 83
Yield
Implied, 83
Dividend Adjustment, 23
Double Barrier Option, 350
Dow Jones, 195
Drift, 27, 32
Risk Neutral, 98
DTB, 56, 310
Dupire
Local Volatility Surface, 483
Volatility, 479
Duration, 485
Key Rate, 78
of a Bond, 291
of a Floater, 286
of a Swap, 296
of a Zero Bond, 284
of an Annuity Loan, 305
Macaulay, 74, 443
Modified, 76, 77, 487
of a Bond, 291
of a Floater, 286
of a Swap, 296
of a Zero Bond, 284

of an Annuity Loan, 305

Early Exercise, 322
Eigenvalue, 401, 543
Eigenvector, 401, 543
Decomposition, 403
Elasticity, 195
Embedded Option, 342
Embeddos, 58
Equilibrium Model, 244, 276
Error, 496
Propagation, 499
Quadratic, 499
Statistical, 355, 498
Systematic, 355, 498
Type I, 454
Type-II, 454
Estimator, 496
Euler's Number, 19
EUREX, 56, 243, 310, 322
EURIBOR, 49
Euro Bobl Future, 309
Euro Bund Future, 309
Euro Dollar Future, 243
Euro EUR Future, 243
Euro Future, 468
Euro Interest Convention, 51
Euro Schatz Future, 309
Euro Swiss Future, 243
EWMA, 535
Ex-Dividend Day, 326
Exchange Option, 176, 342
Exchange Rate, 324
Forward, 325
Exercise
of American Options, 86
Probability, 162
Expectation, 96, 496, 566, 570
Conditional, 508
of the Binomial Distribution, 578
of the Chi-Squared Distribution, 591
of the Gamma Distribution, 587
of the Normal distribution, 579
of the Uniform Distribution, 575
Unconditional, 508
Explicit Method, 115, 121
Exponential Function, 19
Extendable Option, 342, 356

Factor 3, 456
Factor Analysis, 539
Factor Model, 230
Fairway Bond, 346
Fairway Floater, 346
Federal Bond, 310
Federal Debt Obligation, 310
Federal Treasury Note, 310
Filter, 550
Linear, 551
Financial Future, 56
Financial Instrument, 47
Financial Product, 47
Finanzierungsschätze, 52
Finite Difference, 113
Method, 113
Fixed Deposit, 50
Fixed Leg, 468
Fixing, 285
Floater, 49, 285, 429
Reverse, 50
Floating Leg, 468
Floating Rate Note, 49
Floor, 59, 237, 253, 327, 329
Floorlet, 237, 327, 439
Floortion, 343
Fokker-Planck Equation, 38
Forecast, 505, 525
Linear, 525
Best, 526
Optimal, 525
Forecasting Error, 525
Forecasting Test, 558
Foreign Currency, 324
Option, 324
Forward, 56
Discount Factor, 22
Forward Bond, 431
Forward Contract, 81
Forward Equation, 38, 42
Forward Floater, 312
Forward Induction, 248, 252
Forward Par Rate, 317

Forward Price, 81, 82, 84, 233
Dirty, 432
Forward Rate, 21, 55, 307
Instantaneous, 230
Forward Rate Agreement, 55, 237, 307, 430
Forward Swap, 311, 430
Rate, 313
Forward Transaction
Conditional, 55
Unconditional, 55
Forward Volatility, 111, 240
Fourier Transformation, 415, 573
Fast, 418
Inverse, 417, 573
FRA, 55, 307, 430
Rate, 55
FRN, 49
Full Valuation, 420, 421
Future, 56, 84
on a Coupon Bond, 430
on a Zero Bond, 430
Future Price, 233
Future Strip, 468
Future Styled, 179, 242
Futures Exchange, 310
FX Future, 324
FX Option, 324
FX Risk, 446

G30, 7
Report, 7
Study, 5
Gamma, 76, 192
Matrix, 400
Gamma Distribution, 585
Gamma Function, 586
GARCH Process, 513
GARCH(1,1) Process, 516, 533
GARCH(p,q) Process, 513, 528
Gaussian Distribution, 578
Gaussian Elimination, 144
GBM, 376
Geometric Sequence, 288
Geometric Series, 288, 534
Girsanov Theorem, 213
Greeks, 108, 144, 159, 179, 197
Green Zone, 456
Green's Function, 249
Grid, 113
Non-Uniform, 114
Uniform, 114
Group of Thirty, 7

Heat Equation, 98
Heath-Jarrow-Morton Model, 273
Hedge, 181, 208, 211
Ratio, 182
for arbitrary Hedging Instruments, 189
for Spot Transactions, 182
Hedging, 60
Hedging Effect, 448
Hessian Matrix, 392
Hindsight Option, 340
Historical Simulation, 422
HJM Model, 273
Ho-Lee, 258, 275
Holiday Calender, 469
Hull-White, 275
Hump, 547
Hypothesis, 454
Hypothesis Test, 454
Binomial, 455
Inflection Point, 554
Percentile, 554
QQ plot, 556
Sign, 553

Identity Matrix, 387
iid, 389, 512
Imaginary Number, 415, 573
IMM, 56
Implicit Method, 115, 121
Implied Dividend Yield, 323
Importance Sampling, 272
In Sample, 558
In The Money, 58
Independence
of Random Variables, 389, 584
Independent Identically Distributed, 389
Index, 195

Future, 322
Option, 322
Induction
Backward, 246
Forward, 248, 252
Initial Condition, 92, 122
Inner Product, 402
Integral Form, 104, 105
Integral Measure, 203
Riemann, 203
Interest, 16
Interest Proceeds, 365
Interest Rate, 20
Curve, 461
Instantaneous, 230
Interest Rate Future, 308
Interest Rate Guarantee, 328
Interest Rate Instrument, 47, 283
Interest Rate Option, 326
Interest Rate Risk, 427
Interest Rate Tree, 252
Internal Model, 5, 9, 379
Internal Rate of Return, 69
International Monetary Market, 56
Interpolation, 440, 472, 525
Exponential, 473
Linear, 440, 471
Intrinsic Value, 85
IRG, 328
IRR, 69
Issuer Risk, 373
Ito Process, 34
Ito's Lemma, 34

Jacobian Determinant, 100
Johnson Transformations, 414
Joint Probability Distribution, 583
Jumbo Bond, 54

Key Rate, 78
Knock-In Option, 346
Knock-Out Option, 346
Kolmogorov Equation, 39
Kronecker Delta, 387, 402, 569, 584
Kuiper Statistic, 555
KWG, 5

L-U Decomposition, 125
Ladder Option, 350
Lag, 506
Lagrange
Function, 542
Multiplier, 542
Method of, 542
Lamberton-Lapeyre Procedure, 136, 145
Laplace Transformation, 571
Law of Large Numbers, 496
Leap Year, 12
Least Squares Fit, 477
Leg, 364
Fixed, 54
Floating, 54
of a Swap, 294
Legal Risk, 373
Level of Confidence, 374
LIBOR, 49, 285
LIBOR Flat, 286
LIBOR Strip, 468
LIBOR-In-Arrears
FRA, 241
Swap, 241
LIFFE, 56, 243
Likelihood, 517
Conditional, 518
of an AR(p) Process, 518
Linear Algebra, 401
Liquidation Period, 9, 377
Lock-In Option, 351
Log-Likelihood, 518
Logarithm, 30, 584
Logistic Curve, 549
Lognormal Distribution, 32, 584
Lognormal Model, 258
Look-Back Option, 340, 360

MA, 535
MaH, 5
Margin, 84
Account, 57, 84
Additional, 57
Call, 57
Initial, 57
Risk-Based, 57

Variation, 57, 184, 188
Mark-to-Market, 420
Market Price of Risk, 222
Market Rate Model, 230
Market Risk Factor, 373
Markov
Chain, 522
Process, 27
Property, 508
Martingale, 204
Measure, 203
Probability, 207
Matrix
Algebra, 388
Inverse, 389
Product, 388
Singular, 401
Maturity, 47, 58, 339
Band, 8
Date, 15
Maximum Likelihood, 517, 558
Mean, 496, 578
Mean Return, 32
Historical, 495
Mean Reversion, 275
Mean Square Error, 525
Measure, 203, 211
Forward-Neutral, 234
Risk-Neutral, 232
Medium Term Note, 49
Method
English, 51
French, 51
German, 51
International, 51
MGF, 408, 571
Minimum Requirements
for Trading Activities, 7
MMY, 52
Model
Multi-Factor, 212
One-Factor, 212
Modified Following, 469
Moivre-Laplace, 160
Moment Generating Function, 408, 571
of the Binomial Distribution, 577
of the Chi-Squared Distribution, 589
Non-Central, 592
of the Gamma Distribution, 586
of the Normal Distribution, 582
of the Uniform Distribution, 575
Moment Matching, 157, 414
Moments, 570
Central, 413, 570
of the Chi-Squared Distribution, 591
of the Gamma distribution, 587
of the Normal Distribution, 582
Money Market, 48
Account, 232
Future, 242, 468
Rate, 468
Yield, 52
Moneyness, 270, 476
Monotonous Function, 375
Monte Carlo Simulation, 167, 419
on a Tree, 271
Mortgage Bond, 54
Moving Average, 535
Exponentially Weighted, 535
Simple, 535
MSE, 525
Multi-Factor Option, 341
Municipal Bond, 54

Natural Number, 19
Net Present Value, 69
Newton-Raphson Method, 264, 479
Nominal Rate, 51
Nominal Yield, 51
Non-Central Parameter, 592
Normal Distribution, 31, 578
Bivariate, 583
Cumulative, 579
Multivariate, 583
Standard, 31, 579
Cumulative, 580
Normal Distribution Assumption, 27
Normal Model, 258
Normal Plot, 556
Normalization, 205
Normalizing, 204
NPV, 69

Numeraire, 205, 211

OAT, 54
Obligation Assimilable du Tresor, 54
ODAX, 322
OFDX, 322
Omega, 192, 194
Optimal Exercise Boundary, 94
Optimization
  Combinatorical, 522
  Global, 521
  Local, 521
  Numerical, 521
Option, 58, 339
  American, 58, 321, 339
    Plain Vanilla, 85
  Asian, 341, 353
  Atlantic, 339
  Bermudan, 339
  Binary, 344
  Digital, 344
  Embedded, 58
  European, 58, 320, 339
    Plain Vanilla, 85
  Exchange Traded, 58
  on Bond Futures, 326, 436
  on Bonds, 432
  on Coupon Bonds, 326
  on Forwards, 110
  on Futures, 109
  on Zero bonds, 326
  OTC, 58
  Plain Vanilla, 320, 339
  Premium, 58
  Price
    Normalized, 204
Option Stripping, 345, 351
OTC, 58
Out of Sample, 557
Out Of The Money, 58
Outlier, 452
Over Hedging, 449
Over The Counter, 58
Overfitting, 558

Par, 72
Par Rate, 72
  of a Coupon Bond, 290
  of a Forward Bond, 317
  of an Annuity Loan, 302
Par Swap, 313
Par Yield Curve, 73, 290
Parallel Shift, 547
  of the Term Structure, 546
Parity
  Put-Call, 88
Path, 174, 352
Path Dependence
  Short Rate Models, 250
Path Dependent, 175, 352
Pay Later Option, 343
Payer Swap, 293
Payment Date, 15, 48
Payoff, 339
PCA, 539
PDE, 113
pdf, 375, 566
Percentile, 143, 374
  Plot, 555
Performance, 70
Performance Index, 24, 323
Permutation, 576
Persistence
  of Volatility, 515
Phantom Structure, 535
Phase Space, 272
Pi, 169
Plain Vanilla, 50, 339
Portfolio
  Replicating, 148, 181
  Sensitivity, 393
Position Mapping, 394
Power Option, 340
Pre-Treatment
  of Time Series, 505, 549
Prediction, 525
Present Value, 67
  of a Bond, 287
  of a Floater, 285
  of a Forward Swap, 312
  of a FRA, 307
  of a Future

on a Zero Bond, 308
of a Swap, 294
of a Zero Bond, 284
of an Annuity Loan, 299
Present Value Weighting, 485
Previsible, 211, 260
Process, 210
Price Volatility, 442, 446, 487
Interpolation of, 440
Pricing, 4
Principal Issue, 364
Principle Axes, 542
Principle Component, 542
Principle Component Analysis, 539
Principle I, 5, 8
Probability, 565
Cumulative, 566
Density, 566
Function, 375, 566
Joint, 567
Multivariate, 567
Distribution, 566
Cumulative, 566
Empirical, 556
Multivariate, 567
Empirical, 556
Risk Neutral, 203
Probability Measure, 203
Risk Neutral, 98
Product Rule, 410
Profit Loss, 451
Clean, 452
Cleaned, 451
Dirty, 451
Prognosis
Linear
Best, 526
Optimal, 525
Promissory Note, 54
Put, 58, 339
Put-Call Equivalence
for Foreign Currency Options, 324
Put-Call Parity, 88
for Caps and Floors, 331
for Swaptions, 335, 336
Puttable Bond, 58
Pythagoras, 171
Pythagorean Theorem, 170

QQ Plot, 555
Quantile, 374
of the Normal Distribution, 379
Plot, 555

Rainbow Option, 341
Two-Color, 341
Random, 565
Number, 171, 565
Uncorrelated, 569
Variable, 27, 565
Independent, 389, 584
Uncorrelated, 583
Random Number Generator, 171, 593
Random Walk, 24, 512
Range, 374
Range Floater, 346
Ratchet Option, 351
Rating, 306
Rebate, 127, 347
Receiver Swap, 293
Recursion, 173
Red Zone, 454
Redemption, 283
Reference Currency, 487
Reference Rate, 285
Relative Price, 211
Rendleman-Barter, 275
Replicating Portfolio, 148
Residual Debt, 71
of an Annuity Loan, 297
Residues, 505, 553
Return
Mean, 32
Reverse Floater, 50
Rho, 192
Risk, 428
Risk, 373
Credit, 306
Delta, 9
Gamma, 9
Interest Rate, 47
Market, 373

Mitigation, 47
Operational, 373
Transfer, 47
Vega, 9
Risk Factor, 5, 11, 373
Risk Metrics, 377
Risk Neutral
Drift, 98, 157
Exercise probability, 162
Probability, 98, 155, 203
World, 155
Yield, 157
Risk Neutrality, 95, 98, 155, 203
Risk Ratios, 193
Rollover, 15
Date, 48
Day, 15
Rotation
of the Term Structure, 547

Scalar, 401
Scenario, 373
Schatz Future, 309
Schuldscheindarlehen, 54
SDE, 37
Season, 549
Seasonal Component
of Time Series, 505
Security
Coupon Bearing, 49
Interest Bearing, 49
Self Financing, 209, 211
Self Quanto Option, 340, 357
Self-Similarity, 26
Sensitivity, 182, 191
of a Portfolio, 392
Settlement
Daily, 57
Short Rate
Instantaneous, 230
Short Rate Model, 244
with Discrete Compounding, 277
Short Selling, 82
Short Sterling Future, 243
Significance
Statististical, 454
Simulated Annealing, 521, 537, 558
Skewness, 414, 570
of the Normal Distribution, 583
Smile, 475
Spectral Representation, 403
Spline, 440
Spot Market, 48
Spot Price, 23
Dividend Adjusted, 23
Spot Rate, 21
Curve, 461
Instantaneous, 230
Term Structure, 461
Spread, 286, 306
Risk, 305
Square Root of Time Law, 26, 382, 513, 536
Standard Deviation, 498, 566
Standard Normal Distribution, 579
Cumulative, 580
Approximation for, 580
Symmetry of, 580
Stationarity, 506
Statistics, 495
std, 566
Step Up Callable Bond, 366
Step Up Coupon, 52
Stochastic Analysis, 34
Stochastic Process, 24, 27
Straddle, 392
Straight Bond, 53
Stress Scenario, 545
Stress Test, 424
Strike, 58, 339
Clean, 326
Dirty, 326
Relative, 476
Strip, 329
Coupon, 54, 364, 468
EURIBOR, 54
LIBOR, 364, 468
Stripping, 4, 363
of a Floater, 286
STRIPS, 364
Structured Instrument, 363
Structured Monte Carlo, 420

Structured Product, 363
Stub, 466
Stub Rate, 468
Surrounding, 473
Swap, 54, 237, 292, 429, 468
  Plain Vanilla, 54, 294
Swap Gap, 471
Swap Rate, 60, 294, 334, 469
  for FX Transactions, 325
Swaption, 60, 293, 333, 439
  Payer, 333
  Receiver, 333
Symmetric Difference, 117
Symmetry
  Call-Put, 88

T-Bill, 11
  Future, 11
TARGET Holiday Calendar, 55
Taylor Series, 35, 75, 391
Term, 48, 52
Term Structure, 21, 229, 461
  Model, 229
Term-to-Maturity, 92
Thermodynamics, 521
Theta, 192
TIGR, 364
Time Bucket, 8
Time Lag, 506
Time Series, 495
  Centered, 505
  Estimated, 505
  Stationary, 505
  Weakly Stationary, 505
Time Series Analysis, 495, 505
  Multivariate, 540
Time to Maturity, 50
Total Return, 21
Total Return Swap, 297
Tower Law, 215
Trace
  of a Matrix, 410
Tradeable Instrument, 237
Trading, 5, 47
Traffic Light Approach, 452, 456
Transformation
  of Random Variables, 593
  of Scale, 552
  Orthonormal, 402
Transition Probability, 38
  Implied, 477
    from Black-Scholes Prices, 479
    Surface, 478
Transpose, 177
  of a Matrix, 386
Treasury Bill, 48
Treasury Note, 54
Tree Model, 244
Trend, 505, 549
Trinomial Tree, 162, 245
Trunc, 153
Twist, 547

U-L Decomposition, 137
Unbiased Estimate, 496
Uncorrelation
  of Random Variables, 584
Underlying, 58
Uniform Distribution, 574
Up and Out Barrier Call, 127
US Treasury Bond, 54
US Treasury Note, 54

Value at Risk, 373, 374, 445
  Relative, 386
Value Date, 14, 48
VaR, 374
VaR Mapping, 395
Variance, 566, 570
  Conditional, 508
  of a Sum, 485, 569
  of the Binomial Distribution, 578
  of the Chi-Squared Distribution, 591
  of the Gamma Distribution, 587
  of the Normal Distribution, 579
  of the Uniform Distribution, 575
  Unconditional, 508
Variance-Covariance Method, 391
Variation Margin, 84
Vasiceck, 275
VDAX, 476
Vector

Orthogonal, 402
Orthonormal, 402
Vega, 192
Vertex, 440, 469, 472
Volatility, 30
Black-76, 270
Black-Derman-Toy, 270
Black-Scholes, 479
Clustering, 508
Dupire, 479
Flat, 329
Forward-Forward, 329
Historical, 495
Ho-Lee, 270
Hull-White, 270
Implied, 475, 479
Index, 476
Local, 479
of a Portfolio, 484
Smile, 475
Surface, 476
Local, 482
Volatility Clustering, 513, 515
Volatility Forecast
GARCH(1,1), 533
GARCH(p,q), 528
Moving Average, 536
Volatility Term Structure, 476
GARCH(1,1), 535
GARCH(p,q), 533
Moving Average, 536

Warrant, 58
Weak Market Efficiency, 27
White Noise, 509, 513
Wiener Process, 31
Worst Case Scenario, 424

Yellow Zone, 453
Yield, 51
Yield Delta, 78
Yield To Maturity, 68
of a Bond, 287
of a Floater, 286
of a Forward Bond, 318
of a Forward Swap, 314
of a Swap, 296
of a Zero Bond, 284
of an Annuity Loan, 302
Yield Volatility, 446, 487
YTM, 68
Yule-Walker Equations, 510

Zero Bond, 52, 283, 363, 428
Issued With Discount, 52
Synthetic, 364
Zero Cost Collar, 331
Zero Coupon Bond, 52
Zone
Green, 456
Red, 454
Yellow, 453