
Preface

For quantitative researchers working in an investment bank, the process of writing a fixed income model usually has two stages. First, a theoretical framework for yield curve dynamics is specified, using the language of mathematics (especially stochastic calculus) to ensure that the underlying model is well-specified and internally consistent. Second, in order to use the model in practice, the equations arising from the first step need to be turned into a working implementation on a computer. While specification of the theoretical model may be seen as the difficult part, in quantitative finance applications the second step is technically and intellectually often more challenging than the first. In the implementation phase, not only does one need to translate abstract ideas into computer code, one also needs to ensure that the resulting numbers being produced are meaningful to a trading desk, are stable and robust, are in line with market observations, and are produced in a timely manner. Many of these requirements are, as it turns out, extremely challenging, and not only demand a strong knowledge of actual market practices (which tend to deviate in significant ways from “textbook” theory), but also require application of a large arsenal of techniques from applied mathematics, chiefly approximation methods and numerical techniques.

While there are many good introductory books on fixed income derivatives on the market, when we hire people who have read them we find that they still require significant training before they become productive members of our quantitative research teams. For one, while existing literature covers some aspects of the first step above, advanced approaches to specifying yield curve dynamics are typically not covered in sufficient detail. More importantly, there is simply too little said in the literature about the process of getting the theory to work in the real world of trading and risk management. An important goal of our book series is to close these gaps in the literature.

As we write this in early 2010, financial markets are still reeling from a severe crisis that has, at least in part, been blamed on over-the-counter (OTC) options markets, the venue where complex derivative securities are transacted. Stricter regulation of some types of OTC derivatives currently seems all but inevitable, and many common OTC securities may in the future either be outlawed or traded only on public exchanges. In the wake of the crisis, opinion of financial engineers and bankers has hit an all-time low, with many in the public convinced that they are peddlers of toxic waste or “weapons of financial destruction”. All things considered, the present may therefore seem like an inauspicious moment to launch a series of monographs on the pricing and risk management of interest rate derivatives. We disagree, for several reasons. First, in defense of OTC derivatives we note that although they certainly can be used inappropriately to create excessive leverage and risk, many complex (or “exotic”) derivatives serve as innovative and cost-effective vehicles for bank clients to reduce their financial risk. Second, irrespective of what will ultimately transpire on the regulatory front, it has become obvious that going forward both regulators and market participants need a better grasp of the management and characterization of complex financial risk. This is perhaps particularly true for the quantitative research professionals (the “quants”, in common parlance) who recently have been taken to task by the press for the failure of their models and their inability to predict the credit crisis. While this simplistic characterization is actually quite unfair, there is no doubt that many derivatives models that worked well enough before the credit crisis are no longer adequate. Indeed, even the simple task of pricing a basic interest rate swap — possibly the simplest of all interest rate derivatives — has recently required major methodology revisions¹. If nothing else, a severe crisis serves to expose weaknesses in the foundation on which models are built, allowing one to reinforce it for future storms. In this light, we feel that the time is just about right for a comprehensive, practical, and up-to-date exposition of interest rate modeling and risk management².

The three volumes of *Interest Rate Modeling* are aimed primarily at practitioners working in the area of interest rate derivatives, but much of the material is quite general and, we believe, will also hold significant appeal to researchers working in other asset classes. Students and academics interested in financial engineering and applied work will find the material particularly useful for its description of real-life model usage and for its expansive discussion of model calibration, approximation theory, and numerical methods. In preparing the books we have drawn on nearly 30 years of combined industry experience, and much of the material has never been exposed in book form before.

¹We cover this in Chapter 6.

²We ought to note that interest rate derivatives (unlike *credit* derivatives) so far have not been directly implicated in the financial crisis.

Quantitative finance attracts students and practitioners from many different academic fields, and with varying levels of preparation in mathematics and computation. (Case in point: L.B.G.A was originally a robotics engineer and V.V.P a probabilist.) To cater to a broad audience, we have kept the exposition fairly informal; graduate students in applied fields such as engineering and physics should feel at home with the level (or lack) of rigor used in the book. We have relied on a proposition-proof format throughout, largely because this facilitates easier cross-referencing in a long text, but acknowledge that the format is occasionally more formal than the results themselves. For instance, we tend to skip over technical regularity conditions in our proofs and also frequently list approximate results in propositions without explicitly specifying the sense in which they approximate true values. Although the exposition is largely self-contained, some previous knowledge of basic option pricing principles (e.g., at the level of Hull [2006]) may be useful.

Interest Rate Modeling divides into three separate volumes. *Volume I* provides the theoretical and computational foundations for the series, emphasizing the construction of efficient grid- and simulation-based methods for contingent claims pricing. Numerical methods serve an extremely important role in the text, so we develop this topic to an advanced level suitable for professional-quality model implementations. Placing this material early in the text allows us to incorporate it into our discussion of individual models in subsequent chapters. The second part of *Volume I* is dedicated to local-stochastic volatility modeling and to the construction of vanilla models for individual swap and Libor rates. Although the focus is eventually turned toward fixed income securities, much of the material in this volume applies to a broad capital market setting and will be of interest to anybody working in the general area of asset pricing.

Volume II is dedicated to in-depth study of term structure models of interest rates. While providing a thorough analysis of classical short rate models, the primary focus of the volume is on multi-factor stochastic volatility dynamics, in the setups of both the separable HJM and Libor market models. Implementation techniques are covered in detail, as are strategies for model parameterization and calibration to market data.

The first half of *Volume III* contains a detailed study of several classes of fixed income securities, ranging from simple vanilla options to highly exotic cancelable and path-dependent trades. The analysis is done in product-specific fashion, covering, among other subjects, risk characterization, calibration strategies, and valuation methods. In its second half, *Volume III* studies the general topic of derivative portfolio risk management, with a particular emphasis on the challenging problem of computing smooth price sensitivities to market input perturbations.

Although much of the material in *Interest Rate Modeling* is focused on the technical and theoretical issues surrounding model implementation on a computer, it is impractical for us to delve into the exercise of writing actual

computer routines. Fortunately, there are several specialized books on how to write good quant code, see, e.g., Hyer [2010] and Joshi [2004]. Both of these books work with C++ which is still the most common computer language used in professional quant libraries. For those that choose to work with C++, we wholeheartedly endorse books by Scott Meyers (see, e.g., Meyers [2005]) and Andrei Alexandrescu (see, e.g., Sutter and Alexandrescu [2004]) as guides to sound and maintainable code.

During the six year process of writing this book series, we have received encouragement and constructive criticism from many people. We particularly wish to thank Peter Carr, Peter Forsyth, Alexandre Antonov, Peter Jäckel, Dominique Bang, Martin Dahlgren, Neil Oliver, Patrick Roome, Regis van Steenkiste, Natasha Bushueva and many members of the research teams at Barclays Capital and Bank of America Merrill Lynch. Natalia Kryzhanovskaya meticulously proofread our first draft, and contributed greatly to the harmonization of notation across what turned out to be a very long manuscript. All remaining errors are, of course, entirely our own. Speaking of errors: with nearly 20,000 equations, it is probable that a few typos remain, despite our best efforts to weed them out. A list of errata will be maintained on www.andersen-piterbarg-book.com where supplemental material and news will also be posted on a running basis. We greatly appreciate reporting of typos or factual errors to our web address, and will list the names of all those who contribute to error spotting in future editions of *Interest Rate Modeling*.

Lastly, we owe a great debt of gratitude to our families for their support and patience, even when our initial plans for a brief book on tips and tricks for working quants ballooned into something more ambitious that consumed many evenings and weekends over the last six years.

London, New York,
June 2004 — August 2010

*Leif B.G Andersen
Vladimir V. Piterbarg*

Table of Contents for All Volumes

VOLUME I Foundations and Vanilla Models

Part I Foundations

1	Introduction to Arbitrage Pricing Theory	3
1.1	The Setup	3
1.2	Trading Gains and Arbitrage	7
1.3	Equivalent Martingale Measures and Arbitrage	8
1.4	Derivative Security Pricing and Complete Markets	10
1.5	Girsanov's Theorem	12
1.6	Stochastic Differential Equations	14
1.7	Explicit Trading Strategies and PDEs	16
1.8	Kolmogorov's Equations and the Feynman-Kac Theorem	18
1.9	Black-Scholes and Extensions	21
1.9.1	Basics	21
1.9.2	Alternative Derivation	25
1.9.3	Extensions	27
1.9.3.1	Deterministic Parameters and Dividends	27
1.9.3.2	Stochastic Interest Rates	28
1.10	Options with Early Exercise Rights	30
1.10.1	The Markovian Case	32

1.10.2	Some General Bounds	34
1.10.3	Early Exercise Premia	36
2	Finite Difference Methods	43
2.1	1-Dimensional PDEs: Problem Formulation	43
2.2	Finite Difference Discretization	45
2.2.1	Discretization in x -Direction. Dirichlet Boundary Conditions	45
2.2.2	Other Boundary Conditions	47
2.2.3	Time-Discretization	49
2.2.4	Finite Difference Scheme	50
2.3	Stability	52
2.3.1	Matrix Methods	52
2.3.2	Von Neumann Analysis	53
2.4	Non-Equidistant Discretization	56
2.5	Smoothing and Continuity Correction	58
2.5.1	Crank-Nicolson Oscillation Remedies	58
2.5.2	Continuity Correction	58
2.5.3	Grid Shifting	59
2.6	Convection-Dominated PDEs	60
2.6.1	Upwinding	61
2.6.2	Other Techniques	62
2.7	Option Examples	63
2.7.1	Continuous Barrier Options	63
2.7.2	Discrete Barrier Options	65
2.7.3	Coupon-Paying Securities and Dividends	67
2.7.4	Securities with Early Exercise	68
2.7.5	Path-Dependent Options	69
2.7.6	Multiple Exercise Rights	70
2.8	Special Issues	72
2.8.1	Mesh Refinements for Multiple Events	72
2.8.2	Analytics at the Last Time Step	75
2.8.3	Analytics at the First Time Step	76
2.9	Multi-Dimensional PDEs: Problem Formulation	78
2.10	Two-Dimensional PDE with No Mixed Derivatives	79
2.10.1	Theta Method	80
2.10.2	The Alternating Direction Implicit (ADI) Method	81
2.10.3	Boundary Conditions and Other Issues	84
2.11	Two-Dimensional PDE with Mixed Derivatives	85
2.11.1	Orthogonalization of the PDE	85
2.11.2	Predictor-Corrector Scheme	88
2.12	PDEs of Arbitrary Order	91

3 Monte Carlo Methods	93
3.1 Fundamentals	93
3.1.1 Generation of Random Samples	95
3.1.1.1 Inverse Transform Method	96
3.1.1.2 Acceptance-Rejection Method	97
3.1.1.3 Composition	99
3.1.2 Correlated Gaussian Samples	100
3.1.2.1 Cholesky Decomposition	101
3.1.2.2 Eigenvalue Decomposition	102
3.1.3 Principal Components Analysis (PCA)	103
3.2 Generation of Sample Paths	104
3.2.1 Example: Asian Basket Options in Black-Scholes Economy	104
3.2.2 Discretization Schemes, Convergence, and Stability	106
3.2.3 The Euler Scheme	108
3.2.3.1 Linear-Drift SDEs	110
3.2.3.2 Log-Euler Scheme	110
3.2.4 The Implicit Euler Scheme	111
3.2.4.1 Implicit Diffusion Term	112
3.2.5 Predictor-Corrector Schemes	113
3.2.6 Ito-Taylor Expansions and Higher-Order Schemes	114
3.2.6.1 Ordinary Taylor Expansion of ODEs	115
3.2.6.2 Ito-Taylor Expansions	116
3.2.6.3 Milstein Second-Order Discretization Scheme	117
3.2.7 Other Second-Order Schemes	119
3.2.8 Bias vs. Monte Carlo Error	120
3.2.9 Sampling of Continuous Process Extremes	122
3.2.10 PCA and Bridge Construction of Brownian Motion Paths	126
3.2.10.1 Brownian Bridge and Quasi-Random Sequences	126
3.2.10.2 PC Construction	128
3.3 Sensitivity Computations	129
3.3.1 Finite Difference Estimates	129
3.3.1.1 Black-Scholes Delta	129
3.3.1.2 General Case	131
3.3.2 Pathwise Estimate	133
3.3.2.1 Black-Scholes Delta	133
3.3.2.2 General Case	134
3.3.2.3 Sensitivity Path Generation	136
3.3.3 Likelihood Ratio Method	136
3.3.3.1 Black-Scholes Delta	137
3.3.3.2 General Case	138
3.3.3.3 Euler Schemes	138

3.3.3.4	Some Remarks	139
3.4	Variance Reduction Techniques	140
3.4.1	Variance Reduction and Efficiency	141
3.4.2	Antithetic Variates	141
3.4.2.1	The Gaussian Case	141
3.4.2.2	General Case	143
3.4.3	Control Variates	143
3.4.3.1	Basic Idea	143
3.4.3.2	Non-Linear Controls	145
3.4.4	Importance Sampling	146
3.4.4.1	Basic Idea	146
3.4.4.2	Density Formulation	147
3.4.4.3	Importance Sampling and SDEs	149
3.4.4.4	More on SDE Path Simulation	150
3.4.4.5	Rare Event Simulation and Linearization	152
3.5	Some Notes on Bermudan Security Pricing	156
3.5.1	Basic Idea	156
3.5.2	Parametric Lower Bound Methods	157
3.5.3	Parametric Lower Bound: An Example	158
3.5.4	Regression-Based Lower Bound	159
3.5.5	Upper Bound Methods	160
3.5.6	Confidence Intervals	161
3.5.7	Other Methods	162
3.A	Appendix: Constants for Φ^{-1} Algorithm	163
4	Fundamentals of Interest Rate Modeling	165
4.1	Fixed Income Notations	165
4.1.1	Bonds and Forward Rates	165
4.1.2	Futures Rates	167
4.1.3	Annuity Factors and Par Rates	168
4.2	Fixed Income Probability Measures	169
4.2.1	Risk Neutral Measure	170
4.2.2	T -Forward Measure	172
4.2.3	Spot Measure	173
4.2.4	Terminal and Hybrid Measures	174
4.2.5	Swap Measures	175
4.3	Multi-Currency Markets	176
4.3.1	Notations and FX Forwards	176
4.3.2	Risk Neutral Measures	177
4.3.3	Other Measures	178
4.4	The HJM Analysis	179
4.4.1	Bond Price Dynamics	179
4.4.2	Forward Rate Dynamics	180
4.4.3	Short Rate Process	181
4.5	Examples of HJM Models	182

4.5.1	The Gaussian Model	182
4.5.2	Gaussian HJM Models with Markovian Short Rate	185
4.5.3	Log-Normal HJM Models	187
5	Fixed Income Instruments	189
5.1	Fixed Income Markets and Participants	189
5.2	Certificates of Deposit and Libor Rates	192
5.3	Forward Rate Agreements (FRA)	193
5.4	Eurodollar Futures	194
5.5	Fixed-for-Floating Swaps	195
5.6	Libor-in-Arrears Swaps	198
5.7	Averaging Swaps	199
5.8	Caps and Floors	199
5.9	Digital Caps and Floors	201
5.10	European Swaptions	201
5.10.1	Cash-Settled Swaptions	203
5.11	CMS Swaps, Caps and Floors	204
5.12	Bermudan Swaptions	205
5.13	Exotic Swaps and Structured Notes	206
5.13.1	Libor-Based Exotic Swaps	207
5.13.2	CMS-Based Exotic Swaps	208
5.13.3	Multi-Rate Exotic Swaps	208
5.13.4	Range Accruals	209
5.13.5	Path-Dependent Swaps	210
5.14	Callable Libor Exotics	211
5.14.1	Definitions	211
5.14.2	Pricing Callable Libor Exotics	213
5.14.3	Types of Callable Libor Exotics	214
5.14.4	Callable Snowballs	214
5.14.5	CLEs Accreting at Coupon Rate	214
5.14.6	Multi-Tranches	215
5.15	TARNs and Other Trade-Level Features	215
5.15.1	Knock-out Swaps	216
5.15.2	TARNs	216
5.15.3	Global Cap	217
5.15.4	Global Floor	217
5.15.5	Pricing and Trade Representation Challenges	217
5.16	Volatility Derivatives	218
5.16.1	Volatility Swaps	218
5.16.2	Volatility Swaps with a Shout	219
5.16.3	Min-Max Volatility Swaps	220
5.16.4	Forward Starting Options and Other Forward Volatility Contracts	220
5.A	Appendix: Day Counting Rules and Other Trivia	221
5.A.1	Libor Rate Definitions	222
5.A.2	Swap Payments	223

Part II Vanilla Models

6	Yield Curve Construction and Risk Management	227
6.1	Notations and Problem Definition	228
6.1.1	Discount Curves	228
6.1.2	Matrix Formulation	230
6.1.3	Construction Principles and Yield Curves	230
6.2	Yield Curve Fitting with N -Knot Splines	232
6.2.1	C^0 Yield Curves: Bootstrapping	232
6.2.1.1	Piecewise Linear Yields	233
6.2.1.2	Piecewise Flat Forward Rates	234
6.2.2	C^1 Yield Curves: Hermite Splines	236
6.2.3	C^2 Yield Curves: Twice Differentiable Cubic Splines	238
6.2.4	C^2 Yield Curves: Twice Differentiable Tension Splines	241
6.3	Non-Parametric Optimal Yield Curve Fitting	243
6.3.1	Norm Specification and Optimization	243
6.3.2	Choice of λ	246
6.3.3	Example	247
6.4	Managing Yield Curve Risk	248
6.4.1	Par-Point Approach	249
6.4.2	Forward Rate Approach	250
6.4.3	From Risks to Hedging: The Jacobian Approach ..	252
6.4.4	Cumulative Shifts and other Common Tricks ..	254
6.5	Various Topics in Discount Curve Construction	256
6.5.1	Curve Overlays and Turn-of-Year Effects	256
6.5.2	Cross-Currency Curve Construction	257
6.5.2.1	Basic Problem	257
6.5.2.2	Separation of Discount and Forward Rate Curves	258
6.5.2.3	Cross-Currency Basis Swaps	260
6.5.2.4	Modified Curve Construction Algorithm ..	261
6.5.3	Tenor Basis and Multi-Index Curve Group Construction	263
6.A	Appendix: Spline Theory	268
6.A.1	Hermite Spline Theory	268
6.A.2	C^2 Cubic Splines	271
6.A.3	C^2 Exponential Tension Splines	272
7	Vanilla Models with Local Volatility	275
7.1	General Framework	276
7.1.1	Model Dynamics	276
7.1.2	Volatility Smile and Implied Density	276

7.1.3	Choice of φ	277
7.2	CEV Model	278
7.2.1	Basic Properties	278
7.2.2	Call Option Pricing	280
7.2.3	Regularization	282
7.2.4	Displaced Diffusion Models	283
7.3	Quadratic Volatility Model	285
7.3.1	Case 1: Two Real Roots to the Left of $S(0)$	285
7.3.2	Case 2: One Real Root to the Left of $S(0)$	289
7.3.3	Extensions and Other Root Configurations	289
7.4	Finite Difference Solutions for General φ	290
7.4.1	Multiple λ and T	291
7.4.2	Forward Equation for Call Options	291
7.5	Asymptotic Expansions for General φ	293
7.5.1	Expansion around Displaced Log-Normal Process	293
7.5.2	Expansion around Gaussian Process	296
7.6	Extensions to Time-Dependent φ	297
7.6.1	Separable Case	297
7.6.2	Skew Averaging	298
7.6.2.1	Examples	302
7.6.2.2	A Caveat About the Process Domain	304
7.6.3	Skew and Convexity Averaging by Small-Noise Expansion	305
7.6.4	Numerical Example	309
8	Vanilla Models with Stochastic Volatility I	313
8.1	Model Definition	313
8.2	Model Parameters	315
8.3	Basic Properties	316
8.4	Fourier Integration	322
8.4.1	General Theory	322
8.4.2	Applications to SV Model	325
8.4.3	Numerical Implementation	328
8.4.4	Refinements of Numerical Implementation	330
8.4.5	Fourier Integration for Arbitrary European Payoffs	334
8.5	Integration in Variance Domain	337
8.6	CEV-Type Stochastic Volatility Models and SABR	341
8.7	Numerical Examples: Volatility Smile Statics	343
8.8	Numerical Examples: Volatility Smile Dynamics	345
8.9	Hedging in Stochastic Volatility Models	350
8.9.1	Hedge Construction, Delta and Vega	350
8.9.2	Minimum Variance Delta Hedging	353
8.9.3	Minimum Variance Hedging: an Example	354
8.A	Appendix: General Volatility Processes	356

9	Vanilla Models with Stochastic Volatility II	359
9.1	Fourier Integration with Time-Dependent Parameters	359
9.2	Asymptotic Expansion with Time-Dependent Volatility	362
9.3	Averaging Methods	366
9.3.1	Volatility Averaging	367
9.3.2	Skew Averaging	369
9.3.3	Volatility of Variance Averaging	370
9.3.4	Calibration by Parameter Averaging	372
9.4	PDE Method	377
9.4.1	PDE Formulation	377
9.4.2	Range for Stochastic Variance	378
9.4.3	Discretizing Stochastic Variance	379
9.4.4	Boundary Conditions for Stochastic Variance	381
9.4.5	Range for Underlying	382
9.4.6	Discretizing the Underlying	383
9.5	Monte Carlo Method	383
9.5.1	Exact Simulation of Variance Process	384
9.5.2	Biased Taylor-Type Schemes for Variance Process	385
9.5.2.1	Euler Schemes	385
9.5.2.2	Higher-Order Schemes	385
9.5.3	Moment Matching Schemes for Variance Process	386
9.5.3.1	Log-normal Approximation	386
9.5.3.2	Truncated Gaussian	387
9.5.3.3	Quadratic-Exponential	388
9.5.3.4	Summary of QE Algorithm	390
9.5.4	Broadie-Kaya Scheme for the Underlying	390
9.5.5	Other Schemes for the Underlying	392
9.5.5.1	Taylor-Type Schemes	392
9.5.5.2	Simplified Broadie-Kaya	392
9.5.5.3	Martingale Correction	392
9.A	Appendix: Proof of Proposition 9.3.4	393
9.B	Appendix: Coefficients for Asymptotic Expansion	397

VOLUME II Term Structure Models

Part III Term Structure Models

10 One-Factor Short Rate Models I	401
10.1 The One-Factor Gaussian Short Rate Model	402
10.1.1 The Ho-Lee Model	402
10.1.1.1 Notations and First Steps	402
10.1.1.2 Fitting the Term Structure of Discount Bonds	403
10.1.1.3 Analysis and Comparison with HJM Approach	405
10.1.2 The Mean-Reverting GSR Model	407
10.1.2.1 The Vasicek Model	407
10.1.2.2 The General One-Factor GSR Model . . .	409
10.1.2.3 Time-Stationarity and Caplet Hump . . .	412
10.1.3 European Option Pricing	414
10.1.3.1 The Jamshidian Decomposition	414
10.1.3.2 Gaussian Swap Rate Approximation . . .	416
10.1.4 Swaption Calibration	417
10.1.5 Finite Difference Methods	418
10.1.5.1 PDE and Spatial Boundary Conditions .	419
10.1.5.2 Determining Spatial Boundary Conditions from PDE	420
10.1.5.3 Upwinding	421
10.1.6 Monte Carlo Simulation	421
10.1.6.1 Exact Discretization	421
10.1.6.2 Approximate Discretization	423
10.1.6.3 Using other Measures for Simulation . .	424
10.2 The Affine One-Factor Model	425
10.2.1 Basic Definitions	425
10.2.1.1 SDE	425
10.2.1.2 Regularity Issues	426
10.2.1.3 Volatility Skew	426
10.2.1.4 Time-Dependent Parameters	427
10.2.2 Discount Bond Pricing and Extended Transform .	427
10.2.2.1 Constant Parameters	428
10.2.2.2 Piecewise Constant Parameters	430
10.2.3 Discount Bond Calibration	431
10.2.3.1 Change of Variables	431
10.2.3.2 Algorithm for $\omega(t)$	432
10.2.4 European Option Pricing	433

10.2.5	Swaption Calibration	435
10.2.5.1	Basic Problem	435
10.2.5.2	Calibration Algorithm	436
10.2.6	Quadratic One-Factor Model	437
10.2.7	Numerical Methods for the Affine Short Rate Model	437
11	One-Factor Short Rate Models II	439
11.1	Log-Normal Short Rate Models	439
11.1.1	The Black-Derman-Toy Model	439
11.1.2	Black-Karasinski Model	441
11.1.3	Issues in Log-Normal Models	441
11.1.4	Sandmann-Sondermann Transformation	442
11.2	Other Short Rate Models	445
11.2.1	Power-Type Models and Empirical Model Estimation	445
11.2.2	The Black Shadow Rate Model	446
11.2.3	Spanned and Unspanned Stochastic Volatility: the Fong and Vasicek Model	448
11.3	Numerical Methods for General One-Factor Short Rate Models	449
11.3.1	Finite Difference Methods	450
11.3.2	Calibration to Initial Yield Curve	451
11.3.2.1	Forward Induction	452
11.3.2.2	Forward-from-Backward Induction	453
11.3.2.3	Yield Curve and Volatility Calibration	455
11.3.2.4	The Dybvig Parameterization	457
11.3.2.5	Link to HJM Models	458
11.3.2.6	The Hagan and Woodward Parameterization	459
11.3.3	Monte Carlo Simulation	462
11.3.3.1	SDE Discretization	462
11.3.3.2	Practical Issues with Monte Carlo Methods	464
11.A	Appendix: Markov-Functional Models	466
11.A.1	State Process and Numeraire Mapping	466
11.A.2	Libor MF Parameterization	467
11.A.3	Swap MF Parameterization	469
11.A.4	Non-Parametric Calibration	470
11.A.5	Numerical Implementation	471
11.A.6	Comments and Comparisons	472

12 Multi-Factor Short Rate Models	473
12.1 The Gaussian Model	474
12.1.1 Development from Separability Condition	474
12.1.1.1 Mean-Reverting State Variables	475
12.1.1.2 Further Changes of Variables	479
12.1.2 Classical Development	481
12.1.2.1 Diagonalization of Mean Reversion Matrix	482
12.1.3 Correlation Structure	484
12.1.4 The Two-Factor Gaussian Model	485
12.1.4.1 Some Basics	485
12.1.4.2 Variance and Correlation Structure	486
12.1.4.3 Volatility Hump	487
12.1.4.4 Another Formulation of the Two-Factor Model	488
12.1.5 Multi-Factor Statistical Gaussian Model	491
12.1.6 Swaption Pricing	496
12.1.6.1 Jamshidian Decomposition	496
12.1.6.2 Gaussian Swap Rate Approximation	500
12.1.7 Calibration via Benchmark Rates	501
12.1.8 Monte Carlo Simulation	504
12.1.9 Finite Difference Methods	505
12.2 The Affine Model	506
12.2.1 Introduction	506
12.2.2 Basic Model	507
12.2.3 Regularity Issues	508
12.2.4 Discount Bond Prices	509
12.2.5 Some Concrete Models	511
12.2.5.1 Fong-Vasicek Model	511
12.2.5.2 Longstaff-Schwartz Model	512
12.2.5.3 Multi-Factor CIR Models	513
12.2.6 Brief Notes on Option Pricing	514
12.3 The Quadratic Gaussian Model	514
12.3.1 Quadratic Gaussian Models are Affine	515
12.3.2 The Basics	516
12.3.3 Parameterization	518
12.3.3.1 Smile Generation	518
12.3.3.2 Quadratic Term	519
12.3.3.3 Linear Term	521
12.3.4 Swaption Pricing	522
12.3.4.1 State Vector Distribution Under the Annuity Measure	522
12.3.4.2 Exact Pricing of European Swaptions	523
12.3.4.3 Approximations for European Swaptions	524
12.3.5 Calibration	527

12.3.6	Spanned Stochastic Volatility	528
12.3.7	Numerical Methods	528
12.A	Appendix: Quadratic Forms of Gaussian Vectors	528
13	The Quasi-Gaussian Model	533
13.1	One-Factor Quasi-Gaussian Model	533
13.1.1	Definition	533
13.1.2	Local Volatility	535
13.1.3	Swap Rate Dynamics	536
13.1.4	Approximate Local Volatility Dynamics for Swap Rate	537
13.1.4.1	Simple Approximation	538
13.1.4.2	Advanced Approximation	538
13.1.5	Linear Local Volatility	541
13.1.6	Linear Local Volatility for a Swaption Strip	543
13.1.7	Volatility Calibration	544
13.1.8	Mean Reversion Calibration	546
13.1.8.1	Effects of Mean Reversion	546
13.1.8.2	Calibrating Mean Reversion to Volatility Ratios	548
13.1.8.3	Calibrating Mean Reversion to Inter-Temporal Correlations	551
13.1.8.4	Final Comments on Mean Reversion Calibration	553
13.1.9	Numerical Methods	554
13.1.9.1	Direct Integration	554
13.1.9.2	Finite Difference Methods	556
13.1.9.3	Monte Carlo Simulation	559
13.1.9.4	Single-State Approximations	559
13.2	One-Factor Quasi-Gaussian Model with Stochastic Volatility	563
13.2.1	Definition	563
13.2.2	Swap Rate Dynamics	564
13.2.3	Volatility Calibration	566
13.2.4	Mean Reversion Calibration	567
13.2.5	Non-Zero Correlation	567
13.2.6	PDE and Monte Carlo Methods	568
13.3	Multi-Factor Quasi-Gaussian Model	568
13.3.1	General Multi-Factor Model	568
13.3.2	Local and Stochastic Volatility Parameterization ..	570
13.3.3	Swap Rate Dynamics and Approximations	572
13.3.4	Volatility Calibration	577
13.3.5	Mean Reversions, Correlations, and Numerical Schemes	578
13.A	Appendix: Density Approximation	579

13.A.1	Simplified Forward Measure Dynamics	579
13.A.2	Effective Volatility	580
13.A.3	The Forward Equation for Call Options	581
13.A.4	Asymptotic Expansion	582
13.A.5	Proof of Theorem 13.1.14	583
14	The Libor Market Model I	585
14.1	Introduction and Setup	586
14.1.1	Motivation and Historical Notes	586
14.1.2	Tenor Structure	587
14.2	LM Dynamics and Measures	587
14.2.1	Setting	587
14.2.2	Probability Measures	588
14.2.3	Link to HJM Analysis	591
14.2.4	Separable Deterministic Volatility Function	592
14.2.5	Stochastic Volatility	594
14.2.6	Time-Dependence in Model Parameters	597
14.3	Correlation	597
14.3.1	Empirical Principal Components Analysis	598
14.3.1.1	Example: USD Forward Rates	599
14.3.2	Correlation Estimation and Smoothing	600
14.3.2.1	Example: Fit to USD Data	603
14.3.3	Negative Eigenvalues	604
14.3.4	Correlation PCA	605
14.3.4.1	Example: USD Data	607
14.3.4.2	Poor Man's Correlation PCA	608
14.4	Pricing of European Options	608
14.4.1	Caplets	609
14.4.2	Swaptions	610
14.4.3	Spread Options	613
14.4.3.1	Term Correlation	614
14.4.3.2	Spread Option Pricing	615
14.5	Calibration	615
14.5.1	Basic Principles	615
14.5.2	Parameterization of $\ \lambda_k(t)\ $	616
14.5.3	Interpolation on the Whole Grid	617
14.5.4	Construction of $\lambda_k(t)$ from $\ \lambda_k(t)\ $	619
14.5.4.1	Covariance PCA	620
14.5.4.2	Correlation PCA	620
14.5.4.3	Discussion and Recommendation	621
14.5.5	Choice of Calibration Instruments	621
14.5.6	Calibration Objective Function	624
14.5.7	Sample Calibration Algorithm	626
14.5.8	Speed-Up Through Sub-Problem Splitting	627
14.5.9	Correlation Calibration to Spread Options	629

14.5.10	Volatility Skew Calibration	631
14.6	Monte Carlo Simulation	631
14.6.1	Euler-Type Schemes	632
14.6.1.1	Analysis of Computational Effort	633
14.6.1.2	Long Time Steps	634
14.6.1.3	Notes on the Choice of Numeraire	636
14.6.2	Other Simulation Schemes	636
14.6.2.1	Special-Purpose Schemes with Drift Predictor-Corrector	637
14.6.2.2	Euler Scheme with Predictor-Corrector	638
14.6.2.3	Lagging Predictor-Corrector Scheme	638
14.6.2.4	Further Refinements of Drift Estimation	640
14.6.2.5	Brownian-Bridge Schemes and Other Ideas	641
14.6.2.6	High-Order Schemes	643
14.6.3	Martingale Discretization	644
14.6.3.1	Deflated Bond Price Discretization	645
14.6.3.2	Comments and Alternatives	646
14.6.4	Variance Reduction	647
14.6.4.1	Antithetic Sampling	647
14.6.4.2	Control Variates	648
14.6.4.3	Importance Sampling	648
15	The Libor Market Model II	651
15.1	Interpolation	651
15.1.1	Back Stub, Simple Interpolation	652
15.1.2	Back Stub, Arbitrage-Free Interpolation	653
15.1.3	Back Stub, Gaussian Model	655
15.1.4	Front Stub, Zero Volatility	656
15.1.5	Front Stub, Exogenous Volatility	657
15.1.6	Front Stub, Simple Interpolation	660
15.1.7	Front Stub, Gaussian Model	661
15.2	Advanced Swap Pricing via Markovian Projection	662
15.2.1	Advanced Formula for Swap Rate Volatility	664
15.2.2	Advanced Formula for Swap Rate Skew	666
15.2.3	Skew and Smile Calibration in LM Models	668
15.3	Near-Markov LM Models	670
15.4	Swap Market Models	670
15.5	Evolving Separate Discount and Forward Rate Curves	672
15.5.1	Basic Ideas	673
15.5.2	HJM Extension	674
15.5.3	Applications to LM Models	677
15.5.4	Deterministic Spread	681
15.6	SV Models with Non-Zero Correlation	681
15.7	Multi-Stochastic Volatility Extensions	683

15.7.1	Introduction	683
15.7.2	Setup	684
15.7.3	Pricing Caplets and Swaptions	685
15.7.4	Spread Options.	686
15.7.5	Another Use of Multi-Dimensional Stochastic Volatility	687

VOLUME III Products and Risk Management

Part IV Products

16 Single-Rate Vanilla Derivatives	691
16.1 European Swaptions	691
16.1.1 Smile Dynamics	692
16.1.2 Adjustable Backbone.	693
16.1.3 Stochastic Volatility Swaption Grid	696
16.1.4 Calibrating Stochastic Volatility Model to Swaptions	697
16.1.5 Some Other Interpolation Rules	699
16.2 Caps and Floors.	700
16.2.1 Basic Problem	700
16.2.2 Setup and Norms	701
16.2.3 Calibration Procedure.	702
16.3 Terminal Swap Rate Models	703
16.3.1 TSR Basics	703
16.3.2 Linear TSR Model	705
16.3.3 Exponential TSR Model	708
16.3.4 Swap-Yield TSR Model	709
16.4 Libor-in-Arrears	710
16.5 Libor-with-Delay	713
16.5.1 Swap-Yield TSR Model	714
16.5.2 Other Terminal Swap Rate Models	715
16.5.3 Approximations Inspired by Term Structure Models	715
16.5.4 Applications to Averaging Swaps	716
16.6 CMS and CMS-Linked Cash Flows	717
16.6.1 The Replication Method for CMS.	718
16.6.2 Annuity Mapping Function as a Conditional Expected Value	720
16.6.3 Swap-Yield TSR Model	722
16.6.4 Linear and Other TSR Models	722
16.6.5 The Quasi-Gaussian Model	724
16.6.6 The Libor Market Model	725
16.6.7 Correcting Non-Arbitrage-Free Methods	728
16.6.8 Impact of Annuity Mapping Function and Mean Reversion	729
16.6.9 CDF and PDF of CMS Rate in Forward Measure. .	730
16.6.10 SV Model for CMS Rate	734

16.6.11	Dynamics of CMS Rate in Forward Measure	735
16.6.12	Cash-Settled Swaptions	738
16.7	Quanto CMS	740
16.7.1	Overview	740
16.7.2	Modeling the Joint Distribution of Swap Rate and Forward Exchange Rate	742
16.7.3	Normalizing Constant and Final Formula	743
16.8	Eurodollar Futures	744
16.8.1	Fundamental Results on Futures.....	745
16.8.2	Motivations and Plan	747
16.8.3	Preliminaries.....	748
16.8.4	Expansion Around the Futures Value	748
16.8.5	Forward Rate Variances	751
16.8.6	Forward Rate Correlations	753
16.8.7	The Formula	754
16.9	Convexity and Moment Explosions	755
17	Multi-Rate Vanilla Derivatives	759
17.1	Introduction to Multi-Rate Vanilla Derivatives	759
17.2	Marginal Distributions and Reference Measure	761
17.3	Dependence Structure via Copulas	762
17.3.1	Introduction to Gaussian Copula Method	762
17.3.2	General Copulas	764
17.3.3	Archimedean Copulas	766
17.3.4	Making Copulas from Other Copulas	767
17.4	Copula Methods for CMS Spread Options	770
17.4.1	Normal Model for the Spread	770
17.4.2	Gaussian Copula for Spread Options	771
17.4.3	Spread Volatility Smile Modeling with the Power Gaussian Copula	774
17.4.4	Copula Implied From Spread Options	775
17.5	Rates Observed at Different Times.....	778
17.6	Numerical Methods for Copulas	779
17.6.1	Numerical Integration Methods.....	780
17.6.2	Dimensionality Reduction for CMS Spread Options	783
17.6.3	Dimensionality Reduction for Other Multi-Rate Derivatives	785
17.6.4	Dimensionality Reduction by Conditioning.....	787
17.6.5	Dimensionality Reduction by Measure Change ...	791
17.6.6	Monte Carlo Methods	793
17.7	Limitations of the Copula Method	795
17.8	Stochastic Volatility Modeling for Multi-Rate Options....	796
17.8.1	Measure Change by Drift Adjustment	797
17.8.2	Measure Change by CMS Caplet Calibration	798
17.8.3	Impact of Correlations on the Spread Smile	799

XXVIII Contents

17.8.4	Connection to Term Structure Models	800
17.9	CMS Spread Options in Term Structure Models	802
17.9.1	Libor Market Model	802
17.9.2	Quadratic Gaussian Model	804
17.A	Appendix: Implied Correlation in Displaced Log-Normal Models	805
17.A.1	Preliminaries	805
17.A.2	Implied Log-Normal Correlation	806
17.A.3	A Few Numerical Results	807
18	Callable Libor Exotics	809
18.1	Model Calibration for Callable Libor Exotics	809
18.1.1	Risk Factors for CLEs	810
18.1.2	Model Choice and Calibration	813
18.2	Valuation Theory	814
18.2.1	Preliminaries	814
18.2.2	Recursion for Callable Libor Exotics	815
18.2.3	Marginal Exercise Value Decomposition	816
18.3	Monte Carlo Valuation	817
18.3.1	Regression-Based Valuation of CLEs, Basic Scheme	817
18.3.2	Regression for Underlying	819
18.3.3	Valuing CLE as a Cancelable Note	821
18.3.4	Using Regressed Variables for Decision Only	822
18.3.5	Regression Valuation with Boundary Optimization	824
18.3.6	Lower Bound via Regression Scheme	825
18.3.7	Iterative Improvement of Lower Bound	827
18.3.8	Upper Bound	830
18.3.8.1	Basic Ideas	830
18.3.8.2	Nested Simulation (NS) Algorithm	831
18.3.8.3	Bias and Computational Cost of NS Algorithm	834
18.3.8.4	Confidence Intervals and Practical Usage	836
18.3.8.5	Non-Analytic Exercise Values	837
18.3.8.6	Improvements to NS Algorithm	839
18.3.8.7	Other Upper Bound Algorithms	841
18.3.9	Regression Variable Choice	842
18.3.9.1	State Variables Approach	842
18.3.9.2	Explanatory Variables	843
18.3.9.3	Explanatory Variables with Convexity	846
18.3.10	Regression Implementation	848
18.3.10.1	Automated Explanatory Variable Selection	848
18.3.10.2	Suboptimal Point Exclusion	850
18.3.10.3	Two Step Regression	851

18.3.10.4	Robust Implementation of Regression Algorithm	852
18.4	Valuation with Low-Dimensional Models	856
18.4.1	Single-Rate Callable Libor Exotics	856
18.4.2	Calibration Targets for the Local Projection Method	856
18.4.3	Review of Suitable Local Models	857
18.4.4	Defining a Suitable Analog for Core Swap Rates	859
18.4.5	PDE Methods for Path-Dependent CLEs	861
18.4.5.1	CLEs Accreting at Coupon Rate	862
18.4.5.2	Snowballs	864
19	Bermudan Swaptions	867
19.1	Definitions	867
19.2	Local Projection Method	868
19.3	Smile Calibration	870
19.4	Amortizing, Accreting, Other Non-Standard Swaptions	872
19.4.1	Relationship Between Non-Standard and Standard Swap Rates	874
19.4.2	Same-Tenor Approach	875
19.4.3	Representative Swaption Approach	876
19.4.4	Basket Approach	879
19.4.5	Super-Replication for Non-Standard Bermudan Swaptions	882
19.4.6	Zero-Coupon Bermudan Swaptions	886
19.4.7	American Swaptions	887
19.4.7.1	American Swaptions vs. High- Frequency Bermudan Swaptions	888
19.4.7.2	The Proxy Libor Rate Method	889
19.4.7.3	The Libor-as-Extra-State Method	890
19.4.8	Mid-Coupon Exercise	891
19.5	Flexi-Swaps	892
19.5.1	Purely Global Bounds	893
19.5.2	Purely Local Bounds	893
19.5.3	Marginal Exercise Value Decomposition	895
19.5.4	Narrow Band Limit	896
19.6	Monte Carlo Valuation	897
19.6.1	Regression Methods	897
19.6.2	Parametric Boundary Methods	898
19.6.2.1	Sample Exercise Strategies for Bermudan Swaptions	898
19.6.2.2	Some Numerical Tests	901
19.6.2.3	Additional Comments	904
19.7	Other Topics	904

19.7.1	Robust Bermudan Swaption Hedging with European Swaptions	904
19.7.2	Carry and Exercise	907
19.7.3	Fast Pricing via Exercise Premia Representation	908
19.A	Appendix: Forward Volatility and Correlation	912
19.B	Appendix: A Primer on Moment Matching.....	913
19.B.1	Basics	913
19.B.2	Example 1: Asian Option in BSM Model	914
19.B.3	Example 2: Basket Option in BSM Model	916
20	TARNs, Volatility Swaps, and Other Derivatives	919
20.1	TARNs	919
20.1.1	Definitions and Examples	919
20.1.2	Valuation and Risk with Globally Calibrated Models	921
20.1.3	Local Projection Method	922
20.1.4	Volatility Smile Effects	923
20.1.5	PDE for TARNs.....	925
20.2	Volatility Swaps	927
20.2.1	Local Projection Method	928
20.2.2	Shout Options	929
20.2.3	Min-Max Volatility Swaps	932
20.2.4	Impact of Volatility Dynamics on Volatility Swaps	934
20.3	Forward Swaption Straddles	939
21	Out-of-Model Adjustments	945
21.1	Adjusting the Model	946
21.1.1	Calibration to Coupons	946
21.1.2	Adjusters	948
21.1.3	Path Re-Weighting	950
21.1.4	Proxy Model Method	955
21.1.5	Asset-Based Adjustments	957
21.1.6	Mapping Function Adjustments	959
21.2	Adjusting the Market	959
21.3	Adjusting the Trade	960
21.3.1	Fee Adjustments	961
21.3.2	Fee Adjustment Impact on Exotic Derivatives	962
21.3.3	Strike Adjustment	963

22	Introduction to Risk Management	969
22.1	Risk Management and Sensitivity Computations	970
22.1.1	Basic Information Flow	970
22.1.2	Risk: Theory and Practice	972
22.1.3	Example: the Black-Scholes Model	974
22.1.4	Example: Black-Scholes Model with Time-Dependent Parameters	977
22.1.5	Actual Risk Computations	979
22.1.6	What about Θ_{prm} and Θ_{num} ?	980
22.1.7	A Note on Trading P&L and the Computation of Implied Volatility	981
22.2	P&L Analysis	984
22.2.1	P&L Predict	985
22.2.2	P&L Explain	987
22.2.2.1	Waterfall Explain	987
22.2.2.2	Bump-and-Reset Explain	988
22.3	Value-at-Risk	989
22.A	Appendix: Alternative Proof of Lemma 22.1.1	992
23	Payoff Smoothing and Related Methods	995
23.1	Issues with Discretization Schemes	995
23.1.1	Problems with Grid Dimensioning	996
23.1.2	Grid Shifts Relative to Payout	996
23.1.3	Additional Comments	999
23.2	Basic Techniques	1000
23.2.1	Adaptive Integration	1000
23.2.2	Adding Singularities to the Grid	1001
23.2.3	Singularity Removal	1003
23.2.4	Partial Analytical Integration	1004
23.3	Payoff Smoothing For Numerical Integration and PDEs	1006
23.3.1	Introduction to Payoff Smoothing	1006
23.3.2	Payoff Smoothing in One Dimension	1008
23.3.2.1	Box Smoothing	1009
23.3.2.2	Other Smoothing Methods	1012
23.3.3	Payoff Smoothing in Multiple Dimensions	1013
23.4	Payoff Smoothing for Monte Carlo	1016
23.4.1	Tube Monte Carlo for Digital Options	1016
23.4.2	Tube Monte Carlo for Barrier Options	1018
23.4.3	Tube Monte Carlo for Callable Libor Exotics	1023
23.4.4	Tube Monte Carlo for TARNs	1023
23.A	Appendix: Delta Continuity of Singularity-Enlarged Grid Method	1024
23.B	Appendix: Conditional Independence for Tube Monte Carlo	1026

24 Pathwise Differentiation	1029
24.1 Pathwise Differentiation: Foundations	1029
24.1.1 Callable Libor Exotics	1029
24.1.1.1 CLE Greeks	1030
24.1.1.2 Keeping the Exercise Time Constant	1032
24.1.1.3 Noise in CLE Greeks	1033
24.1.2 Barrier Options	1034
24.2 Pathwise Differentiation for PDE Based Models	1038
24.2.1 Model and Setup	1038
24.2.2 Bucketed Deltas	1039
24.2.3 Survival Density	1042
24.3 Pathwise Differentiation for Monte Carlo Based Models	1045
24.3.1 Pathwise Derivatives of Forward Libor Rates	1045
24.3.2 Pathwise Deltas of European Options	1048
24.3.2.1 Pathwise Deltas of the Numeraire	1048
24.3.2.2 Pathwise Deltas of the Payoff	1049
24.3.3 Adjoint Method For Greeks Calculation	1050
24.3.4 Pathwise Delta Approximation for Callable Libor Exotics	1052
24.4 Notes on Likelihood Ratio and Hybrid Methods	1054
25 Importance Sampling and Control Variates	1057
25.1 Importance Sampling In Short Rate Models	1057
25.2 Payoff Smoothing by Importance Sampling	1059
25.2.1 Binary Options	1059
25.2.2 TARNs	1062
25.2.3 Removing the First Digital	1062
25.2.4 Smoothing All Digitals by One-Step Survival Conditioning	1063
25.2.5 Simulating Under the Survival Measure Using Conditional Gaussian Draws	1066
25.2.6 Generalized Trigger Products in Multi-Factor LM Models	1068
25.3 Model-Based Control Variates	1071
25.3.1 Low-Dimensional Markov Approximation for LM models	1072
25.3.2 Two-Dimensional Extension	1075
25.3.3 Approximating Volatility Structure	1076
25.3.4 Markov Approximation as a Control Variate	1078
25.4 Instrument-Based Control Variates	1080
25.5 Dynamic Control Variates	1084
25.6 Control Variates and Risk Stability	1087

26 Vegas in Libor Market Models	1089
26.1 Basic Problem of Vega Computations	1089
26.2 Review of Calibration	1091
26.3 Vega Calculation Methods	1092
26.3.1 Direct Vega Calculations	1092
26.3.1.1 Definition and Analysis	1092
26.3.1.2 Numerical Example	1095
26.3.2 What is a Good Vega?	1096
26.3.3 Indirect Vega Calculations	1099
26.3.3.1 Definition and Analysis	1099
26.3.3.2 Numerical Example and Performance Analysis	1102
26.3.4 Hybrid Vega Calculations	1105
26.3.4.1 Definition and Analysis	1105
26.3.4.2 Numerical Example	1107
26.4 Skew and Smile Vegas	1107
26.5 Vegas and Correlations	1109
26.5.1 Term Correlation Effects	1109
26.5.2 What Correlations should be Kept Constant?	1110
26.5.3 Vegas with Fixed Term Correlations	1112
26.5.4 Numerical Example	1113
26.6 Deltas with Backbone	1114
26.7 Vega Projections	1116
26.8 Some Notes on Computing Model Vegas	1118

Appendix

A Markovian Projection	1123
A.1 Marginal Distributions of Ito Processes	1123
A.2 Approximations for Conditional Expected Values	1128
A.2.1 Gaussian Approximation	1128
A.2.2 Least-Squares Projection	1130
A.3 Applications to Local Stochastic Volatility Models	1131
A.3.1 Markovian Projection onto an SV Model	1131
A.3.2 Fitting the Market with an LSV Model	1133
A.3.3 On Calculating Proxy Local Volatility	1137
A.4 Basket Options in Local Volatility Models	1139
A.5 Basket Options in Stochastic Volatility Models	1143
A.A Appendix: $E(\sqrt{z_n(t)z_m(t)})$ and $E(\sqrt{z_n(t)})$	1146
A.A.1 Proof of Proposition A.A.1	1147
A.A.1.1 Step 1. Reduction to Covariance	1147
A.A.1.2 Step 2. Linear Approximation	1148
A.A.1.3 Step 3. Coefficients	1148
A.A.1.4 Step 4. Order of Approximation	1149
A.A.2 Proof of Lemma A.A.2	1149

Introduction to Arbitrage Pricing Theory

For reference, this chapter reviews selected results from stochastic calculus and from the modern theory of asset pricing. The material in this chapter is well covered in existing literature, so we keep the chapter brief and the mathematical treatment informal. For a more rigorous treatment we refer to Duffie [2001] or Musiela and Rutkowski [1997]. Most of the necessary mathematical foundation for the theory is available in Karatzas and Shreve [1997], Øksendal [1992], and Protter [2005].

The treatment in this chapter focuses on asset pricing in general; we shall specialize it to interest rate securities in Chapter 4. Chapter 5 introduces fixed income markets in detail.

1.1 The Setup

Unless otherwise noted, in this book we shall always consider an economy with continuous and frictionless trading taking place inside a finite horizon $[0, T]$. We assume the existence of traded dividend-free assets with prices characterized by a p -dimensional vector-valued stochastic process $X(t) = (X_1(t), \dots, X_p(t))^\top$. Uncertainty and information arrival is modeled by a probability space (Ω, \mathcal{F}, P) , with Ω being a sample space with outcome elements ω ; \mathcal{F} being a σ -algebra on Ω ; and P being a probability measure on the measure space (Ω, \mathcal{F}) . Information is revealed over time according to a filtration $\{\mathcal{F}_t, t \in [0, T]\}$, a family of sub- σ -algebras of \mathcal{F} satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ whenever $s \leq t$. We can loosely think of \mathcal{F}_t as the information available at time t . We assume that the process $X(t)$ is adapted to $\{\mathcal{F}_t\}$, i.e. that $X(t)$ is fully observable at time t . For technical reasons, we require that the filtration satisfies the “usual conditions”¹. Let $E^P(\cdot)$ be the expectation

¹To satisfy the “usual conditions”, \mathcal{F}_t must be right-continuous for all t , and \mathcal{F}_0 must contain all the null-sets of \mathcal{F} , i.e. all subsets of sets of zero P -probability.

operator for the measure P ; when conditioning on information at time t , we will use the notation $E_t^P(\cdot) = E^P(\cdot | \mathcal{F}_t)$.

In all of the models in this book, we specialize the abstract setup above to the situation where information is generated by a d -dimensional vector-valued *Brownian motion* (or *Wiener process*) $W(t) = (W_1(t), \dots, W_d(t))^\top$, where W_i is independent of W_j for $i \neq j$. Brownian motions are treated in detail in Karatzas and Shreve [1997]; here, we just recall that a scalar Brownian motion W_i is a continuous stochastic process starting at 0 (i.e. $W_i(0) = 0$), having independent Gaussian increments: $W_i(t) - W_i(s) \sim \mathcal{N}(0, t-s)$, $t \geq s$. The filtration we consider is normally always the one *generated* by W , $\mathcal{F}_t = \sigma\{W(u), 0 \leq u \leq t\}$, possibly augmented to satisfy the usual conditions. We will generally assume that the price vector $X(t)$ is described by a vector-valued *Ito process*:

$$X(t) = X(0) + \int_0^t \mu(s, \omega) ds + \int_0^t \sigma(s, \omega) dW(s), \quad (1.1)$$

or, in differential notation,

$$dX(t) = \mu(t, \omega) dt + \sigma(t, \omega) dW(t), \quad (1.2)$$

where $\mu : \mathbb{R} \times \Omega \rightarrow \mathbb{R}^p$ and $\sigma : \mathbb{R} \times \Omega \rightarrow \mathbb{R}^{p \times d}$ are processes of dimension p and $p \times d$, respectively. We assume that both μ and σ are adapted to $\{\mathcal{F}_t\}$ and are in L^1 and L^2 respectively, in the sense that for all $t \in [0, T]$,

$$\int_0^t |\mu(s, \omega)| ds < \infty, \quad (1.3)$$

$$\int_0^t |\sigma(s, \omega)|^2 ds < \infty, \quad (1.4)$$

almost surely². In (1.4), we have defined

$$|\sigma(t, \omega)|^2 = \text{tr} (\sigma(t, \omega) \sigma(t, \omega)^\top). \quad (1.5)$$

We notice that the sample paths of X generated by (1.1) are almost surely continuous, with no jumps in asset prices.

A technical treatment of Ito processes and the Ito integral with respect to Brownian motion can be found in Karatzas and Shreve [1997]. For our needs, it suffices to think of the Ito integral as

$$\int_0^t \sigma(s, \omega) dW(s) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sigma((i-1)\delta, \omega) [W(i\delta) - W((i-1)\delta)], \quad (1.6)$$

²An event holds “almost surely” — often abbreviated by “a.s.” — if the probability of the event is one.

where $\delta \triangleq t/n$. We note that the integrand σ is here always evaluated at the *left* of each interval $[(i-1)\delta, i\delta]$. Other choices are possible³, but, as we shall see, the “non-anticipative” structure of the Ito integral gives rise to a number of useful results and makes it particularly useful as a model of trading gains (see Section 1.2).

We list a few relevant definitions and results below.

Definition 1.1.1 (Martingale). *Let $Y(t)$ be an adapted vector-valued process with $E^P(|Y(t)|) < \infty$ for all $t \in [0, T]$. We say that $Y(t)$ is a martingale under measure P if for all $s, t \in [0, T]$ with $t \leq s$,*

$$E_t^P(Y(s)) = Y(t), \quad a.s.$$

If we replace the equality sign in this equation with \leq or \geq , $Y(t)$ is said to be a *supermartingale* or a *submartingale*, respectively.

Definition 1.1.2 (Space H^2). *Let $|\sigma(t, \omega)|^2$ be as defined in (1.5). We say that σ is in H^2 , if for all $t \in [0, T]$ we have*

$$E^P \left(\int_0^t |\sigma(s, \omega)|^2 ds \right) < \infty.$$

The importance of Definition 1.1.2 becomes clear from the following result:

Theorem 1.1.3 (Properties of Ito Integral). *Define $I(t) = \int_0^t \sigma(s, \omega) dW(s)$ and assume that σ is in H^2 . Then*

1. $I(t)$ is \mathcal{F}_t -measurable.
2. $I(t)$ is a continuous martingale. In particular, $E^P(I(t)) = 0$ for all $t \in [0, T]$.
3. $E^P(|I(t)|^2) = E^P(\int_0^t |\sigma(s, \omega)|^2 ds) < \infty$.
4. $E^P(I(t)I(s)^\top) = E^P(\int_0^{\min(t,s)} \sigma(u, \omega) \sigma(u, \omega)^\top du)$.

A proof of Theorem 1.1.3 can be found in, e.g., Karatzas and Shreve [1997]. The equality in the third item of Theorem 1.1.3 is known as the *Ito isometry*. Due to the inequality in the third item, we say that the martingale defined in the process is a *square-integrable martingale*.

While it is common in applied work to simply assume that Ito integrals are martingales, without technical regularity conditions on $\sigma(t, \omega)$ (such as the H^2 restriction in Theorem 1.1.3), we should note that Ito integrals involving general processes in L^2 can, in fact, only be guaranteed to be *local martingales*. A process X is said to be a local martingale if there exists a

³The Stratonovich stochastic integral evaluates σ at the mid-point of each interval.

sequence of stopping times⁴ $\{\tau_n\}_{n=1}^\infty$, with $\tau_n \rightarrow \infty$ as $n \rightarrow \infty$, such that $X(\min(t, \tau_n))$, $t \geq 0$, is a martingale for all n . In other words, all “driftless” Ito processes of the type

$$dY(t) = \sigma(t, \omega) dW(t) \quad (1.7)$$

are local martingales, but not necessarily martingales. Interestingly, a converse result holds as well; all local martingales adapted to the filtration generated by the Brownian motion W can be represented as Ito processes of the form (1.7):

Theorem 1.1.4 (Martingale Representation Theorem). *If Y is a local martingale adapted to the filtration generated by a Brownian motion W , then there exists a process σ such that (1.7) holds. If Y is a square-integrable martingale, then σ is in H^2 .*

The proof of Theorem 1.1.4 can be found in Karatzas and Shreve [1997].

In the manipulation of functionals of Ito processes, the key result is a famous result by K. Ito:

Theorem 1.1.5 (Ito's Lemma). *Let $f(t, x)$, $x = (x_1, \dots, x_p)^\top$, denote a continuous function, $f : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}$, with continuous partial derivatives $\partial f / \partial t = f_t$, $\partial f / \partial x_i = f_{x,i}$, $\partial^2 f / \partial x_i \partial x_j = f_{x,x,j}$. Let $X(t)$ be given by the Ito process (1.2) and define a scalar process $Y(t) = f(t, X(t))$. Then $Y(t)$ is an Ito process with stochastic differential*

$$\begin{aligned} dY(t) &= f_t(t, X(t)) dt + f_{x,t}(t, X(t)) \mu(t, \omega) dt + f_x(t, X(t)) \sigma(t, \omega) dW(t) \\ &\quad + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p f_{x,x,j}(t, X(t)) (\sigma(t, \omega) \sigma(t, \omega)^\top)_{i,j} dt, \end{aligned}$$

where $f_x = (f_{x,1}, \dots, f_{x,p})$.

For easy reference, the result below lists Ito's lemma for the special case where $p = d = 1$.

Corollary 1.1.6. *For the case $p = d = 1$, Ito's lemma becomes*

$$\begin{aligned} dY(t) &= \left(f_t(t, X(t)) + f_x(t, X(t)) \mu(t, \omega) + \frac{1}{2} f_{xx}(t, X(t)) \sigma(t, \omega)^2 \right) dt \\ &\quad + f_x(t, X(t)) \sigma(t, \omega) dW(t). \end{aligned}$$

Ito's lemma can be motivated heuristically from a Taylor expansion. For instance, for the scalar case in Corollary 1.1.6, we write informally

⁴Recall that a stopping time τ is simply a random time adapted to the given filtration, in the sense that the event $\{\tau \leq t\}$ belongs to \mathcal{F}_t .

$$f(t + dt, X(t + dt)) = f(t, X(t)) + f_t dt + f_x dX(t) + \frac{1}{2} f_{xx} (dX(t))^2 + \dots \quad (1.8)$$

Here, we have

$$(dX(t))^2 = \mu(t, \omega)^2 (dt)^2 + \sigma(t, \omega)^2 (dW(t))^2 + 2\mu(t, \omega)\sigma(t, \omega) dt dW(t).$$

As shown earlier, $(dW(t))^2 = dt$ in quadratic mean, whereas all other terms in the expression for $(dX(t))^2$ are of order $O(dt^{3/2})$ or higher and can be neglected for small dt . In the limit, we therefore have $(dX(t))^2 = \sigma(t, \omega)^2 dt$ which can be inserted into (1.8). The result in Corollary 1.1.6 then emerges.

Remark 1.1.7. The quantity $(dX(t))^2$ discussed above is the differential of the *quadratic variation* of $X(t)$, often denoted by $\langle X(t), X(t) \rangle$. That is,

$$d\langle X(t), X(t) \rangle = (dX(t))^2 \Rightarrow \langle X(t), X(t) \rangle = \int_0^t (dX(u))^2.$$

For two different (scalar) Ito processes $X(t)$ and $Y(t)$, we may equivalently define the *quadratic covariation* process $\langle X(t), Y(t) \rangle$ by

$$d\langle X(t), Y(t) \rangle = dX(t) dY(t).$$

Sometimes we also write $d\langle X(t), Y(t) \rangle = \langle dX(t), dY(t) \rangle$. If $X(t)$ is a p -dimensional process and $Y(t)$ is a q -dimensional process, the quadratic covariation $\langle X(t), Y(t)^\top \rangle$ is a $(p \times q)$ -dimensional matrix process whose (i, j) -th element is $\langle X_i(t), Y_j(t) \rangle$, $i = 1, \dots, p$, $j = 1, \dots, q$.

The so-called *Tanaka extension* (see Karatzas and Shreve [1997]) extends Ito's lemma to continuous but non-differentiable functions. At points where the function has a kink, the Tanaka extension (loosely speaking) justifies using the Heaviside (step-) function for the first-order derivative and the Dirac delta function for the second-order derivative. An application of the Tanaka extension can be found in Section 1.9.2 and in Chapter 7, along with further discussion and references.

1.2 Trading Gains and Arbitrage

Working in the setting of Section 1.1 with assets driven by Ito processes, we now consider an investor engaging in a trading strategy involving the p assets X_1, \dots, X_p . Let the trading strategy be characterized by a predictable⁵ adapted process $\phi(t, \omega) = (\phi_1(t, \omega), \dots, \phi_p(t, \omega))^\top$, with $\phi_i(t, \omega)$ denoting

⁵A *predictable* process is one where we, loosely speaking, can “foretell” the value of the process at time t , given all information available up to, but not including, time t . All adapted continuous processes are thus predictable. For a technical definition of predictable processes, see Karatzas and Shreve [1997].

the holdings at time t in the i -th asset X_i . The value $\pi(t)$ of the trading strategy at time t is thus (dropping the dependence on ω in the notation)

$$\pi(t) = \phi(t)^\top X(t). \quad (1.9)$$

The gain from trading over a small time interval $[t, t + \delta]$ is (approximately) $\phi(t)^\top [X(t + \delta) - X(t)]$, suggesting (compare to (1.6)) that the Ito integral

$$\int_0^t \phi(s)^\top dX(s) = \int_0^t \phi(s)^\top \mu(s) ds + \int_0^t \phi(s)^\top \sigma(s) dW(s)$$

is a proper model for trading gains over $[0, t]$. An investment strategy is said to be *self-financing* if, for any $t \in [0, T]$,

$$\pi(t) - \pi(0) = \int_0^t \phi(s)^\top dX(s). \quad (1.10)$$

This relationship simply expresses that changes in portfolio value are solely caused by trading gains or losses, with no funds being added or withdrawn.

Self-financing trading strategies allow investors to turn a certain initial investment $\pi(0)$ into stochastic future wealth $\pi(t)$. Under natural assumptions on possible trading strategies (e.g., that there is finite supply of all assets) we would expect that there should be limitations to the profits that self-financing strategies can create. Most notably, it should be impossible to create “something for nothing”, that is, to turn a zero initial investment into future wealth that is certain to be non-negative and may be positive with non-zero probability. To express this formally, we introduce the concept of an *arbitrage opportunity*:

Definition 1.2.1 (Arbitrage). *An arbitrage opportunity is a self-financing strategy ϕ for which $\pi(0) = 0$ and, for some $t \in [0, T]$,*

$$\pi(t) \geq 0 \text{ a.s., and } P(\pi(t) > 0) > 0, \quad (1.11)$$

with π given in (1.9).

In economic equilibrium, arbitrage strategies cannot exist and precluding (1.11) constitutes a fundamental consistency requirement on the asset processes.

1.3 Equivalent Martingale Measures and Arbitrage

We turn to the question of characterizing the conditions under which the trading economy is free of arbitrage opportunities. A concise way to state these conditions involves *equivalent martingale measures*, a concept we shall work our way up to in a number of steps. First, we recall that two

probability measures P and \widehat{P} on the same measure space (Ω, \mathcal{F}) are said to be *equivalent* if $P(A) = 0 \Leftrightarrow \widehat{P}(A) = 0, \forall A \in \mathcal{F}$; that is, the two measures have the same null-sets. An important result from measure theory states that equivalent measures are uniquely associated through a quantity known as a *Radon-Nikodym derivative*:

Theorem 1.3.1 (Radon-Nikodym Theorem). *Let P and \widehat{P} be equivalent probability measures on the common measure space (Ω, \mathcal{F}) . There exists a unique (a.s.) non-negative random variable R with $E^P(R) = 1$, such that*

$$\widehat{P}(A) = E^P(R1_{\{A\}}), \quad \text{for all } A \in \mathcal{F}.$$

For a proof of Theorem 1.3.1, see e.g. Billingsley [1995]. The random variable R in the theorem is known as a *Radon-Nikodym derivative* and is denoted $d\widehat{P}/dP$. In the theorem we have used an *indicator* $1_{\{A\}}$; this quantity is 1 if the event A comes true, 0 if not.

For later use, we associate any probability measure \widehat{P} with a *density process*

$$\varsigma(t) = E_t^P \left(\frac{d\widehat{P}}{dP} \right), \quad \forall t \in [0, T]. \quad (1.12)$$

Clearly, $\varsigma(t)$ is a P -martingale with $\varsigma(0) = 1$ and $\varsigma(t) = E_t^P(\varsigma(T))$. A simple conditioning exercise demonstrates that for any \mathcal{F}_T -measurable random variable $Y(T)$, with $R = d\widehat{P}/dP$,

$$\begin{aligned} E^{\widehat{P}}(Y(T)|\mathcal{F}_t) &= \frac{1}{E^P(R|\mathcal{F}_t)} E^P(RY(T)|\mathcal{F}_t) \\ &= \varsigma(t)^{-1} E^P(E^P(R|\mathcal{F}_T)Y(T)|\mathcal{F}_t) \\ &= E^P \left(Y(T) \frac{\varsigma(T)}{\varsigma(t)} \middle| \mathcal{F}_t \right). \end{aligned} \quad (1.13)$$

We shall use this result on numerous occasions in this book.

We now introduce the important concept of a *deflator*, a strictly positive Ito process used to normalize the asset prices. Let the deflator be denoted $D(t)$, and define the normalized asset process $X^D(t) = (X_1(t)/D(t), \dots, X_p(t)/D(t))^{\top}$. We say that a measure Q^D is an *equivalent martingale measure induced by D* if $X^D(t)$ is a martingale with respect to Q^D . If Q^D is a martingale measure, we say that a self-financing trading strategy is *permissible* if

$$\int_0^t \phi(s)^{\top} dX^D(s)$$

is a martingale. For the Ito setup discussed earlier, a permissible strategy⁶ is obtained by, say, requiring that $\phi(t)^{\top} \sigma(t)$ is in H^2 ; see Theorem 1.1.3. An

⁶The technical restriction on trading positions imposed by only considering permissible trading strategies rules out certain pathological strategies, such as the

application of Ito's lemma combined with (1.9)–(1.10) implies that $\pi(t)/D(t)$ is a Q^D -martingale when the trading strategy is permissible.

For permissible trading strategies, the importance of equivalent martingale measures follows from the following theorem:

Theorem 1.3.2 (Sufficient Condition for No-Arbitrage). *Restrict attention to permissible trading strategies. If there is a deflator D such that the deflated asset price process allows for an equivalent martingale measure, then there is no arbitrage.*

For a proof we refer to Musiela and Rutkowski [1997]. We note that Theorem 1.3.2 only provides sufficient conditions for the absence of arbitrage, and known (and rather technical) counterexamples demonstrate that the existence of an equivalent martingale measure does not follow from the absence of arbitrage in a setting with permissible trading strategies. A body of results known as the *fundamental theorem of arbitrage* establishes the conditions under which the existence of an equivalent martingale measure is also a necessary condition for the absence of arbitrage. The results are rather technical, but generally state that absence of arbitrage and the existence of an equivalent martingale measure are “nearly” equivalent concepts. The exact notion of “nearly” equivalent is discussed in Duffie [2001] as well as in the authoritative reference⁷ Delbaen and Schachermayer [1994]. For our purposes in this book, we ignore many of these technicalities and often simply treat the absence of arbitrage and the existence of a martingale measure as equivalent concepts.

Finally, if the deflator is one of the p assets, we call the deflator a *numeraire*. Let us, say, assume that X_1 is strictly positive and can be used as a numeraire. Also assume that a deflator D has been identified such that Theorem 1.3.2 holds. As $X_1(t)/D(t)$ is a Q^D -martingale, we can use the Radon-Nikodym theorem to define a new measure Q^{X_1} by the density $\varsigma(t) = (X_1(t)/D(t))/(X_1(0)/D(0))$. For an \mathcal{F}_T -measurable variable $Y(T)$, we then have, from (1.13),

$$X_1(t)E_t^{Q^{X_1}}\left(\frac{Y(T)}{X_1(T)}\right) = D(t)E_t^{Q^D}\left(\frac{Y(T)}{D(T)}\right). \quad (1.14)$$

In particular, if $Y(t)/D(t)$ is a Q^D -martingale, $Y(t)/X_1(t)$ must also be a Q^{X_1} -martingale. In practice, it normally suffices to only consider deflators from the set of available numeraires.

Remark 1.3.3. Some sources define $1/D(t)$ (rather than $D(t)$) as the deflator. The convention used in this book is more natural for our applications.

doubling strategy considered in Harrison and Kreps [1979]. A realistic resource-constrained economy will always bound the size of the positions one can take in an asset, sufficing to ensure that predictable trading strategies are permissible.

⁷In a nutshell, Delbaen and Schachermayer [1994] show that absence of arbitrage implies only the existence of a *local* martingale measure.

1.4 Derivative Security Pricing and Complete Markets

A T -maturity *derivative security* (also known as a *contingent claim*) pays out at time T an \mathcal{F}_T -measurable random variable $V(T)$, and makes no payments before T . We assume that $V(T)$ has finite variance, and say that the derivative security is *attainable* (or sometimes *redundant*) if there exists a permissible trading strategy ϕ such that $V(T) = \phi(T)^\top X(T) = \pi(T)$ a.s. The trading strategy is said to *replicate* the derivative security. Importantly, the absence of arbitrage dictates that the time 0 price of an attainable derivative security $V(0)$ must be equal to the cost of setting up the self-financing strategy, i.e. $V(0) = \pi(0)$. More generally, $V(t) = \pi(t)$, $t \in [0, T]$. This observation is the foundation of *arbitrage pricing* and allows us to price derivative securities as expectations under an equivalent martingale measure. Specifically, consider a deflator D and assume the existence of an equivalent martingale measure Q^D induced by D ; the existence of Q^D guarantees that there are no arbitrages in the market, by Theorem 1.3.2. Now, from the martingale property of $\pi(t)/D(t)$ in the measure Q^D and the relation $V(t) = \pi(t)$ it immediately follows that

$$\frac{V(t)}{D(t)} = E_t^{Q^D} \left(\frac{V(T)}{D(T)} \right)$$

or

$$V(t) = D(t) E_t^{Q^D} \left(\frac{V(T)}{D(T)} \right). \quad (1.15)$$

If all finite-variance \mathcal{F}_T -measurable random variables can be replicated, the market is said to be *complete*. In a complete market, all derivatives are “spanned” and hence have unique prices. Interestingly, a similar uniqueness result holds for equivalent martingale measures:

Theorem 1.4.1. *In the absence of arbitrage, a market is complete if and only if there exists a deflator inducing a unique martingale measure.*

From (1.14) it follows that the martingale measures induced by all numeraires must then be unique as well.

In practical applications, we shall often manipulate the choice of numeraire asset to simplify computations. The following result is useful for this:

Theorem 1.4.2 (Change of Numeraire). *Consider two numeraires $N(t)$ and $M(t)$, inducing equivalent martingale measures Q^N and Q^M , respectively. If the market is complete, then the density of the Radon-Nikodym derivative relating the two measures is uniquely given by*

$$\varsigma(t) = E_t^{Q^N} \left(\frac{dQ^M}{dQ^N} \right) = \frac{M(t)/M(0)}{N(t)/N(0)}.$$

Proof. As the market is complete, all derivatives prices are unique. Consider an integrable \mathcal{F}_T -measurable payout $V(T) = Y(T)M(T)$, with time t price $V(t)$. From Theorem 1.4.1 and (1.15) we must have

$$V(t) = N(t)\mathbb{E}_t^{\mathbb{Q}^N} \left(\frac{M(T)Y(T)}{N(T)} \right) = M(t)\mathbb{E}_t^{\mathbb{Q}^M} \left(\frac{M(T)Y(T)}{M(T)} \right)$$

or

$$\mathbb{E}_t^{\mathbb{Q}^M} (Y(T)) = \mathbb{E}_t^{\mathbb{Q}^N} \left(Y(T) \frac{M(T)/N(T)}{M(t)/N(t)} \right).$$

Comparison with (1.13), and the fact that the density must be scaled to equal 1 at time 0, reveals that the Radon-Nikodym derivative for the measure shift is characterized by the density in the theorem. \square

1.5 Girsanov's Theorem

The last two sections have demonstrated a close link between the concept of arbitrage and the existence and uniqueness of equivalent martingale measures. In this section, we consider i) the conditions on the asset prices that allow for an equivalent martingale measure; and ii) the effect on asset dynamics from a change of probability measure. We consider two measures P and $P(\theta)$ related by a density $\varsigma^\theta(t) = \mathbb{E}_t^P(dP(\theta)/dP)$, where $\varsigma^\theta(t)$ is an *exponential martingale* given by the Ito process

$$d\varsigma^\theta(t)/\varsigma^\theta(t) = -\theta(t)^\top dW(t),$$

where $W(t)$ is a d -dimensional P -Brownian motion. The d -dimensional process θ is known as the *market price of risk*. By an application of Ito's lemma, we can write

$$\begin{aligned} \varsigma^\theta(t) &= \exp \left(- \int_0^t \theta(s)^\top dW(s) - \frac{1}{2} \int_0^t \theta(s)^\top \theta(s) ds \right) \\ &\triangleq \mathcal{E} \left(- \int_0^t \theta(s)^\top dW(s) \right) \end{aligned} \tag{1.16}$$

where $\mathcal{E}(\cdot)$ is the *Doleans exponential*. An often-quoted sufficient condition on $\theta(t)$ for (1.16) to define a proper martingale (and not just a local martingale) is the *Novikov condition*

$$\mathbb{E}^P \left[\exp \left(\frac{1}{2} \int_0^t \theta(s)^\top \theta(s) ds \right) \right] < \infty. \tag{1.17}$$

The Novikov condition can often be difficult to verify in practical applications.

Armed with the notation above, we are now ready to state the main result of this section.

Theorem 1.5.1 (Girsanov's Theorem). Suppose that $\varsigma^\theta(t)$ defined in (1.16) is a martingale. Then for all $t \in [0, T]$

$$W^\theta(t) = W(t) + \int_0^t \theta(s) ds$$

is a Brownian motion under the measure $P(\theta)$.

To discuss a strategy to prove Girsanov's theorem, assume for simplicity that the dimension of the Brownian motion is $d = 1$. One way to construct a proof for Theorem 1.5.1 is to demonstrate that the joint moment-generating function (mgf)⁸ (under $P(\theta)$) of the increments

$$W^\theta(t_1), W^\theta(t_2) - W^\theta(t_1), \dots, W^\theta(t_n) - W^\theta(t_{n-1}), \quad 0 < t_1 < \dots < t_n,$$

is the same as that of n independent Gaussian random variables with expectations 0 and variances $t_1, t_2 - t_1, \dots$. That is, for any positive integer value of n and any set of values $\alpha_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, we need to show that,

$$E^{P^\theta} \left[\exp \left(\sum_{i=1}^n \alpha_i (W^\theta(t_i) - W^\theta(t_{i-1})) \right) \right] = \prod_{i=1}^n \exp (\alpha_i^2 (t_i - t_{i-1}) / 2),$$

where we have defined $t_0 = 0$. While carrying out such a proof is not difficult, we here merely justify the final result by examining the case $n = 1$ only. Specifically, we consider

$$E^{P(\theta)} [\exp (\alpha W^\theta(t))],$$

where $\alpha \in \mathbb{R}$ and $t > 0$. Shifting probability measure, we get

$$\begin{aligned} E^{P(\theta)} [\exp (\alpha W^\theta(t))] &= E^{P(\theta)} \left[\exp \left(\alpha W(t) + \alpha \int_0^t \theta(s) ds \right) \right] \\ &= E^P \left[\exp \left(\alpha W(t) + \alpha \int_0^t \theta(s) ds \right) \mathcal{E} \left(- \int_0^t \theta(s) dW(s) \right) \right] \\ &= e^{\alpha^2 t / 2} E^P \left[\exp \left(\int_0^t (\alpha - \theta(s)) dW(s) - \frac{1}{2} \int_0^t (\alpha - \theta(s))^2 ds \right) \right] \\ &= e^{\alpha^2 t / 2} E^P \left[\mathcal{E} \left(\int_0^t (\alpha - \theta(s)) dW(s) \right) \right] \\ &= e^{\alpha^2 t / 2}, \end{aligned}$$

⁸Recall that the moment-generating function of a random variable Y in some measure P is defined as the expectation $E^P(\exp(\alpha Y))$, $\alpha \in \mathbb{R}$. Unlike the characteristic function, the moment-generating function is not always well-defined for all values of the argument α .

as desired. In the last step, we used the fact that the Doleans exponential is a martingale with initial value 1.

Girsanov's theorem implies that we can shift probability measure to transform an Ito process with a given drift to an Ito process with nearly arbitrary drift. Specifically, we notice that our asset price process (under P)

$$dX(t) = \mu(t) dt + \sigma(t) dW(t)$$

can be written

$$dX(t) = (\mu(t) - \sigma(t)\theta(t)) dt + \sigma(t)dW^\theta(t),$$

where $W^\theta(t)$ is a Brownian measure under the measure $P(\theta)$. This process will be driftless provided that θ satisfies the “spanning condition” $\mu(t) = \sigma(t)\theta(t)$ for all $t \in [0, T]$. This gives us a convenient way to check for the existence of equivalent martingale measures:

Corollary 1.5.2. *For a given numeraire D , assume that the deflated asset process satisfies*

$$dX^D(t) = \mu^D(t) dt + \sigma^D(t) dW(t),$$

where $\sigma^D(t)$ is sufficiently regular to make $\int_0^t \sigma^D(s) dW(s)$ a martingale. Assume also that there exists a θ such that the density ς^θ is a martingale and (a.s.)

$$\sigma^D(t)\theta(t) = \mu^D(t), \quad t \in [0, T], \quad (1.18)$$

then D induces an equivalent martingale measure and there is no arbitrage.

Equation (1.18) is a system of linear equations and we can use rank results from linear algebra to determine the circumstances under which (1.18) will have solutions (no arbitrage) and when these are unique (complete market). For instance, a necessary condition for the market to be complete is that $\text{rank}(\sigma) = d$. Further results along these lines can be found in Musiela and Rutkowski [1997] and Duffie [2001].

We conclude this section by noting that while a change of probability measure affects the drift μ of an Ito process, it does not change the diffusion coefficient σ . This is sometimes known as the *diffusion invariance principle*.

1.6 Stochastic Differential Equations

So far we have defined the asset process vector to be an Ito process with general measurable coefficients $\mu(t, \omega)$ and $\sigma(t, \omega)$. In virtually all applications, however, we restrict our attention to the case where these coefficients

are deterministic functions of time and the state of the asset process⁹. In other words, we consider a *stochastic differential equation* (SDE) of the form

$$dX(t) = \mu(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad X(0) = X_0, \quad (1.19)$$

with $\mu : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}^p$; $\sigma : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}^{p \times d}$; and X_0 an initial condition. A *strong solution*¹⁰ to (1.19) is an Ito process

$$X(t) = X_0 + \int_0^t \mu(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW(s).$$

A number of restrictions on μ and σ are needed to ensure that the solution to (1.19) exists and is unique. A standard result is listed below.

Theorem 1.6.1. *In (1.19) assume that there exists a constant K such that for all $t \in [0, T]$ and all $x, y \in \mathbb{R}^p$,*

$$\begin{aligned} |\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| &\leq K|x - y|, & (\text{Lipschitz condition}), \\ |\mu(t, x)|^2 + |\sigma(t, x)|^2 &\leq K^2(1 + |x|^2), & (\text{growth condition}). \end{aligned}$$

Then there exists a unique solution to (1.19).

We notice that the dynamics of (1.19) do not depend on the past evolution of $X(t)$ beyond the state of X at time t . This lack of path-dependence suggests that X is a *Markov* process. We formalize this as follows.

Definition 1.6.2 (Markov Process). *The \mathbb{R}^p -valued stochastic process $X(t)$ is called a *Markov process* if for all $s, t \in [0, T]$ with $t \leq s$,*

$$\mathbb{P}(X(s) \in B | \mathcal{F}_t) = \mathbb{P}(X(s) \in B | X(t)) \quad (1.20)$$

for all sets B in the p -dimensional σ -algebra of Borel set \mathfrak{B}^p . If (1.20) holds with s replaced by a stopping time, the process is a strong Markov process.

Expressed verbally, the Markov property implies that the past and future become statistically independent when we condition on the present.

Theorem 1.6.3 (Markov Property of SDEs). *Let the coefficient of the SDE for $X(t)$ satisfy the conditions in Theorem 1.6.1. Then $X(t)$ is a strong Markov process.*

⁹In this section, the process X is generic and need not represent financial assets.

¹⁰In a strong solution, the Brownian motion is given and the solution is adapted to the filtration generated by it. If we are free to pick our own Brownian motion on some different probability space, we say that (1.19) holds in a *weak sense*. For financial applications where we normally only need the law of the underlying process, weak solutions are typically sufficient. The distinction between weak and strong solutions is of little importance for our purposes and we shall ignore it going forward.

Let us consider the explicit solutions of a few simple SDEs. First, consider a *linear* SDE

$$dX(t) = (AX(t) + B(t)) dt + C(t) dW(t),$$

where A is a constant $p \times p$ matrix, and B and C are deterministic matrices of dimension $p \times 1$ and $p \times d$, respectively. The solution to this equation can, by Ito's lemma, be verified to be

$$X(t) = e^{At} X(0) + \int_0^t e^{A(t-s)} (B(s) ds + C(s) dW(s)).$$

The term

$$\int_0^t e^{A(t-s)} C(s) dW(s)$$

is distributed as a p -dimensional Gaussian random variable with mean 0 and, from Theorem 1.1.3, covariance matrix

$$\Sigma = \int_0^t e^{A(t-s)} C(s) C(s)^\top e^{A^\top(t-s)} ds.$$

Extensions to time-varying A are straightforward, and basically involve replacing the exponential matrix e^{At} with the solution of a homogeneous ODE with time-dependent coefficients. Details can be found in, e.g., Arnold [1974] and key results are listed in Chapter 12.

Now let us specialize to the scalar case with $p = 1$. An SDE of great importance is the *geometric Brownian motion with drift* (GBMD):

$$dX(t)/X(t) = \mu(t) dt + \sigma(t) dW(t),$$

where $\mu(t)$ and $\sigma(t)$ are *deterministic* (with $\sigma(t)$ having dimension $1 \times d$). An application of Ito's lemma to $\ln(X(t))$ reveals that

$$\begin{aligned} X(t) &= X(0) \exp \left(\int_0^t \left(\mu(s) - \frac{1}{2} \sigma(s) \sigma(s)^\top \right) ds + \int_0^t \sigma(s) dW(s) \right) \\ &= X(0) \exp \left(\int_0^t \mu(s) ds \right) \mathcal{E} \left(\int_0^t \sigma(s) dW(s) \right). \end{aligned} \quad (1.21)$$

Being an exponential of a Gaussian random variable, $X(t)$ follows a *log-normal* distribution, with moments (see Karatzas and Shreve [1997])

$$\mathbb{E}^P(X(t)) = X(0) \exp \left(\int_0^t \mu(s) ds \right), \quad (1.22)$$

$$\mathbb{E}^P(X(t)^2) = \mathbb{E}^P(X(t))^2 \exp \left(\int_0^t \sigma(s) \sigma(s)^\top ds \right). \quad (1.23)$$

1.7 Explicit Trading Strategies and PDEs

After the mathematical interlude of Section 1.6, we now return to financial markets and a more careful analysis of the trading strategies that replicate derivative securities. We have already established that in a complete market such strategies must exist for any given derivative, but it still remains to determine these strategies explicitly. Consider a Markovian setup where the asset vector X satisfies an SDE of the form (1.19). Let there be given a derivative security V paying out at time T an amount $V(T) = g(X(T))$, for some smooth payout function $g : \mathbb{R}^p \rightarrow \mathbb{R}$. The Markovian form of the asset dynamics suggests that the time t derivative price is a function of t and $X(t)$ only, $V(t) = V(t, X(t))$ for some deterministic function $V(t, x)$, $x \in \mathbb{R}^p$. Conjecturing that this function is smooth enough to allow for an application of Ito's lemma for all $t \in [0, T]$, Theorem 1.1.5 implies (suppressing dependence on $X(t)$ for brevity)

$$\begin{aligned} dV(t) &= V_t(t) dt + \sum_{i=1}^p V_{x_i}(t) \mu_i(t) dt \\ &\quad + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p V_{x_i x_j}(t) \Sigma_{i,j}(t) dt + \sum_{i=1}^p V_{x_i}(t) \sigma_i(t) dW(t), \end{aligned} \quad (1.24)$$

where σ_i is the i -th row of the $p \times d$ matrix σ and $\Sigma_{i,j}$ is the (i, j) -th element in $\sigma \sigma^\top$. We recall that subscripts like V_X , denote partial differentiation, see Theorem 1.1.5.

If $V(t)$ can be replicated by a self-financing trading strategy ϕ in the p assets, we must also have, from (1.10),

$$dV(t) = \phi(t)^\top dX(t) = \sum_{i=1}^p \phi_i(t) \mu_i(t) dt + \sum_{i=1}^p \phi_i(t) \sigma_i(t) dW(t). \quad (1.25)$$

Comparing terms in (1.24) and (1.25) we see that both equations will hold, provided that for all $t \in [0, T]$

$$\phi_i(t) = \frac{\partial V(t, X(t))}{\partial x_i}, \quad i = 1, \dots, p, \quad (1.26)$$

and

$$\frac{\partial V(t, x)}{\partial t} + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \frac{\partial^2 V(t, x)}{\partial x_i \partial x_j} \Sigma_{i,j}(t, x) = 0. \quad (1.27)$$

To the extent that the system above allows for a solution (it may not if the market is not complete, from (1.26) we see that the trading strategy that replicates the derivative V holds $\partial V(t, X(t))/\partial x_i$ units of asset X_i at time

t. The quantity $\partial V / \partial x_i$ is often known as the *delta* with respect to X_i ¹¹. Note that, from (1.9) and (1.26) we have that

$$V(t, X(t)) = \sum_{i=1}^p \frac{\partial V(t, X(t))}{\partial x_i} X_i(t). \quad (1.28)$$

Besides identifying an explicit replication strategy, the arguments above have also produced (1.27), a partial differential equation (PDE) for the value function $V(t, x)$. The PDE is a second-order parabolic equation in p spatial variables, with known terminal condition $V(T, x) = g(x)$ (a so-called *Cauchy problem*). Solving this PDE provides an alternative way to price the derivative, as compared to the purely probabilistic expectations-based methods outlined earlier (see (1.15)). We shall investigate the link between expectations and PDEs in more detail in Section 1.8.

Inspection of the valuation PDE (1.27) reveals that the drifts μ_i of the asset price SDE (1.19) are notably absent, making the price of the derivative security independent of drifts. This is typical of derivatives in complete markets and follows from the fact that derivatives can be priced preference-free, by arbitrage arguments. In contrast, for the elements of the fundamental asset price vector, risk-averse investors would demand that assets with high volatilities $|\sigma_i|$ be rewarded with higher drifts (more precisely, higher rates of return) as compensation for the additional uncertainty.

1.8 Kolmogorov's Equations and the Feynman-Kac Theorem

In earlier sections, we have seen that derivatives prices can be expressed as expectations under certain probability measures or as solutions to PDEs. This hints at a deeper connection between expectations and PDEs, a connection we shall explore in this section. As part of this exploration, we list results for transition densities that will be useful later in model calibration.

As in Section 1.6, we consider a Markov vector SDE of the type (see (1.19))

$$dX(t) = \mu(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad X(0) = X_0, \quad (1.29)$$

where the coefficients are assumed smooth enough to allow for a unique solution (see Theorem 1.6.1). Now define a functional

$$u(t, x) = \mathbb{E}^P(g(X(T)) | X(t) = x),$$

¹¹Note that taking a position in V and following a trading strategy with $\phi_i = -\partial V / \partial x_i$, $i = 1, \dots, p$ will effectively remove any exposure to V (as we simultaneously take a long position in V and, through a trading strategy, a short position in V). This strategy is known as a *delta hedge*.

for a function $g : \mathbb{R}^p \rightarrow \mathbb{R}$. Under regularity conditions on g , it is easy to see that the process $u(t, X(t))$, being a conditional expectation, must be a martingale. Proceeding informally, an application of Ito's lemma gives, for $t \in [0, T)$ (suppressing dependence on $X(t)$),

$$du(t) = u_t(t) dt + \sum_{i=1}^p u_{x_i}(t) \mu_i(t) dt + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p u_{x_i x_j}(t) \Sigma_{i,j}(t) dt + O(dW(t)),$$

where as before $\Sigma_{i,j}$ is the (i, j) -th element of $\sigma\sigma^\top$. From earlier results, we know that for $u(t, X(t))$ to be a martingale, the term multiplying dt in the equation above must be zero. Defining the operator

$$\mathcal{A} = \sum_{i=1}^p \mu_i(t, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \Sigma_{i,j}(t, x) \frac{\partial^2}{\partial x_i \partial x_j},$$

we deduce that $u(t, x)$ satisfies the PDE

$$\frac{\partial u(t, x)}{\partial t} + \mathcal{A}u(t, x) = 0, \quad (1.30)$$

with terminal condition $u(T, x) = g(x)$. The equation above is known as the *Kolmogorov backward equation* for the SDE (1.29). The operator \mathcal{A} is known as the *generator* or *infinitesimal operator* of the SDE, and can be identified as

$$\mathcal{A}u(t, x) = \lim_{h \downarrow 0} \frac{\mathbb{E}^P(u(t+h, X(t+h)) | X(t) = x) - u(t, x)}{h}.$$

In arriving at (1.30) we made several implicit assumptions, most notably that the function $u(t, x)$ exists and is twice differentiable. Sufficient conditions for the validity of (1.30) can be found in Karatzas and Shreve [1997], for instance. A relevant result is listed below.

Theorem 1.8.1. *Let the process $X(t)$ be given by the SDE (1.29), where the coefficients μ and σ are continuous in x and satisfy the Lipschitz and growth conditions of Theorem 1.6.1. Consider a continuous function $g(x)$ that is either non-negative or satisfies a polynomial growth condition, meaning that for some positive constants K and q*

$$g(x) \leq K(1 + |x|^q), \quad x \in \mathbb{R}^p.$$

If $u(t, x)$ solves (1.30) with boundary condition $u(T, x) = g(x)$, and $u(t, x)$ satisfies a polynomial growth condition in x , then

$$u(t, x) = \mathbb{E}^P(g(X(T)) | X(t) = x), \quad t \in [0, T]. \quad (1.31)$$

Conditions required to ensure existence of a solution to (1.30) are more involved, and we just refer to Karatzas and Shreve [1997] and the references therein.

A family of functions g of particular importance to many of our applications is

$$g(x) = e^{ik^T x}, \quad k \in \mathbb{R}^p,$$

where $i = \sqrt{-1}$ is the imaginary unit. In this case $u(t, x)$ becomes the *characteristic function* of $X(T)$, conditional on $X(t) = x$. We refer to any standard statistics textbook (e.g. Ochi [1990]) for the many useful properties of characteristic functions.

For the Markov process $X(t)$ in (1.29), let us now introduce a *transition density*, given heuristically by

$$p(t, x; s, y) dy \triangleq \mathbb{P}(X(s) \in [y, y + dy] | X(t) = x), \quad 0 \leq t \leq s \leq T.$$

We can loosely think of the transition density as a special case of the functional $u(t, x)$ above, with boundary condition $u(s, x) = \delta(x - y)$, where $\delta(\cdot)$ is the Dirac delta function. Sometimes $p(\cdot, \cdot; \cdot, \cdot)$ is called a *Green's function* or a *fundamental solution* to (1.30). Under certain regularity conditions discussed in Karatzas and Shreve [1997], the transition density solves the Kolmogorov backward equation

$$\frac{\partial p(t, x)}{\partial t} + \mathcal{A}p(t, x) = 0, \quad (s, y) \text{ fixed},$$

subject to the boundary condition $p(s, x; s, y) = \delta(x - y)$. Further, the general expectation $u(t, x) = \mathbb{E}^P(g(X(T)) | X(t) = x)$ in Theorem 1.8.1 can be written

$$u(t, x) = \int_{\mathbb{R}^p} g(y)p(t, x; T, y) dy, \quad t \in [0, T]. \quad (1.32)$$

In many applications, it is useful to have a result that produces transition densities at future times $s \geq t$ from a known state at time t , rather than vice-versa. For this, we first define an operator \mathcal{A}^* by

$$\mathcal{A}^* f(s, y) = - \sum_{i=1}^p \frac{\partial [\mu_i(s, y) f(s, y)]}{\partial y_i} + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \frac{\partial^2 [\Sigma_{i,j}(s, y) f(s, y)]}{\partial y_i \partial y_j}.$$

In the transition density $p(t, x; s, y)$ now consider (t, x) fixed and let \mathcal{A}^* operate on the resulting function of s and y . Under additional regularity conditions, we then have the *forward Kolmogorov equation*

$$-\frac{\partial p(s, y)}{\partial s} + \mathcal{A}^* p(s, y) = 0, \quad (t, x) \text{ fixed}, \quad (1.33)$$

subject to the boundary condition $p(t, x; t, y) = \delta(x - y)$.

The forward Kolmogorov equation is sometimes known as the *Fokker-Planck* equation. We stress that the backward equation is more general than the forward equation, in the sense that the former holds for general terminal conditions $g(x)$, whereas the latter only holds for δ -type initial conditions.

We round off this section by a useful extension to the Kolmogorov backward equation. Specifically, consider extending the PDE (1.30) to

$$\frac{\partial u(t, x)}{\partial t} + \mathcal{A}u(t, x) + h(t, x) = r(t, x)u(t, x), \quad (1.34)$$

where $h, r : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}$. Given the boundary condition $u(T, x) = g(x)$, the *Feynman-Kac solution* to (1.34), should it exist, is given by

$$u(t, x) = \mathbb{E}^P \left(\psi(t, T)g(X(T)) + \int_t^T \psi(s, X(s))h(s, X(s)) ds \middle| X(t) = x \right), \quad (1.35)$$

where

$$\psi(t, T) = \exp \left(- \int_t^T r(s, X(s)) ds \right), \quad t \in [0, T].$$

The result is easily understood from an application of Ito's lemma, similar to the one used above to motivate the backward Kolmogorov equation. Sufficient regularity conditions for the Feynman-Kac result to hold are identical to those of Theorem 1.8.1, supplemented with the requirement that r be nonnegative and continuous in x ; and the requirement that h be continuous in x and either be nonnegative or satisfy a polynomial growth requirement in x . See Duffie [2001] for further details about the often delicate regularity issues surrounding the Feynman-Kac result.

For later use, let us finally note that when $g(x) = \delta(x - y)$ and $h(t, x) = 0$, $u(t, x)$ in (1.35) will equal

$$G(t, x; T, y) \triangleq \mathbb{E}^P \left(e^{-\int_t^T r(s, X(s)) ds} \delta(X(T) - y) | X(t) = x \right).$$

The function G is known as a *state-price density* or as an *Arrow-Debreu security price* function. In particular, notice that for an arbitrary $g(x)$, we then have

$$\mathbb{E}^P \left(e^{-\int_t^T r(s, X(s)) ds} g(X(T)) | X(t) = x \right) = \int_{\mathbb{R}} G(t, x; T, y) g(y) dy. \quad (1.36)$$

Comparison with (1.32) shows that the state-price density is, essentially, equivalent to a Green's function with built-in discounting.

1.9 Black-Scholes and Extensions

In reviews of asset pricing theory, a discussion of the seminal *Black-Scholes-Merton* model (sometimes just known as the *Black-Scholes* model) of Black and Scholes [1973] and Merton [1973] is nearly mandatory. As the Black-Scholes-Merton (BSM) model constitutes a well-behaved setting in which to tie elements of previous sections together, our text is no exception. To provide a smoother transition to material that follows, we do, however, extend the usual analysis to include a simple case of stochastic interest rates.

1.9.1 Basics

In the basic BSM economy, two assets are traded: a money market account β and a stock S . In previous notations, $X(t) = (\beta(t), S(t))^\top$ and $p = 2$. The money market account value is 1 at time 0 and accrues risk-free interest at a continuously compounded, non-negative rate of r , initially assumed constant. The dynamics for β are thus given by an ordinary differential equation (ODE)

$$d\beta(t)/\beta(t) = r dt, \quad \beta(0) = 1,$$

implying that simply $\beta(t) = \beta(0)e^{rt}$.

The stock dynamics are assumed to satisfy GBMD under measure P:

$$dS(t)/S(t) = \mu dt + \sigma dW(t), \quad (1.37)$$

where W is a Brownian motion of dimension $d = 1$, and μ and σ are constants.

Taking first a probabilistic approach, we notice that β is positive and can be used as a numeraire. Let $S^\beta(t) = S(t)/\beta(t)$ be the stock price deflated by β . By Ito's lemma,

$$dS^\beta(t)/S^\beta(t) = (\mu - r) dt + \sigma dW(t).$$

Applying Girsanov's theorem (see Theorem 1.5.1) and Corollary 1.5.2, we see that if $\sigma \neq 0$, β will induce a unique equivalent martingale measure, with the measure shift characterized by the density process¹²

$$d\varsigma(t)/\varsigma(t) = -\theta dW(t), \quad \theta = \frac{\mu - r}{\sigma}.$$

Clearly, $\varsigma(t)$ defines an exponential martingale. The probability measure induced by the money market account β is called the *risk-neutral martingale measure* and is traditionally denoted Q. Under Q, $W^\beta(t) = W(t) + \theta t$ is a Brownian motion, and

$$\begin{aligned} dS^\beta(t)/S^\beta(t) &= \sigma dW^\beta(t), \\ dS(t)/S(t) &= r dt + \sigma dW^\beta(t), \end{aligned} \quad (1.38)$$

or, from (1.21),

$$S(T) = S(t)e^{(r - \frac{1}{2}\sigma^2)(T-t) + \sigma(W^\beta(T) - W^\beta(t))}, \quad t \in [0, T]. \quad (1.39)$$

We note that under Q, the drift μ of the stock process is replaced by the risk-free interest rate r . That is, under Q agents in the economy will

¹²The reader may recognize the market price of risk θ as the *Sharpe ratio* of the stock S , a measure of how well the risk of stock (represented by σ) is compensated by excess return (represented by $\mu - r$).

appear to be indifferent (“neutral”) to the risk of the stock, content with an average growth rate of the stock equal to that of the money market account.

Before proceeding with the BSM analysis, we wish to emphasize that the drift restriction imposed on the stock in the risk-neutral measure Q is a general result. In a larger setting with a p -dimensional vector asset process X , if the Q -dynamics of the components of X are all of the form

$$dX_i(t) = rX_i dt + O(dW(t)), \quad i = 1, \dots, p,$$

there is no arbitrage. This result holds unchanged if the interest rate is random (see Section 1.9.3).

Returning to the BSM setting, we note that the risk-neutral measure is unique, whereby the market is complete and all derivative securities on S (and β) are attainable. Let us consider a few such securities. First, we consider a security paying at time T \$1 for certain. Such a security is a *discount bond* and we shall denote its time t price by $P(t, T)$, $t \in [0, T]$. If the interest rate is positive, we would expect $P(t, T) \leq 1$ as a reflection of the time value of money, with equality only holding for $t = T$. Application of the basic derivative pricing equation (1.15) immediately gives

$$P(t, T) = \beta(t) E_t^Q \left(\frac{1}{\beta(T)} \right) = E_t^Q \left(e^{-r(T-t)} \right) = e^{-r(T-t)}.$$

This result is trivial, as it is easily seen that the amount $e^{-r(T-t)}$ invested in the money market account at time t will grow to exactly \$1 at time T .

Second, consider a derivative V paying $V(T) = S(T) - K$ at time T , with K being an arbitrary constant. Proceeding as above, at time $t \leq T$ the arbitrage-free price must be

$$\begin{aligned} V(t) &= E_t^Q \left(e^{-r(T-t)} (S(T) - K) \right) \\ &= e^{-r(T-t)} \left(E_t^Q (S(T)) - K \right) = S(t) - K P(t, T), \end{aligned} \quad (1.40)$$

where the last equality follows from property (1.22) of GBMD. We notice that $V(t) = 0$ if $K = S(t)/P(t, T)$. This value of K is known as the time t *forward price of $S(T)$* ¹³.

Third, consider the derivative that was the main focus of the original BSM analysis, a *European call option* paying¹⁴ $c(T) = (S(T) - K)^+$, with K being a positive *strike price*. Following (1.40), we can write

$$c(t) = P(t, T) E_t^Q \left((S(T) - K)^+ \right). \quad (1.41)$$

From the representation (1.39), basic probability theory allows us to write this expectation as

¹³We shall touch on the closely related concept of a *futures price* in Section 4.1.2.

¹⁴We use the notations $x^+ = \max(x, 0)$, $x^- = \min(x, 0)$ throughout this book.

$$c(t) = P(t, T) \int_{-\infty}^{\infty} \left(S(t) e^{(r - \frac{1}{2}\sigma^2)(T-t) + z\sigma\sqrt{T-t}} - K \right)^+ \phi(z) dz, \quad (1.42)$$

where $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ is the standard Gaussian density. A straightforward evaluation of the integral leads to the famous *Black-Scholes-Merton call pricing formula*:

Theorem 1.9.1. *In the BSM economy, the arbitrage-free time t price of a K -strike call option maturing at time T is*

$$c(t) = S(t)\Phi(d_+) - K P(t, T)\Phi(d_-), \quad (1.43)$$

$$d_{\pm} \triangleq \frac{\ln(S(t)/K) + (r \pm \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad t < T,$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function.

A formula for a *European put option* $p(t)$ paying $(K - S(T))^+$ can be obtained from (1.43) by *put-call parity*:

$$c(t) - p(t) = V(t),$$

where $V(t)$ is the forward contract defined above.

Remark 1.9.2. At time t , call and put options with strikes equal to $S(t)$ are said to be *at-the-money* (ATM). If $S(t) > K$, the call option is *in-the-money* (ITM) and the put option is *out-of-the-money* (OTM). If $S(t) < K$, the call is OTM and the put is ITM. The ATM, ITM, and OTM monikers are sometimes used to refer to the ordering of the *forward value* $E_t(S(T)) = S(t)e^{r(T-t)}$ (for a T -maturity option) rather than the spot $S(t)$, relative to the strike K .

In deriving (1.43), the choice of β as numeraire was arbitrary. If we instead use S (which is also strictly positive) as numeraire, we can write

$$c(t) = S(t)E_t^{Q^S} \left(\frac{(S(T) - K)^+}{S(T)} \right) = S(t)E_t^{Q^S} \left((1 - K/S(T))^+ \right), \quad (1.44)$$

where Q^S is the martingale measure induced by S . To identify the measure shift involved in moving from P to Q^S , consider that $\beta^S(t) = \beta(t)/S(t)$ must be a martingale in Q^S . By Ito's lemma, in measure P we have

$$d\beta^S(t)/\beta^S(t) = (r - \mu + \sigma^2) dt - \sigma dW(t),$$

such that $dW^S(t) = ((r - \mu)/\sigma + \sigma) dt + dW(t)$ is a Brownian motion under Q^S . Application of Ito's lemma on $1/S(t)$ yields, after a few rearrangements,

$$dS(t)^{-1}/S(t)^{-1} = -rdt - \sigma dW^S(t),$$

which is a GBMD as before. Evaluation of the expectation (1.44) can be verified to recover the BSM formula (1.43).

Our derivation of the BSM formula was so far entirely probabilistic. Writing $c(t) = c(t, \beta, S)$, the arguments in Section 1.7 allow us to write c as a solution to the PDE (see (1.27))

$$\frac{\partial c}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 c}{\partial S^2} = 0, \quad (1.45)$$

subject to the boundary condition $c(T, \beta, S) = (S - K)^+$. From (1.28) we also have that the replication positions in β and S are $\frac{\partial c}{\partial \beta}$ and $\frac{\partial c}{\partial S}$, respectively. That is,

$$c(t, \beta, S) = \frac{\partial c}{\partial \beta} \beta + \frac{\partial c}{\partial S} S. \quad (1.46)$$

As β is deterministic, we can actually eliminate c -dependence on this variable by a change of variables $\tilde{c}(t, S) = c(t, \beta, S)$. By the chain rule

$$\frac{\partial \tilde{c}}{\partial t} = \frac{\partial c}{\partial t} + \frac{\partial c}{\partial \beta} \frac{\partial \beta}{\partial t} = \frac{\partial c}{\partial t} + \frac{\partial c}{\partial \beta} r\beta = \frac{\partial c}{\partial t} + rc - \frac{\partial c}{\partial S} rS$$

where the last equation follows from (1.46). Inserting this into (1.45) yields the original *Black-Scholes PDE*

$$\frac{\partial \tilde{c}}{\partial t} + rS \frac{\partial \tilde{c}}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \tilde{c}}{\partial S^2} = r\tilde{c}, \quad (1.47)$$

with $\tilde{c}(T, S) = (S - K)^+$. We can solve this equation by classical methods (see Lipton [2001] for several techniques), or we can use the Feynman-Kac result to write it as an expectation. We leave it as an exercise to the reader to verify that Feynman-Kac leads to the same expectation as derived earlier by probabilistic means (see (1.41)).

A final note: the derivation of the Black-Scholes PDE above was somewhat non-standard due to the initial assumption of option price being a function of the deterministic numeraire β . A more conventional (but entirely equivalent) argument sets up a portfolio of the call option and a position in the stock, and demonstrates that the stock position can be set such that the total portfolio growth is deterministic (risk-free) on $[t, t+dt]$. Equating the portfolio growth with the risk-free rate yields the Black-Scholes PDE (1.47). See Hull [2006] for details of this approach.

1.9.2 Alternative Derivation

We have already demonstrated several different ways of proving the BSM call pricing formula, but as shown in Andreasen et al. [1998] there are many more. One particularly enlightening proof is based on the concept of *local time* and shall briefly be discussed in this section. The proof, which borrows

from the results in Carr and Jarrow [1990], will also allow us to demonstrate the Tanaka extension of Ito's lemma, mentioned earlier in Section 1.1.

As above, we assume that the stock price process is as in (1.38), and define the forward stock price $F(t) \triangleq S(t)/P(t, T)$. Clearly,

$$dF(t)/F(t) = \sigma dW^\beta(t), \quad t \leq T, \quad (1.48)$$

where W^β is a Brownian motion in the risk-neutral measure. Define the random variable $I(t) = (F(t) - K)^+$. The first derivative of I with respect to F is an indicator function $1_{\{F(t)>K\}}$ and the second derivative can be interpreted as the Dirac delta function, $\delta(F(t) - K)$. As I is clearly not twice differentiable, Ito's lemma formally does not apply, but the Tanaka extension nevertheless gives us permission to write

$$\begin{aligned} dI(t) &= 1_{\{F(t)>K\}} dF(t) + \frac{1}{2}\sigma^2 F(t)^2 \delta(F(t) - K) dt \\ &= 1_{\{F(t)>K\}} \sigma F(t) dW^\beta(t) + \frac{1}{2}\sigma^2 K^2 \delta(F(t) - K) dt. \end{aligned}$$

In integrated form,

$$I(T) = I(t) + \int_t^T 1_{\{F(u)>K\}} \sigma F(u) dW^\beta(u) + \frac{1}{2}\sigma^2 K^2 \int_t^T \delta(F(u) - K) du.$$

The second integral in this expression is a random variable known as the *local time of F spent at the level K* , on the interval $[t, T]$. Taking expectations, it follows that

$$E_t^Q(I(T)) = I(t) + \frac{1}{2}\sigma^2 K^2 \int_t^T E_t^Q(\delta(F(u) - K)) du.$$

Here, if $p(t, y; u, x)$ is the density of $F(u)$ given $F(t) = y$, $u \geq t$, then obviously

$$E_t^Q(\delta(F(u) - K)) = p(t, F(t); u, K).$$

By the definition of $F(T)$ we have $F(T) = S(T)$, such that $I(T) = (S(T) - K)^+$. From (1.41), we may therefore write the time t European call option price as

$$\begin{aligned} c(t) &= P(t, T) E_t^Q(I(T)) \\ &= (S(t) - KP(t, T))^+ + \frac{P(t, T)}{2}\sigma^2 K^2 \int_t^T p(t, F(t); u, K) du. \quad (1.49) \end{aligned}$$

The formula (1.49) decomposes the call option into a sum of two terms, the *intrinsic value* and the *time value*, respectively. The time value can be made more explicit by observing from the representation (1.39) that¹⁵

¹⁵This also follows directly from the fact that $F(u)$ is a log-normal random variable with moments given by (1.22) and (1.23).

$$p(t, F(t); u, K) = \frac{1}{K\sigma\sqrt{u-t}\sqrt{2\pi}} \exp\left(-\frac{1}{2}d(u)^2\right),$$

$$d(u) \triangleq \frac{\ln(F(t)/K) - \frac{1}{2}\sigma^2(u-t)}{\sigma\sqrt{u-t}}.$$

In other words, we have arrived at the following result.

Proposition 1.9.3. *The European call option price $c(t)$ on the process (1.38) can be written as*

$$c(t) = (S(t) - KP(t, T))^+ + \frac{P(t, T)\sigma K}{2} \int_t^T \frac{\phi(d(u))}{\sqrt{u-t}} du, \quad (1.50)$$

where $\phi(x)$ is the Gaussian density.

Explicit evaluation of the integral in (1.50) can be verified to produce the BSM formula in Theorem 1.9.1. We leave this as an exercise to the reader.

1.9.3 Extensions

1.9.3.1 Deterministic Parameters and Dividends

In our basic BSM setup, consider now first a simple extension to a deterministic interest rate $r(t)$ and a deterministic volatility $\sigma(t)$. Carrying out the analysis as before, we see that discount bond prices now become

$$P(t, T) = e^{-\int_t^T r(s) ds}. \quad (1.51)$$

The BSM call pricing formula (1.43) holds unchanged provided $P(t, T)$ is changed according to (1.51), and we redefine

$$d_{\pm} \triangleq \frac{\ln(S(t)/K) + \int_t^T (r(s) \pm \sigma(s)^2/2) ds}{\sqrt{\int_t^T \sigma(s)^2 ds}}.$$

Let us further assume that the stock pays dividends at a deterministic rate of $q(t)$. Our framework so far, however, has assumed that assets pay no cash over $[0, T]$. To salvage the situation, consider a fictitious asset S^* obtained by reinvesting all dividends into the stock S itself. It is easily seen that

$$S^*(t) = S(t)e^{\int_0^t q(s) ds},$$

and clearly $S^*(t)$ satisfies the requirements of generating no cash flows on $[0, T]$. Stating the call option payout as

$$c(T) = (S(T) - K)^+ = \left(S^*(T)e^{-\int_0^T q(s) ds} - K \right)^+$$

and performing the pricing analysis of Section 1.9.1 on $S^*(t)$, rather than $S(t)$, results in a dividend-extended BSM call option formula:

$$c(t) = S(t)e^{-\int_t^T q(s) ds} \Phi(d_+) - KP(t, T)\Phi(d_-),$$

$$d_{\pm} \triangleq \frac{\ln(S(t)/K) + \int_t^T (r(s) - q(s)) \pm \sigma(s)^2/2 ds}{\sqrt{\int_t^T \sigma(s)^2 ds}}.$$

When the stock pays a dividend rate of $q(t)$, note that the risk-neutral process for $S(t)$ is

$$dS(t)/S(t) = (r(t) - q(t)) dt + \sigma(t) dW^{\beta}(t),$$

which extends (1.38). Note that for the special case where $r(t) = q(t)$, $S(t)$ becomes a martingale and the call option price formula simplifies to

$$c(t) = P(t, T)(S(t)\Phi(d_+) - K\Phi(d_-)), \quad (1.52)$$

where now

$$d_{\pm} \triangleq \frac{\ln(S(t)/K) \pm \frac{1}{2} \int_t^T \sigma(s)^2 ds}{\sqrt{\int_t^T \sigma(s)^2 ds}}.$$

Remark 1.9.4. The martingale call formula (1.52) typically emerges when pricing options on futures and forward prices (see (1.48)) and is often called the *Black formula*, in honor of the work in Black [1976].

1.9.3.2 Stochastic Interest Rates

We now get even more ambitious and wish to consider call option pricing in the case where the interest rate r is stochastic. The money market account β becomes

$$\beta(t) = e^{\int_0^t r(s) ds},$$

and is now assumed an \mathcal{F}_t -measurable random variable. Proceeding as in Section 1.9.1, we find that under the risk-neutral measure Q , the call option price expression is (assuming that the stock pays no dividends)

$$c(t) = \beta(t) \mathbb{E}_t^Q \left(\frac{1}{\beta(T)} (S(T) - K)^+ \right) = \mathbb{E}_t^Q \left(e^{-\int_t^T r(s) ds} (S(T) - K)^+ \right). \quad (1.53)$$

In (1.53), we emphasize that the numeraire no longer can be pulled out from the expectation. Still, to simplify call option computations, it would be convenient to somehow remove the term $\exp(-\int_t^T r(s) ds)$ from the expectation in (1.53). By substituting 1 for $(S(T) - K)^+$ in the expression above, we first notice that

$$P(t, T) = \mathbb{E}_t^Q \left(e^{-\int_t^T r(s) ds} \right).$$

This inspires us to perform a new measure shift, where we use the discount bond $P(t, T)$, rather than $\beta(t)$, as our numeraire. Let the martingale measure induced by $P(t, T)$ be denoted Q^T , often termed the *T-forward measure*. By the standard result (1.15) we have

$$\begin{aligned} c(t) &= P(t, T) \mathbb{E}_t^{Q^T} \left(P(T, T)^{-1} (S(T) - K)^+ \right) \\ &= P(t, T) \mathbb{E}_t^{Q^T} \left((S(T) - K)^+ \right), \end{aligned}$$

where we have used that $P(T, T) = 1$. From Theorem 1.4.2, Q^T and Q are related by the density

$$\varsigma(t) = \mathbb{E}_t^Q \left(\frac{dQ^T}{dQ} \right) = \frac{P(t, T)/P(0, T)}{\beta(t)}. \quad (1.54)$$

To proceed, we need to add more structure to the model by making assumptions about the stochastic process for $P(t, T)$. We shall spend considerable effort in subsequent chapters on this issue, but for this initial application we simply assume that $P(t, T)$ has Q dynamics

$$dP(t, T)/P(t, T) = r(t) dt - \sigma_P(t, T) dW_P(t), \quad (1.55)$$

where $\sigma_P(t, T)$ is deterministic and $W_P(t)$ is a Brownian motion correlated to the stock Brownian motion. Notice that the drift of $P(t, T)$ under Q is not freely specifiable and must be equal to the risk-free rate; see the discussion following (1.38). For clarity, let the stock Brownian motion be renamed $W_S(t)$, and assume that the correlation between $W_P(t)$ and $W_S(t)$ is a constant ρ . In the setting of vector-valued Brownian motion with independent components used in earlier sections, we can introduce correlation by writing $W(t) = (W_1(t), W_2(t))^\top$ and setting, say,

$$\begin{aligned} W_P(t) &= W_1(t), \\ W_S(t) &= \rho W_1(t) + \sqrt{1 - \rho^2} W_2(t). \end{aligned}$$

The filtration $\{\mathcal{F}_t\}$ of our extended BSM setting is the one generated by the 2-dimensional $W(t)$.

Under Q^T , the deflated process $S^P(t) = S(t)/P(t, T)$ is a martingale. An application of Ito's lemma combined with the Diffusion Invariance Principle shows that the Q^T process for $S^P(t)$ is

$$dS^P(t)/S^P(t) = \sigma_P(t, T) dW_1(t) + \sigma(t) \left(\rho dW_1(t) + \sqrt{1 - \rho^2} dW_2(t) \right), \quad (1.56)$$

where $\sigma(t)$ as before is the deterministic volatility of the stock S . We recognize $S^P(t)$ as a drift-free geometric Brownian motion with instantaneous variance of

$$\sigma_P(t, T)^2 + \sigma(t)^2 + 2\rho\sigma(t)\sigma_P(t, T).$$

Exploiting the convenient fact that $S^P(T) = S(T)$ and $c(T) = (S^P(T) - K)^+$ (as $P(T, T) = 1$), we get

$$\begin{aligned} c(t) &= P(t, T) E_t^{Q^T} \left((S^P(T) - K)^+ \right) \\ &= P(t, T) \int_{-\infty}^{\infty} \left(S^P(t) e^{-\frac{1}{2}v(t, T) + z\sqrt{v(t, T)}} - K \right)^+ \phi(z) dz, \end{aligned} \quad (1.57)$$

where we have defined the “term”, or total, variance

$$v(t, T) \triangleq \int_t^T (\sigma_P(s, T)^2 + \sigma(s)^2 + 2\rho\sigma(s)\sigma_P(s, T)) ds. \quad (1.58)$$

Completing the integration (compare with (1.42)) and using $S^P(t) = S(t)/P(t, T)$, we arrive at a modified BSM-type call option formula:

Proposition 1.9.5. *Consider a BSM economy with stochastic interest rates evolving according to (1.55). Define term variance $v(t, T)$ as in (1.58). Then, the T -maturity European call option price is*

$$\begin{aligned} c(t) &= S(t)\Phi(d_+) - KP(t, T)\Phi(d_-), \\ d_{\pm} &= \frac{\ln(S(t)/(KP(t, T))) \pm \frac{1}{2}v(t, T)}{\sqrt{v(t, T)}}. \end{aligned}$$

Proposition 1.9.5 was originally derived in Merton [1973], using PDE methods. Extensions to dividend-paying stocks are straightforward and follow the arguments shown in Section 1.9.3.1.

1.10 Options with Early Exercise Rights

In our previous definition of a contingent claim, we assumed that the claim involved a single \mathcal{F}_T -measurable payout at time T . In reality, a number of derivative contracts may have intermediate cash payments from, say, scheduled coupons or through “rebates” for barrier-style options. Mostly, such complications are straightforwardly incorporated; see for instance Section 2.7.3. Of particular interest from a theoretical perspective are the claims that allow the holder to accelerate payments through *early exercise*. Derivative securities with early exercise are characterized by an adapted payout process $U(t)$, payable to the option holder at a stopping time (or *exercise policy*) $\tau \leq T$, chosen by the holder. If early exercise can take place at any time in some interval, we say that the derivative security is an *American option*; if exercise can only take place on a discrete set of dates, we say that it is a *Bermudan option*.

Let the allowed (and deterministic) set of exercise dates larger than or equal to t be denoted $\mathcal{D}(t)$, and suppose that we are given at time 0 a particular exercise policy τ taking values in $\mathcal{D}(0)$, as well as a pricing numeraire N inducing a unique martingale measure Q^N . Let $V^\tau(0)$ be the time 0 value of a derivative security that pays $U(\tau)$. Under some technical conditions on $U(t)$, we can write for the value of the derivative security

$$V^\tau(0) = E^{Q^N} \left(\frac{U(\tau)}{N(\tau)} \right), \quad (1.59)$$

where we have assumed, with no loss of generality, that $N(0) = 1$. Let $\mathcal{T}(t)$ be the time t set of (future) stopping times taking value in $\mathcal{D}(t)$. In the absence of arbitrage, the time 0 value of a security with early exercise into U must then be given by the *optimal stopping problem*

$$V(0) = \sup_{\tau \in \mathcal{T}(0)} V^\tau(0) = \sup_{\tau \in \mathcal{T}(0)} E^{Q^N} \left(\frac{U(\tau)}{N(\tau)} \right), \quad (1.60)$$

reflecting the fact that a rational investor would choose an exercise policy to optimize the value of his claim.

We can extend (1.60) to future times t by

$$V(t) = N(t) \sup_{\tau \in \mathcal{T}(t)} E_t^{Q^N} \left(\frac{U(\tau)}{N(\tau)} \right), \quad (1.61)$$

where $\sup_{\tau \in \mathcal{T}(t)} E_t^{Q^N} (U(\tau)/N(\tau))$ is known as the *Snell envelope* of U/N under Q^N . The process $V(t)$ must here be interpreted as the value of the option with early exercise, *conditional* on exercise not having taken place before time t . To make this explicit, let $\tau^* \in \mathcal{T}(0)$ be the optimal exercise policy, as seen from time 0. We can then write, for $0 < t \leq T$,

$$V(0) = E^{Q^N} (1_{\{\tau^* \geq t\}} V(t)/N(t)) + E^{Q^N} (1_{\{\tau^* < t\}} U(\tau^*)/N(\tau^*)), \quad (1.62)$$

where we break the time 0 value into two components: one from the time t value of the option, should it not have been exercised before time t ; and one from the right to exercise on $[0, t]$. As we can always elect — possibly suboptimally — to never exercise on $[0, t]$, from (1.62) we see that

$$V(0) \geq E^{Q^N} (V(t)/N(t)),$$

which establishes that $V(t)/N(t)$ is a *supermartingale* under Q^N . This result also follows directly from known properties of the Snell envelope; see, e.g., Musiela and Rutkowski [1997].

For later use, focus now on the Bermudan case and assume that $\mathcal{D}(0) = \{T_1, T_2, \dots, T_B\}$, where $T_1 > 0$ and $T_B = T$. For $t < T_{i+1}$, define $H_i(t)$ as the time t value of the Bermudan option when exercise is restricted to the dates $\mathcal{D}(T_{i+1}) = \{T_{i+1}, T_{i+2}, \dots, T_B\}$. That is

$$H_i(t) = N(t) \mathbb{E}_t^{\mathbb{Q}^N} (V(T_{i+1})/N(T_{i+1})) , \quad i = 1, \dots, B-1.$$

At time T_i , $H_i(T_i)$ can be interpreted as the *hold value* of the Bermudan option, that is, the value of the Bermudan option if not exercised at time T_i . If an optimal exercise policy is followed, clearly we must have at time T_i

$$V(T_i) = \max (U(T_i), H_i(T_i)) , \quad i = 1, \dots, B, \quad (1.63)$$

such that

$$H_i(t) = N(t) \mathbb{E}_t^{\mathbb{Q}^N} (\max (U(T_{i+1}), H_{i+1}(T_{i+1})) / N(T_{i+1})) , \quad i = 1, \dots, B-1. \quad (1.64)$$

Starting with the terminal condition $H_B(T) = 0$, (1.64) defines a useful iteration backwards in time for the value $V(0) = H_0(0)$. We shall use this later for the purposes of designing valuation algorithms in Chapter 18, and for computing price sensitivities (deltas) in Chapter 24.

We note that the idea behind (1.63) is often known as *dynamic programming* or the *Bellman principle*. Loosely speaking, we here work “from the back” to price the Bermudan option. As we shall see later (in Chapter 2), this idea is particularly well-suited for numerical methods that proceed backwards in time, such as finite difference methods.

1.10.1 The Markovian Case

We now specialize to the Markovian case where $U(t) = g(t, x(t))$, where $g : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and

$$dx(t) = \mu(t, x(t)) dt + \sigma(t, x(t)) dW(t) \quad (1.65)$$

is an n -dimensional Markovian process, where μ and σ satisfy the regularity conditions of Theorem 1.6.1. The n -dimensional process¹⁶ $x(t)$ here defines the state of the exercise value $U(t)$, so we say that $x(t)$ is a *state variable process*. For concreteness let our numeraire $N(t)$ be the money market account

$$N(t) = \beta(t) = e^{\int_0^t r(u, x(u)) du},$$

where the short interest rate $r : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ is here assumed a function of time and the state variable vector x . In (1.65), $W(t)$ is understood to be a d -dimensional Brownian motion in the risk-neutral measure \mathbb{Q} .

Writing $V(t) = V(t, x(t))$, we have from (1.61)

$$V(t, x) = \sup_{\tau \in \mathcal{T}(t)} \mathbb{E}^{\mathbb{Q}} \left(e^{-\int_t^\tau r(u, x(u)) du} g(\tau, x(\tau)) \middle| x(t) = x \right). \quad (1.66)$$

¹⁶Note that $x(t)$ is an abstract construct, and does not necessarily coincide with any asset price process.

For dates $t \in \mathcal{D}(0)$, clearly $V(t, x) \geq g(t, x)$, with equality holding only when time t exercise is optimal. This leads us to define the concept of an *exercise region* as

$$\mathcal{X} = \{(t, x) \in \mathcal{D}(0) \times \mathbb{R}^n : V(t, x) = g(t, x)\}.$$

Similarly, we define the complement of \mathcal{X} ,

$$\mathcal{C} = \{(t, x) \in [0, T] \times \mathbb{R}^n : (t, x) \notin \mathcal{X}\},$$

to be the *continuation region*, i.e. the region where we wait (either because exercise is not optimal or because it is not allowed, $t \notin \mathcal{D}(0)$) rather than exercise the option.

For Markovian systems, rather than solving the optimization problem (1.66) directly, it is often particularly convenient to invoke the Bellman principle. Extending the ideas presented earlier, let us, somewhat loosely, state the Bellman principle as follows: for any $t \in \mathcal{D}(0)$,

$$V(t, x) = \lim_{\Delta \downarrow 0} \max \left(g(t, x), \mathbb{E}_t^Q \left(e^{-\int_t^{t+\Delta} r(u, x(u)) du} V(t + \Delta, x(t + \Delta)) \right) \right). \quad (1.67)$$

Again, this simply says that the option value at time t is the maximum of the exercise value and the hold value, that is, the present value of continuing to hold on to the option for a small period of time. As we have seen above, for a Bermudan option, (1.67) also holds for finite Δ (namely up to the next exercise date).

The Bellman principle provides us with a link between present (time t) and future (time $t + \Delta$) option values that we can often exploit in a numerical scheme. For this, however, we need further characterization of $V(t, x)$ in the continuation region. By earlier arguments, we realize that $V(t, x)/\beta(t)$ must be a Q-martingale on the continuation region. Assuming sufficient smoothness for an application of Ito's lemma, this leads to a PDE formulation, to hold for $(t, x) \in \mathcal{C}$,

$$\mathcal{J}V(t, x) = 0, \quad (1.68)$$

where

$$\mathcal{J} = \frac{\partial}{\partial t} + \mu(t, x) \frac{\partial}{\partial x} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\sigma(t, x) \sigma(t, x)^\top)_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} - r(t, x).$$

Assume first that our option is of the Bermudan type, and let T_i and T_{i+1} be subsequent exercise dates in the exercise schedule. For any function f of time, define $f(t \pm)$ to be the limits $\lim_{\epsilon \downarrow 0} f(t \pm \epsilon)$, and assume that $V(T_{i+1}-, x)$ is known for all x . As all values of $t \in (T_i, T_{i+1})$ by definition must be in the continuation region, we can use (1.68) to solve for $V(T_i+, x)$. Applying the Bellman principle (1.67) at time T_i then leads to the condition

$$V(T_i-, x) = \max(g(T_i, x), V(T_i+, x)).$$

In PDE parlance, this is a so-called *jump condition* which is straightforward to incorporate into a numerical solution; see Section 2.7.4 for details.

For American-style options, (1.68) continues to apply on \mathcal{C} . The Bellman principle here leads to the characterization that

$$\mathcal{J}V(t, x) < 0,$$

for $(t, x) \in \mathcal{X}$, i.e. we exercise when the rate of return from holding the option strictly fails to match $r(t, x)$. The American option pricing problem is often conveniently summarized in a *variational inequality*, to hold on $\mathcal{X} \cup \mathcal{C}$,

$$V(t, x) \geq g(t, x), \quad \mathcal{J}V(t, x) \leq 0, \quad (V(t, x) - g(t, x)) \mathcal{J}V(t, x) = 0, \quad (1.69)$$

and subject to the boundary condition $V(T, x) = g(T, x)$. The first of these three conditions expresses that the option is always worth at least its exercise value; the second expresses the supermartingale property of $V(t, x)$; and the third implies (after a little thought) that $\mathcal{J}V(t, x) = 0$ on \mathcal{C} and $\mathcal{J}V(t, x) < 0$ on \mathcal{X} . The system (1.69) is discussed more carefully in Duffie [2001], where additional discussion of regularity issues may also be found.

1.10.2 Some General Bounds

In many cases of practical interest, solving PDEs and/or variational inequalities is not computationally feasible. In such situations, we may be interested in at least bounding the value of an option with early exercise rights. Providing a lower bound is straightforward: postulate an exercise policy τ and compute the price $V^\tau(0)$ by direct methods. From (1.60), clearly this provides a lower bound

$$V^\tau(0) \leq V(0). \quad (1.70)$$

The closer the postulated exercise policy τ is to the optimal exercise policy τ^* , the tighter this bound will be. We shall later study a number of numerical techniques to generate good exercise strategies for fixed income options with early exercise rights, see Chapter 18.

To produce an upper bound, we can rely on duality results established in Rogers [2001], Haugh and Kogan [2004] and Andersen and Broadie [2004]. Let \mathcal{K} denote the space of adapted martingales M for which $\sup_{\tau \in [0, T]} \mathbb{E}^{Q^N} |M(\tau)| < \infty$. For a martingale $M \in \mathcal{K}$, we then write

$$\begin{aligned} V(0) &= \sup_{\tau \in \mathcal{T}(0)} \mathbb{E}^{Q^N} \left(\frac{U(\tau)}{N(\tau)} \right) \\ &= \sup_{\tau \in \mathcal{T}(0)} \mathbb{E}^{Q^N} \left(\frac{U(\tau)}{N(\tau)} + M(\tau) - M(\tau) \right) \\ &= M(0) + \sup_{\tau \in \mathcal{T}(0)} \mathbb{E}^{Q^N} \left(\frac{U(\tau)}{N(\tau)} - M(\tau) \right). \end{aligned}$$

In the second equality, we have relied on the *optional sampling theorem*, a result that states that the martingale property is satisfied up to a bounded random stopping time, i.e. that $E^{Q^N}(M(\tau)) = M(0)$; see Karatzas and Shreve [1997] for details. We now turn the above result into an upper bound by forming a pathwise maximum at all possible future exercise dates $\mathcal{D}(0)$:

$$\begin{aligned} V(0) &= M(0) + \sup_{\tau \in \mathcal{T}(0)} E^{Q^N} \left(\frac{U(\tau)}{N(\tau)} - M(\tau) \right) \\ &\leq M(0) + E^{Q^N} \left(\max_{t \in \mathcal{D}(0)} \left(\frac{U(t)}{N(t)} - M(t) \right) \right). \end{aligned} \quad (1.71)$$

With (1.70) and (1.71) we have, as desired, established upper and lower bounds for values of options with early exercise rights. Let us consider how to make these bounds tight. As mentioned earlier, to tighten the lower bound we need to pick exercise strategies close to the optimal one. Tightening the upper bound is a bit more involved and requires the following basic theorem, proven in Karatzas and Shreve [1997]:

Theorem 1.10.1 (Doob-Meyer Decomposition). *Let $\{Y(t), t \in [0, T]\}$ be a positive \mathcal{F}_t -adapted supermartingale process with right-continuous sample paths. Then we can write*

$$Y(t) = m(t) - A(t),$$

where $m(t)$ is a martingale process with $m(0) = Y(0)$ and $A(t)$ is an increasing predictable process with $A(0) = 0$.

Applying the Doob-Meyer decomposition on the supermartingale process $V(t)/N(t)$ under Q^N shows that

$$V(t)/N(t) = m(t) - A(t),$$

and $V(0) = m(0)$. Consider taking $M(t) = m(t)$ in equation (1.71), to get

$$\begin{aligned} V(0) &\leq V(0) + E^{Q^N} \left(\max_{t \in \mathcal{D}(0)} \left(\frac{U(t)}{N(t)} - m(t) \right) \right) \\ &= V(0) + E^{Q^N} \left(\max_{t \in \mathcal{D}(0)} \left(\frac{U(t)}{N(t)} - \frac{V(t)}{N(t)} - A(t) \right) \right) \\ &\leq V(0). \end{aligned}$$

The last inequality follows from the fact that $V(t) \geq U(t)$ and $A(t) \geq 0$. In conclusion, we have arrived at a *dual* formulation of the option price

$$V(0) = \inf_{M \in \mathcal{K}} \left\{ M(0) + E^{Q^N} \left(\max_{t \in \mathcal{D}(0)} \left(\frac{U(t)}{N(t)} - M(t) \right) \right) \right\}, \quad (1.72)$$

and have demonstrated that the infimum is attained when the martingale M is set equal to the martingale component of the deflated price process

$V(t)/N(t)$. In practice, we are obviously not privy to $V(t)/N(t)$ (which is a quantity that we are trying to estimate), but we are nevertheless provided with a strategy to make the upper bound (1.71) tight: use a martingale that is “close” to the martingale component of the true deflated option price process. In Chapter 18 we shall demonstrate how to make this strategy operational.

1.10.3 Early Exercise Premia

We finish our discussion of options with early exercise rights by listing some known results for puts and calls, including an interesting decomposition of American and Bermudan option prices into the sum of a European option price and an *early exercise premium*. For convenience, we work in a Markovian setting where the single state variable, denoted $S(t)$, follows one-dimensional GBMD. Specifically, we assume that

$$dS(t)/S(t) = (r - q) dt + \sigma dW^\beta(t), \quad (1.73)$$

with $W^\beta(t)$ being a one-dimensional Brownian motion in the risk-neutral measure, i.e. the measure induced by the money market account $\beta(t) = e^{rt}$. For simplicity we assume that the interest rate r , the dividend yield q , and the volatility σ are all constants; the extension to time-dependent parameters is straightforward.

Let $c(t)$, $C_A(t)$, and $C_B(t)$ be the time t European, American, and Bermudan prices of the call option with terminal maturity T , conditional on no exercise prior to time t . While obviously $c(t) \leq C_B(t) \leq C_A(t)$, in some cases these inequalities are equalities, as the following straightforward lemma shows.

Lemma 1.10.2. *Suppose that $r \geq 0$ and $q \leq 0$ in (1.73). It is then never optimal to exercise a call option early, and*

$$c(t) = C_A(t) = C_B(t).$$

Proof. Notice that, by Jensen’s inequality,

$$\begin{aligned} c(t) &= e^{-r(T-t)} \mathbb{E}_t^Q ((S(T) - K)^+) \\ &\geq e^{-r(T-t)} \left((\mathbb{E}_t^Q (S(T)) - K)^+ \right) = (e^{-q(T-t)} S(t) - e^{-r(T-t)} K)^+. \end{aligned}$$

It is therefore clear that if $r \geq 0$ and $q \leq 0$, then for any value of $T - t$,

$$c(t) \geq (S(t) - K)^+,$$

i.e. the European call option price dominates the exercise value. As the hold value of American and Bermudan options must be at least as large as the European option price, it follows that the option to exercise early is worthless. \square

Remark 1.10.3. For the put option, early exercise is never optimal if $r \leq 0$ and $q \geq 0$. As this situation rarely happens in practice, American put options nearly always trade at a premium to their European counterparts.

Lemma 1.10.2 demonstrates the well-known fact that American or Bermudan call options on stocks that pay no dividends ($q = 0$) should never be exercised early. On the other hand, if the stock does pay dividends, for an American call option there will, at time t , be a critical value of the stock, $S_A(t)$, at which the value of the stream of dividends paid by the stock will compensate for the cost of accelerating the payment of the strike K . In other words, an American option should be exercised at time t , provided that $S(t) \geq S_A(t)$. The deterministic curve $S_A(t)$ is known as the *early exercise boundary* and marks the boundary between the exercise and continuation regions, \mathcal{X} and \mathcal{C} . Writing $C_A(t) = C_A(t, S(t))$, we formally have

$$S_A(t) = \inf \left\{ S : C_A(t, S) = (S - K)^+ \right\}, \quad t \leq T.$$

For a Bermudan option, we may similarly define

$$S_B(t) = \inf \left\{ S : C_B(t, S) = (S - K)^+ \right\}, \quad t \in \mathcal{D}(0),$$

where we recall that $\mathcal{D}(0)$ is the (discrete) set of allowed exercise dates for the Bermudan option.

The following important result characterizes the exercise boundary of American call options.

Proposition 1.10.4. *For the American call option on a stock that follows (1.73), we have*

$$\frac{\partial S_A(t)}{\partial t} \leq 0, \quad t < T, \tag{1.74}$$

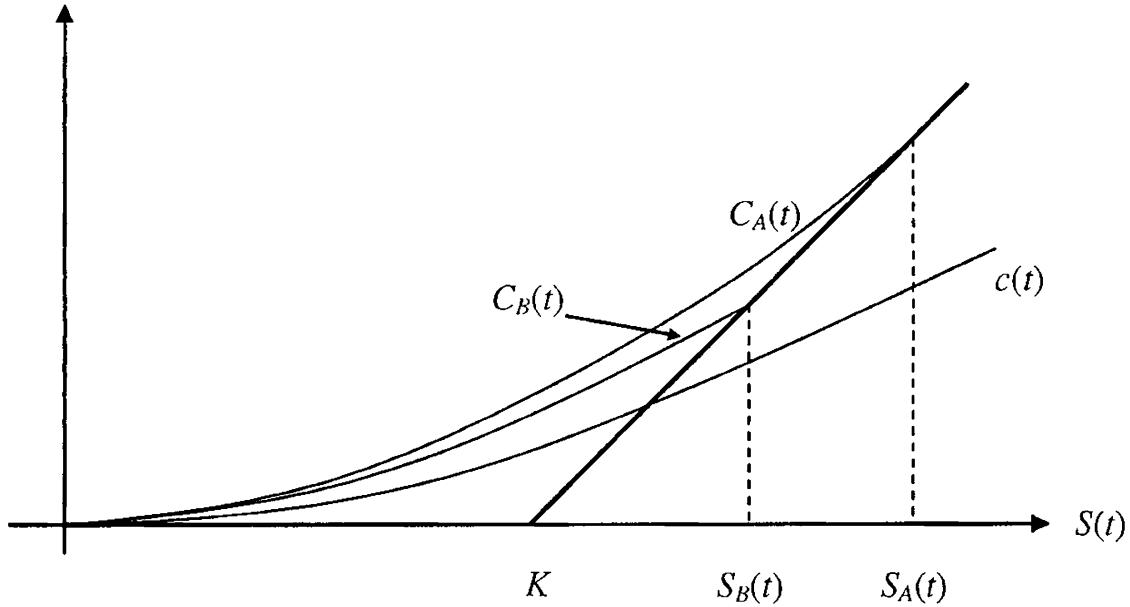
and

$$\left. \frac{\partial C_A(t, S)}{\partial S} \right|_{S=S_A(t)} = 1, \quad t < T. \tag{1.75}$$

Equation (1.74) states that the exercise boundary decreases as we approach maturity, a result that is easily understood. Statement (1.75) is more subtle, however, and amounts to a tangency condition that ensures that the American call option value transitions smoothly from hold value to exercise value across the early exercise boundary. As a consequence, (1.75) is often known as the *smooth pasting condition* or the *high contact condition*. A similar tangency condition does *not* hold for the Bermudan option value, which is not differentiable at the boundary but instead transitions into the exercise region at a “kink”:

$$\lim_{\varepsilon \downarrow 0} \left. \frac{C_B(t, S) - C_B(t, S - \varepsilon)}{\varepsilon} \right|_{S=S_B(t)} < 1, \quad t \in \mathcal{D}(0).$$

Fig. 1.1. Call Option Prices



Notes: Time t prices of American, Bermudan, and European call options, as a function of the asset price. The Bermudan option is assumed to be exercisable at time t .

Figure 1.1 shows a typical value profile for a Bermudan call, along with the corresponding profiles for the European and American options.

Smooth pasting is essentially an optimality condition, which is how Proposition 1.10.4 is traditionally derived (see, e.g., Merton [1973] or the more recent Brekke and Øksendal [1991]). A more descriptive proof based on hedging arguments is given in Tavella and Randall [2000] and Wilmott et al. [1993]. Loosely speaking, the idea is here that a delta hedger should not be able to make riskless profits when the underlying asset crosses into the exercise region. This requires that the delta is continuous across the boundary, which is (1.75).

Remark 1.10.5. For the American put option, $\partial S_A(t)/\partial t \geq 0$ and the high contact condition states that the delta equals -1 at the exercise boundary.

Establishing the boundary $S_A(t)$ will virtually always require numerical methods, although asymptotic results are known for t close to T (see for instance Lipton [2001]). One simple result is listed below.

Lemma 1.10.6. *Assume that $r \geq 0$ and $q \geq 0$, such that the early exercise boundary exists for the American call option. The exercise boundary just prior to maturity is then*

$$\lim_{\varepsilon \downarrow 0} S_A(T - \varepsilon) = K \max \left(1, \frac{r}{q} \right).$$

Proof. An informal proof of Lemma 1.10.6 proceeds as follows. At time $T - dt$, assume that $S(T - dt) > K$; otherwise it clearly makes no sense to exercise the option. If we exercise the option, we receive $S(T - dt) - K$ at time $T - dt$. On the other hand, if we postpone exercise, at time $T - dt$ our hold value is

$$\begin{aligned} e^{-r dt} \mathbb{E}_{T-dt}^Q (S(T) - K) &= S(T - dt) e^{-q dt} - K e^{-r dt} \\ &= S(T - dt) - K - S(T - t) q dt + K r dt. \end{aligned}$$

Clearly, we should then only exercise if

$$S(T - dt) - K > S(T - dt) - K - S(T - t) q dt + K r dt$$

or if

$$S(T - dt) q > K r.$$

□

Notice that since clearly $S_A(T) = K$, the call option exercise boundary will have a *discontinuity* at time T , if $q < r$.

One might guess that complete knowledge of the curve $S_A(t)$ should suffice to price the American option analytically. This intuition is confirmed by the following result due to Jamshidian [1992], Carr et al. [1992], Kim [1990], and Jacka [1991].

Proposition 1.10.7. *The American option price $C_A(t)$ satisfies*

$$C_A(t) = c(t) + E_A(t), \quad t \leq T, \quad (1.76)$$

where the (American) early exercise premium $E_A(t)$ is defined as

$$E_A(t) = \int_t^T e^{-r(u-t)} \mathbb{E}_t^Q (1_{\{S(u) \geq S_A(u)\}} (qS(u) - rK)) du \quad (1.77)$$

$$= \int_t^T \left(qS(t) e^{-q(u-t)} \Phi(d_+(u)) - rK e^{-r(u-t)} \Phi(d_-(u)) \right) du, \quad (1.78)$$

where

$$d_{\pm}(u) = \frac{\ln(S(t)/S_A(u)) + (r - q \pm \frac{1}{2}\sigma^2)(u - t)}{\sigma\sqrt{u - t}}.$$

Proof. Due to the smooth pasting condition in Proposition 1.10.4, we are justified¹⁷ in applying Ito's lemma. In informal notation,

$$\begin{aligned} dC_A(t) &= 1_{\{S(t) \geq S_A(t)\}} dS(t) \\ &+ 1_{\{S(t) < S_A(t)\}} \left\{ \frac{\partial C_A(t)}{\partial t} dt + \frac{\partial C_A(t)}{\partial S} dS(t) + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C_A(t)}{\partial S^2} dt \right\}, \end{aligned} \quad (1.79)$$

¹⁷In particular, there is no local time contribution to $dC_A(t)$ at the boundary.

where we have used the fact that

$$1_{\{S(t) \geq S_A(t)\}} C_A(t) = 1_{\{S(t) \geq S_A(t)\}} (S(t) - K).$$

In the continuation region, $C_A(t, S)$ satisfies the PDE (1.47), i.e.

$$\frac{\partial C_A(t, S)}{\partial t} + (r - q)S \frac{\partial C_A(t, S)}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C_A(t, S)}{\partial S^2} = rC_A(t, S).$$

Inserting this into (1.79) we get, after a few rearrangements,

$$\begin{aligned} dC_A(t) &= rC_A(t) dt + 1_{\{S(t) < S_A(t)\}} (r - q)S(t) \frac{\partial C_A(t)}{\partial S} dW^\beta(t) \\ &\quad + 1_{\{S(t) \geq S_A(t)\}} \{ ((r - q)S(t) - rC_A(t)) dt + \sigma S(t) dW^\beta(t) \} \\ &= rC_A(t) dt + 1_{\{S(t) < S_A(t)\}} (r - q)S(t) \frac{\partial C_A(t)}{\partial S} dW^\beta(t) \\ &\quad + 1_{\{S(t) \geq S_A(t)\}} \{ (rK - qS(t)) dt + \sigma S(t) dW^\beta(t) \} \end{aligned}$$

Setting $y(t) = C_A(t)/\beta(t)$, it follows from Ito's lemma that

$$\begin{aligned} dy(t) &= e^{-rt} 1_{\{S(t) < S_A(t)\}} (r - q)S(t) \frac{\partial C_A(t)}{\partial S} dW^\beta(t) \\ &\quad + 1_{\{S(t) \geq S_A(t)\}} e^{-rt} \{ (rK - qS(t)) dt + \sigma S(t) dW^\beta(t) \}. \end{aligned}$$

Integrating and taking expectations leads to

$$E_t^Q(y(T)) = y(t) + \int_t^T e^{-ru} E_t^Q \left(1_{\{S(u) \geq S_A(u)\}} (rK - qS(u)) \right) du.$$

Applying the definition of $y(t)$ and the fact that $y(T) = e^{-rT}(S(T) - K)^+$ proves (1.76). The explicit form of the early exercise premium in (1.78) follows from the properties of GBMD. \square

Remark 1.10.8. Combining results from Lemma 1.10.6 and Proposition 1.10.4, it follows that $E_A(t) \geq 0$, so $C_A(t) \geq c(t)$ as expected.

The integral representation of the American call option in Proposition 1.10.7 forms the basis for a number of proposed computational methods for American option pricing. Loosely speaking, these methods are based on the idea of iteratively estimating the exercise boundary $S_A(t)$, often working backwards from $t = T$, after which an application of Proposition 1.10.7 will yield the American option price. A representative example of these methods can be found in Ju [1998]. See Chiarella et al. [2004] for a survey of the literature, and Section 19.7.3 for an application in interest rate derivative pricing.

For a Bermudan option, an integral representation such as that in Proposition 1.10.7 is not possible. Nevertheless, it is still possible to break the

Bermudan call option into the sum of a European option and an early exercise premium. To show this, assume that the allowed exercise dates are $\mathcal{D}(0) = \{T_1, T_2, \dots, T_B\}$, and let $S_B(T_i)$ be the exercise level above which the Bermudan option should be exercised at time T_i , $i = 1, \dots, B$. Notice that if at time T_i we have $S(T_i) > S_B(T_i)$, then C_B will *jump down* in value when time progresses past time T_i , as a reflection of the missed exercise opportunity. Indeed, in the earlier notation of hold and exercise values, we have

$$\begin{aligned} C_B(T_i) &= \max(U(T_i), H(T_i)), \\ C_B(T_i+) &= H(T_i), \end{aligned}$$

which makes the jump in value evident. Given the existence of these jumps, we may write

$$\begin{aligned} dC_B(t) &= rC_B(t) dt + dM(t) \\ &\quad + \sum_{i=1}^B 1_{\{S(T_i) > S_B(T_i)\}} \delta(T_i - t) (H(T_i) - U(T_i)) dt, \end{aligned}$$

where $H(T_i) = C_B(T_i+)$ is the hold value at time T_i , $U(T_i) = S(T_i) - K$, and $M(t)$ is a martingale,

$$dM(t) = \frac{\partial C_B(t)}{\partial S} (r - q) S(t) dW^\beta(t).$$

Deflating C_B by the money market account and forming expectations, we get, since $c(t) = e^{-r(T-t)} \mathbb{E}_t^Q(C_B(T))$,

$$C_B(t) = c(t) + \sum_{T_i \geq t} e^{-r(T_i-t)} \mathbb{E}_t^Q \left(1_{\{S(T_i) > S_B(T_i)\}} (U(T_i) - H(T_i)) \right).$$

As $H(T_i)$ must be less than the exercise value $U(T_i)$ whenever $S(T_i) > S_B(T_i)$ we can simplify this expression to the following result that we, in Section 18.2.3, call the *marginal exercise value decomposition*.

Proposition 1.10.9. *The Bermudan option price $C_B(t)$ satisfies*

$$C_B(t) = c(t) + E_B(t), \quad t \leq T,$$

where the (Bermudan) early exercise premium $E_B(t)$ is defined as

$$E_B(t) = \sum_{T_i \geq t} e^{-r(T_i-t)} \mathbb{E}_t^Q \left((U(T_i) - H(T_i))^+ \right),$$

with $T_1 < T_2 < \dots < T_B = T$ being the set of exercise dates.

As shown in Section 18.2.3, the result in Proposition 1.10.9 may be extended to more complicated processes and payouts than those considered here.

Finite Difference Methods

In Chapter 1 we described how the pricing of a derivative security typically requires either the solution of a parabolic partial differential equation (PDE) or the evaluation of an expectation of a random variable. In realistic applications, both of these price formulations often do not allow for closed-form solution, in which case we must resort to either analytical approximations or, more generally, numerical techniques. In the next two chapters we will describe a number of numerical algorithms useful in derivatives pricing. Analytical approximations will receive ample treatment later in this book, in the context of specific problems.

Our treatment of numerical methods is broken into two main subjects. In this chapter, we cover finite difference solutions of PDEs; and in Chapter 3 we turn to Monte Carlo evaluation of expectations. Many excellent specialist books exist on both topics, including Mitchell and Griffiths [1980], Tavella and Randall [2000], and Glasserman [2004]; our treatment only surveys the most important concepts, as required for our needs in this book. We do provide, however, a number of schemes rarely described in detail in the finance literature and also supplement our analysis with a number of “tricks of the trade”, particularly in the application of finite difference grids.

The analysis of numerical PDE solutions in this chapter is arranged in two blocks. First, in Sections 2.1–2.8 we study the basic mechanics of the finite difference grid method for one-dimensional PDEs. Subsequently, Sections 2.9–2.12 then apply operator splitting techniques to extend the finite difference method to PDE of dimensions two and higher. The analysis culminates with a presentation of *ADI schemes* for multi-dimensional PDEs with mixed partial derivatives.

2.1 1-Dimensional PDEs: Problem Formulation

Initially, we will consider the numerical solution of the general one-dimensional terminal value PDE problem

$$\frac{\partial V}{\partial t} + \mathcal{L}V = 0, \quad (2.1)$$

where \mathcal{L} is the operator

$$\mathcal{L} = \mu(t, x) \frac{\partial}{\partial x} + \frac{1}{2}\sigma(t, x)^2 \frac{\partial^2}{\partial x^2} - r(t, x),$$

and where $V = V(t, x)$ satisfies a terminal condition $V(T, x) = g(x)$. We recognize the PDE as being an extension of the Black-Scholes PDE (1.47) to general time- and state-dependent drift (μ), volatility (σ), and interest rate (r). Underneath the PDE lies a physical model where a state variable process $x(\cdot)$ follows an SDE of the form

$$dx(t) = \mu(t, x(t)) dt + \sigma(t, x(t)) dW(t) \quad (2.2)$$

where $W(t)$ is a Brownian motion in the risk-neutral probability measure Q . Let the range of values attainable by $x(t)$ on $t \in [0, T]$ be denoted $\mathcal{B} \subseteq \mathbb{R}$, and assume that the functions $\mu, \sigma, r : [0, T] \times \mathcal{B} \rightarrow \mathbb{R}$ are sufficiently regular to make (2.1) and (2.2) meaningful (see Chapter 1).

The terminal value problem above is, as discussed earlier, a *Cauchy problem* to be solved for $V(t, x)$ on $(t, x) \in [0, T] \times \mathcal{B}$. In many cases of practical interest, further boundary conditions are applied in the spatial (x) domain. If such boundary conditions are expressed directly in terms of V (rather than its derivatives) we have a *Dirichlet boundary problem*. For instance, a so-called *up-and-out barrier option* will pay out $g(x(T))$ at time T if and only if $x(t)$ stays strictly below a contractually specified barrier level H at all times $t \leq T$. If, on the other hand, $x(t)$ touches H at any time during the life of the contract, it will expire worthless (or “knock out”). In this case, the PDE is only to be solved on $(t, x) \in [0, T] \times (\mathcal{B} \cap (-\infty, H))$ and is subject to the Dirichlet boundary condition

$$V(t, H) = 0, \quad t \in [0, T],$$

which expresses that the option has no value for $x \geq H$. We note that it is not uncommon to encounter options where the spatial domain boundaries are functions of time, a situation we shall deal with in Section 2.7.1. Also, as we shall see shortly, sometimes boundary conditions are conveniently expressed in terms of derivatives of V .

For numerical solution of the PDE (2.1), we often need to assume that the domain of the state variable x is finite, even in situations where (2.1) is supposed to hold for an infinite domain. Suitable truncation of the domain can often be done probabilistically, based on a confidence interval for $x(T)$. To illustrate the procedure, consider the Black-Scholes PDE (1.47) applied to a call option with strike K . A common first step is to use the transformation $x = \ln S$, such that the PDE has constant coefficients,

$$\frac{\partial V}{\partial t} + \left(r - \frac{1}{2}\sigma^2\right) \frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} - rV = 0, \quad (2.3)$$

with terminal value (for a call option) $V(T, x) = (e^x - K)^+$. The domain of x is here the entire real line, $\mathcal{B} = \mathbb{R}$. We know (from (1.39)) that

$$x(T) = x(0) + \left(r - \frac{1}{2}\sigma^2\right)T + \sigma(W(T) - W(0)), \quad (2.4)$$

which is a Gaussian random variable with mean $\bar{x} = x(0) + (r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$. Consider now replacing the domain $(-\infty, \infty)$ with the finite interval $[\bar{x} - \alpha\sigma\sqrt{T}, \bar{x} + \alpha\sigma\sqrt{T}]$ for some positive constant α . The likelihood of $x(T)$ falling outside of this interval is easily seen to be $2\Phi(-\alpha)$ (where, as always, $\Phi(z)$ is the standard Gaussian cumulative distribution function). If, say, we set α to 4, $2\Phi(-4) = 6.3 \times 10^{-5}$, which is an insignificant probability for most applications. Larger (smaller) values of α will make the truncation error smaller (larger) and will ultimately require more (less) effort in a numerical scheme. We recommend values of α somewhere between 3 and 5 for most applications. For the Black-Scholes case, a rigorous estimate of the error imposed by domain truncation is given in Kangro and Nicolaides [2000].

In many cases of practical interest, it is not possible to write down an exact confidence interval for $x(T)$. In such cases, one instead may use an approximate confidence interval, found by, for instance, using “average” values for $\mu(t, x)$ and $\sigma(t, x)$. High precision in these estimates is typically not needed.

2.2 Finite Difference Discretization

In order to solve the PDE (2.1) numerically, we now wish to discretize it on the rectangular domain $(t, x) \in [0, T] \times [\underline{M}, \overline{M}]$, where \overline{M} and \underline{M} are finite constants, possibly found by a truncation procedure such as the one outlined above. We first introduce two equidistant¹ grids $\{t_i\}_{i=0}^n$ and $\{x_j\}_{j=0}^{m+1}$ where $t_i = iT/n \triangleq i\Delta_t$, $i = 0, 1, \dots, n$, and $x_j = \underline{M} + j(\overline{M} - \underline{M})/(m+1) \triangleq \underline{M} + j\Delta_x$, $j = 0, 1, \dots, m+1$. The terminal value $V(T, x) = g(x)$ is imposed at $t_n = T$, and spatial boundary conditions are imposed at x_0 and x_{m+1} .

2.2.1 Discretization in x -Direction. Dirichlet Boundary Conditions

We first focus on the spatial operator \mathcal{L} and restrict x to take values in the interior of the spatial grid $x \in \{x_j\}_{j=1}^m$. Consider replacing the first- and second-order partial derivatives with first- and second-order difference operators:

¹Non-equidistant grids are often required in practice and will be covered in Section 2.4.

$$\delta_x V(t, x_j) \triangleq \frac{V(t, x_{j+1}) - V(t, x_{j-1})}{2\Delta_x}, \quad (2.5)$$

$$\delta_{xx} V(t, x_j) \triangleq \frac{V(t, x_{j+1}) + V(t, x_{j-1}) - 2V(t, x_j)}{\Delta_x^2}. \quad (2.6)$$

These operators are accurate to second order. Formally²,

Lemma 2.2.1.

$$\delta_x V(t, x_j) = \frac{\partial V(t, x_j)}{\partial x} + O(\Delta_x^2),$$

$$\delta_{xx} V(t, x_j) = \frac{\partial^2 V(t, x_j)}{\partial x^2} + O(\Delta_x^2).$$

Proof. A Taylor expansion of $V(t, x)$ around the point $x = x_j$ gives

$$\begin{aligned} V(t, x_{j+1}) &= V(t, x_j) + \Delta_x \frac{\partial V(t, x_j)}{\partial x} \\ &\quad + \frac{1}{2} \Delta_x^2 \frac{\partial^2 V(t, x_j)}{\partial x^2} + \frac{1}{6} \Delta_x^3 \frac{\partial^3 V(t, x_j)}{\partial x^3} + O(\Delta_x^4), \end{aligned}$$

and

$$\begin{aligned} V(t, x_{j-1}) &= V(t, x_j) - \Delta_x \frac{\partial V(t, x_j)}{\partial x} \\ &\quad + \frac{1}{2} \Delta_x^2 \frac{\partial^2 V(t, x_j)}{\partial x^2} - \frac{1}{6} \Delta_x^3 \frac{\partial^3 V(t, x_j)}{\partial x^3} + O(\Delta_x^4). \end{aligned}$$

Insertion of these expressions into (2.5) and (2.6) gives the desired result.

□

In other words, if we introduce the discrete operator

$$\widehat{\mathcal{L}} = \mu(t, x)\delta_x + \frac{1}{2}\sigma(t, x)^2\delta_{xx} - r(t, x),$$

we have, for $x \in \{x_j\}_{j=1}^m$,

$$\mathcal{L}V(t, x) = \widehat{\mathcal{L}}V(t, x) + O(\Delta_x^2).$$

With attention restricted to values on the grid $\{x_j\}_{j=1}^m$, we can view $\widehat{\mathcal{L}}$ as a matrix, once we specify the side boundary conditions at x_0 and x_{m+1} . For the Dirichlet case, assume for instance that

$$V(x_0, t) = \underline{f}(t, x_0), \quad V(x_{m+1}, t) = \overline{f}(t, x_{m+1}),$$

²Recall that a function $f(h)$ is of order $O(e(h))$ if $|f(h)|/|e(h)|$ is bounded from above by a positive constant in the limit $h \rightarrow 0$.

for given functions $\underline{f}, \bar{f} : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$. With³ $\mathbf{V}(t) \triangleq (V(t, x_1), \dots, V(t, x_m))^\top$ and, for $j = 1, \dots, m$,

$$c_j(t) \triangleq -\sigma(t, x_j)^2 \Delta_x^{-2} - r(t, x_j), \quad (2.7)$$

$$u_j(t) \triangleq \frac{1}{2}\mu(t, x_j)\Delta_x^{-1} + \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}, \quad (2.8)$$

$$l_j(t) \triangleq -\frac{1}{2}\mu(t, x_j)\Delta_x^{-1} + \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}, \quad (2.9)$$

we can write

$$\hat{\mathcal{L}}\mathbf{V}(t) = \mathbf{A}(t)\mathbf{V}(t) + \boldsymbol{\Omega}(t), \quad (2.10)$$

where \mathbf{A} is a *tri-diagonal matrix*

$$\mathbf{A}(t) = \begin{pmatrix} c_1(t) & u_1(t) & 0 & 0 & 0 & \dots & 0 \\ l_2(t) & c_2(t) & u_2(t) & 0 & 0 & \dots & 0 \\ 0 & l_3(t) & c_3(t) & u_3(t) & 0 & \dots & 0 \\ 0 & 0 & l_4(t) & c_4(t) & u_4(t) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & l_{m-1}(t) & c_{m-1}(t) & u_{m-1}(t) \\ 0 & 0 & 0 & 0 & 0 & l_m(t) & c_m(t) \end{pmatrix} \quad (2.11)$$

and $\boldsymbol{\Omega}(t)$ is a vector containing boundary values

$$\boldsymbol{\Omega}(t) = \begin{pmatrix} l_1(t)\underline{f}(t, x_0) \\ 0 \\ \vdots \\ 0 \\ u_m(t)\bar{f}(t, x_{m+1}) \end{pmatrix}.$$

As discussed earlier, sometimes one or both of the functions \bar{f} and \underline{f} are explicitly imposed as part of the option specification (as is the case for a knock-out options). In other cases, asymptotics may be necessary to establish these functions. For instance, for the case of a simple call option on a stock paying no dividends, we can set

$$\begin{aligned} \bar{f}(t, x) &= e^x - Ke^{-r(T-t)}, \\ \underline{f}(t, x) &= 0, \end{aligned}$$

where we, as before, have set $x = \ln S$ (S being the stock price) and assumed that the strike K is positive. The result for \underline{f} is obvious; the result for \bar{f} follows from the fact that a deep in-the-money call option will almost certainly pay at maturity the stock (the present value of which is just $S = e^x$) minus the strike (the present value of which is $Ke^{-r(T-t)}$).

³For clarity, this chapter uses boldface type for all vectors and matrices.

2.2.2 Other Boundary Conditions

Deriving asymptotic Dirichlet conditions can be quite involved for complicated option payouts and is often inconvenient in implementations. Rather than having to perform an asymptotic analysis for each and every type of option payout, it would be preferable to have a general-purpose mechanism for specifying the boundary condition. One common idea involves making assumptions on the form of the functional dependency between V and x at the grid boundaries, often from specification of relationships between spatial derivatives. For instance, if we impose the condition that the second derivative of V is zero at the upper boundary (x_{m+1}) — that is, V is a linear function of x — we can write (effectively using a downward discretization of the second derivative)

$$\frac{V(t, x_{m+1}) + V(t, x_{m-1}) - 2V(t, x_m)}{\Delta_x^2} = 0 \\ \Rightarrow V(t, x_{m+1}) = 2V(t, x_m) - V(t, x_{m-1}).$$

A similar assumption at the lower spatial boundary yields

$$V(t, x_0) = 2V(t, x_1) - V(t, x_2).$$

For PDEs discretized in the logarithm of some asset, it may be more natural to assume that $V(t, x) \propto e^x$ at the boundaries; equivalently, we can assume that $\partial V / \partial x = \partial^2 V / \partial x^2$ at the boundary. When discretized in downward fashion at the upper boundary (x_{m+1}), this implies that

$$\frac{V(t, x_{m+1}) - V(t, x_m)}{\Delta_x} = \frac{V(t, x_{m+1}) + V(t, x_{m-1}) - 2V(t, x_m)}{\Delta_x^2}$$

or (assuming that $\Delta_x \neq 1$)

$$V(t, x_{m+1}) = V(t, x_{m-1}) \frac{1}{\Delta_x - 1} + V(t, x_m) \frac{\Delta_x - 2}{\Delta_x - 1}.$$

Similarly,

$$V(t, x_0) = V(t, x_1) \frac{2 + \Delta_x}{1 + \Delta_x} - V(t, x_2) \frac{1}{\Delta_x + 1}.$$

Common for both methods above — and for the Dirichlet specification discussed earlier — is that they give rise to boundary specifications through simple linear systems of the general form

$$V(t, x_{m+1}) = k_m(t)V(t, x_m) + k_{m-1}(t)V(t, x_{m-1}) + \bar{f}(t, x_{m+1}), \quad (2.12)$$

$$V(t, x_0) = k_1(t)V(t, x_1) + k_2(t)V(t, x_2) + \underline{f}(t, x_0). \quad (2.13)$$

This boundary specification can be captured in the matrix system (2.10) by simply rewriting a few components of $\mathbf{A}(t)$; specifically, we must set

$$\begin{aligned}
c_m(t) &= -\sigma(t, x_m)^2 \Delta_x^{-2} - r(t, x_m) + k_m(t) u_m(t), \\
l_m(t) &= -\frac{1}{2} \mu(t, x_m) \Delta_x^{-1} + \frac{1}{2} \sigma(t, x_m)^2 \Delta_x^{-2} + k_{m-1}(t) u_m(t), \\
c_1(t) &= -\sigma(t, x_j)^2 \Delta_x^{-2} - r(t, x_j) + k_1(t) l_1(t), \\
u_1(t) &= \frac{1}{2} \mu(t, x_1) \Delta_x^{-1} + \frac{1}{2} \sigma(t, x_1)^2 \Delta_x^{-2} + k_2(t) l_1(t).
\end{aligned}$$

All other components of \mathbf{A} remain as in (2.11); note that \mathbf{A} remains tri-diagonal.

An alternative approach to specification of boundary conditions in the x -domain involves using the PDE itself to determine the boundary conditions, through replacement of all central difference operators with one-sided differences at the boundaries. Section 10.1.5.2 contains a detailed example of this idea; ultimately, this approach leads to boundary conditions that can also be written in the form (2.12)–(2.13).

2.2.3 Time-Discretization

To simplify notation, assume for now that $\Omega(t) = 0$ for all t , as will be the case if, say, we use the linear or linear-exponential boundary conditions outlined earlier. On the spatial grid, our original PDE can be written

$$\frac{\partial \mathbf{V}(t)}{\partial t} = -\mathbf{A}(t)\mathbf{V}(t) + O(\Delta_x^2)$$

which, ignoring the error term⁴, defines a system of coupled ordinary differential equations (ODEs).

A number of methods are available for the numerical solution of coupled ODEs; see, e.g., Press et al. [1992]. We here only consider basic two-level time-stepping schemes, where grid computations at time t_i involve only PDE values at times t_i and t_{i+1} . Focusing the attention on a particular bucket $[t_i, t_{i+1}]$, the choice for the finite difference approximation of $\partial V / \partial t$ is obvious:

$$\frac{\partial \mathbf{V}}{\partial t} \approx \frac{\mathbf{V}(t_{i+1}) - \mathbf{V}(t_i)}{\Delta_t}.$$

Not so obvious, however, is to which time in the interval $[t_i, t_{i+1}]$ we should associate this derivative. To be general, consider picking a time $t_i^{i+1}(\theta) \in [t_i, t_{i+1}]$, given by

$$t_i^{i+1}(\theta) = (1 - \theta)t_{i+1} + \theta t_i, \quad (2.14)$$

where $\theta \in [0, 1]$ is a parameter. We then write

$$\frac{\partial \mathbf{V}(t_i^{i+1}(\theta))}{\partial t} \approx \frac{\mathbf{V}(t_{i+1}) - \mathbf{V}(t_i)}{\Delta_t}.$$

⁴Note that the error term $O(\Delta_x^2)$ is here to be interpreted as an m -dimensional vector. We will use such short-hand notation throughout this chapter.

By a Taylor expansion, it is easy to see that this expression is first-order accurate in the time step when $\theta \neq \frac{1}{2}$, and second-order accurate when $\theta = \frac{1}{2}$. Written compactly,

$$\frac{\partial \mathbf{V}(t_i^{i+1}(\theta))}{\partial t} = \frac{\mathbf{V}(t_{i+1}) - \mathbf{V}(t_i)}{\Delta_t} + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t) + O(\Delta_t^2). \quad (2.15)$$

This result on the convergence order is intuitive since only in the case $\theta = \frac{1}{2}$ is the difference coefficient precisely central; for all other cases, the difference coefficient is either predominantly backward in time or predominantly forward in time.

The time-discretization technique introduced above is known as a *theta scheme*. The special cases of $\theta = 1$, $\theta = 0$, and $\theta = \frac{1}{2}$ are known as the *fully implicit scheme*, the *fully explicit scheme*, and the *Crank-Nicolson scheme*, respectively. In light of the convergence result (2.15), one may wonder why anything other than the Crank-Nicolson scheme is ever used. The CN method is, indeed, often the method of choice, but there are situations where a straight application of the Crank-Nicolson scheme can lead to oscillations in the numerical solution or its spatial derivatives. Judicial application of the fully implicit method can often alleviate these problems, as we shall discuss later. The fully explicit method should never be used due to poor convergence and stability properties (see Section 2.3), but has nevertheless managed to survive in a surprisingly large number of finance texts and papers.

2.2.4 Finite Difference Scheme

We now proceed to combine the discretizations (2.10) and (2.15) into a complete finite difference scheme. First, we expand

$$\begin{aligned} \mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_i^{i+1}(\theta)) &= \theta \mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_i) \\ &\quad + (1 - \theta) \mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_{i+1}) + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t) + O(\Delta_t^2), \end{aligned}$$

such that our PDE can be represented as

$$\begin{aligned} \frac{\mathbf{V}(t_{i+1}) - \mathbf{V}(t_i)}{\Delta_t} + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t) + O(\Delta_t^2) \\ &= -\mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_i^{i+1}(\theta)) + O(\Delta_x^2) \\ &= -\theta \mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_i) - (1 - \theta) \mathbf{A}(t_i^{i+1}(\theta)) \mathbf{V}(t_{i+1}) \\ &\quad + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t) + O(\Delta_t^2) + O(\Delta_x^2). \end{aligned}$$

Multiplying through with Δ_t gives rise to the complete finite difference representation of the PDE solution at times t_i and t_{i+1} :

Proposition 2.2.2. *On the grid $\{x_j\}_{j=1}^m$, the solution to (2.1) at times t_i and t_{i+1} is characterized by*

$$(\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \mathbf{V}(t_i) = (\mathbf{I} + (1 - \theta) \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \mathbf{V}(t_{i+1}) + e_i^{i+1}, \quad (2.16)$$

where \mathbf{I} is the $m \times m$ identity matrix, and e_i^{i+1} is an error term

$$e_i^{i+1} = \Delta_t O(\Delta_x^2) + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t^2) + O(\Delta_t^3). \quad (2.17)$$

Let $\widehat{\mathbf{V}}(t_i, x_j)$ denote the approximation to the true solution $V(t_i, x_j)$ obtained by using (2.16) without the error term. Defining

$$\widehat{\mathbf{V}}(t) = \left(\widehat{V}(t, x_1), \dots, \widehat{V}(t, x_m) \right)^\top,$$

we have

$$(\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \widehat{\mathbf{V}}(t_i) = (\mathbf{I} + (1 - \theta) \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \widehat{\mathbf{V}}(t_{i+1}). \quad (2.18)$$

For a known value of $\widehat{\mathbf{V}}(t_{i+1})$, (2.18) defines a simple linear system of equations that can be solved for $\widehat{\mathbf{V}}(t_i)$ by standard methods. Simplifying matters is the fact that the matrix $(\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i^{i+1}(\theta)))$ is tri-diagonal, allowing us to solve (2.18) in only $O(m)$ operations; see Press et al. [1992] for an algorithm⁵.

Starting from the prescribed terminal condition $V(t_n, x_j) = g(x_j)$, $j = 1, \dots, m$, we can now use (2.18) to iteratively step backward in time until we ultimately recover $\widehat{\mathbf{V}}(0)$. This procedure is known as *backward induction*.

Proposition 2.2.3. *The theta scheme (2.18) recovers $\widehat{\mathbf{V}}(0)$ in $O(mn)$ operations. If the scheme converges, the error on $\widehat{\mathbf{V}}(0)$ compared to the exact solution $\mathbf{V}(0)$ is of order*

$$O(\Delta_x^2) + 1_{\{\theta \neq \frac{1}{2}\}} O(\Delta_t) + O(\Delta_t^2).$$

Proof. The backward induction algorithm requires the solution of n tri-diagonal systems, one per time step, for a total computational cost of $O(mn)$. The local truncation error on $\widehat{\mathbf{V}}(t_i)$ is e_i^{i+1} , making the global truncation error after n time steps of order ne_i^{i+1} . Combining (2.17) with the fact that $n = T/\Delta_t = O(\Delta_t^{-1})$ gives the order result listed in the proposition. \square

⁵The special case of an explicit scheme ($\theta = 0$) provides us with a direct expression for $V(t_i, x_j)$ in terms of $V(t_{i+1}, x_{j-1})$, $V(t_{i+1}, x_j)$, and $V(t_{i+1}, x_{j+1})$, a scheme that is easily visualized as a “trinomial tree”. The intuitive nature of the explicit scheme coupled with the fact that no matrix equation must be solved may explain the popularity of this scheme in the finance literature, despite its poor numerical qualities (see Section 2.3). We stress that the workload of the explicit scheme is still $O(m)$ per time step, as is the case for all theta schemes.

It follows from Proposition 2.2.3 that the Crank-Nicolson scheme is second-order convergent in the time step, and all other theta schemes are first-order convergent in the time step. All theta-schemes are second-order convergent in the spatial step Δ_x .

In deriving (2.18), we assumed earlier that the boundary vector was zero, $\Omega(t) = 0$. Including a non-zero boundary vector into the scheme is, however, straightforward and results in a time-stepping scheme of the form

$$\begin{aligned} (\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \hat{\mathbf{V}}(t_i) &= (\mathbf{I} + (1 - \theta) \Delta_t \mathbf{A}(t_i^{i+1}(\theta))) \hat{\mathbf{V}}(t_{i+1}) \\ &\quad + (1 - \theta) \Omega(t_{i+1}) + \theta \Omega(t_i). \end{aligned} \quad (2.19)$$

Again, this system is easily solved for $\hat{\mathbf{V}}(t_i)$ by a standard tri-diagonal equation solver.

As a final point, we stress that the finite difference scheme above ultimately yields a full vector of values $\hat{\mathbf{V}}(0)$ at time 0, with one element per value of x_j , $j = 1, \dots, m$. In general, we are mainly interested in $V(0, x(0))$, where $x(0)$ is the known value of x at time 0. There is no need to include $x(0)$ in the grid, as we can simply employ an interpolator (e.g., a cubic spline) on this vector $\hat{\mathbf{V}}(0)$ to compute $V(0, x(0))$. Clearly, such an interpolator should be at least second-order accurate to avoid interfering with the overall $O(\Delta_x^2)$ convergence of the finite difference scheme. Assuming the interpolator is sufficiently smooth, we can also use it to compute various partial derivatives with respect to x that we may be interested in. Alternatively, these can be computed by the same type of finite difference coefficients discussed in Section 2.2.1. The derivative $\partial V(0, x(0))/\partial t$ — the *time decay* — can be picked up from the grid in the same fashion.

Remark 2.2.4. The scheme (2.18) may, without affecting convergence order, be replaced with

$$(\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i)) \hat{\mathbf{V}}(t_i) = (\mathbf{I} + (1 - \theta) \Delta_t \mathbf{A}(t_{i+1})) \hat{\mathbf{V}}(t_{i+1}).$$

2.3 Stability

2.3.1 Matrix Methods

Ignoring the contributions from boundary conditions, the finite difference scheme developed in the previous section can be rewritten

$$\hat{\mathbf{V}}(t_i) = \mathbf{B}_i^{i+1} \hat{\mathbf{V}}(t_{i+1}), \quad (2.20)$$

where

$$\mathbf{B}_i^{i+1} \triangleq (\mathbf{I} - \theta \Delta_t \mathbf{A}(t_i^{i+1}(\theta)))^{-1} (\mathbf{I} + (1 - \theta) \Delta_t \mathbf{A}(t_i^{i+1}(\theta))).$$

That is, for any $0 \leq k < n$,

$$\widehat{\mathbf{V}}(t_k) = \mathbf{B}_k^n \widehat{\mathbf{V}}(t_n), \quad \mathbf{B}_k^n \triangleq \mathbf{B}_k^{k+1} \mathbf{B}_{k+1}^{k+2} \dots \mathbf{B}_{n-1}^n.$$

We say that the scheme is *stable* if $|\widehat{\mathbf{V}}(t_k)|$ is bounded for all $0 \leq k < n$. Assuming $|\widehat{\mathbf{V}}(T)| < \infty$, a necessary and sufficient condition for stability is that there exists a constant K such that for all $0 \leq k < n$

$$|\mathbf{B}_k^n| \leq K, \quad (2.21)$$

where $|\cdot|$ is any matrix norm, e.g. the spectral norm or the infinity norm⁶. See Mitchell and Griffiths [1980] for further details.

2.3.2 Von Neumann Analysis

For simple problems with time- and space-independent coefficients, it may be possible to establish the spectral norm of \mathbf{B}_k^n by direct methods (see e.g. Mitchell and Griffiths [1980], Kraaijevanger et al. [1987], Lenferink and Spijker [1991], Spijker and Straetemans [1997]), but generally the stability criterion (2.21) is difficult to evaluate. While certain somewhat simpler matrix-based methods exist to establish necessary conditions for stability (again, see Mitchell and Griffiths [1980]), we shall here only consider a “local” method, known as the *von Neumann method*. In principle, the von Neumann method only holds for finite difference schemes where the underlying PDE has constant coefficients, but there is much numerical evidence to support wider application⁷. The von Neumann method does not directly consider the effect of boundary conditions on stability, but (for constant coefficient problems) provides a necessary condition for stability irrespective of the type of boundary condition.

The basis for the von Neumann analysis is the observation that a real function sampled on a finite number of points is uniquely defined by a complex Fourier series. For our PDE solution sampled on the spatial grid, the precise result is

$$V(t_k, x_j) = \sum_l H_l(t_k) e^{-i\omega_l j \Delta_x},$$

where $H_l(t_k)$ and ω_l are the amplification factor (discrete Fourier transform) and wave number for the l -th mode, respectively. Notice that i here denotes

⁶The spectral norm of a matrix \mathbf{C} is defined as the largest absolute eigenvalue of $(\mathbf{C}^\top \mathbf{C})^{1/2}$. The infinity norm is defined as $\max_i \sum_j |C_{i,j}|$.

⁷In the application to PDEs with non-constant coefficients, it may help to think of the von Neumann analysis as being applied to the PDE locally with “frozen” coefficients, followed by an examination of the worst case among all frozen coefficients.

the imaginary unit, $i^2 = -1$, with k (momentarily) having taken the role of the time index in the finite difference grid. For the constant coefficient case, a key fact for our PDE problem is that

$$H_l(t_k) = H_l(t_{k+1})\xi_l^{-1},$$

where ξ_l is a mode-specific *amplification factor* independent of time. To determine how a solution is propagated back through the finite difference grid, it thus suffices to consider a test function of the form

$$v(t_k, x_j) = \xi(\omega)^{n-k} e^{i\omega j \Delta_x}. \quad (2.22)$$

According to the Von Neumann criterion, stability of (2.20) requires that the *modulus of the amplification factor* $\xi(\omega)$ is less or equal to one, independent of the wave number:

$$\forall \omega : |\xi(\omega)| \leq 1. \quad (2.23)$$

This criterion is natural and merely expresses that all eigenmodes should be damped, and not exponentially amplified, by the finite difference scheme.

Turning to our system (2.20), assume for simplicity that $r(t, x) = 0$. A positive interest rate (we will nearly always have $r(t, x) > 0$) introduces some extra dampening through discounting effects and will, if anything, lead to better stability properties than the case of zero interest rates. Writing $v(t_k, x_j) = v_{k,j}$, $\sigma(t_k^{k+1}(\theta), x_j) = \sigma_{k,j}$, and $\mu(t_k^{k+1}(\theta), x_j) = \mu_{k,j}$, the von Neumann analysis gives the following result:

Proposition 2.3.1. Define $\alpha = \Delta_t / (\Delta_x)^2$. For (2.20) with $r(t, x) = 0$, the von Neumann stability criterion is

$$1 \geq \theta \geq \frac{1}{2} - \frac{1}{\alpha} \left(\frac{\sigma_{k,j}^2}{\sigma_{k,j}^4 + \mu_{k,j}^2 \Delta_x^2 + |\mu_{k,j}^2 \Delta_x^2 - \sigma_{k,j}^4|} \right), \quad (2.24)$$

to hold for all $k = 0, 1, \dots, n-1$, $j = 1, 2, \dots, m$.

Proof. Define $\varsigma_{k,j}^\pm = \sigma_{k,j}^2 \pm \Delta_x \mu_{k,j}$. A local application of (2.20) gives

$$\begin{aligned} v_{k,j-1} \left(-\frac{\alpha\theta}{2} \varsigma_{k,j}^- \right) + v_{k,j} (1 + \alpha\theta\sigma_{k,j}^2) + v_{k,j+1} \left(-\frac{\alpha\theta}{2} \varsigma_{k,j}^+ \right) = \\ v_{k+1,j} \left(\frac{\alpha(1-\theta)}{2} \varsigma_{k,j}^- \right) + v_{k+1,j} (1 - \alpha(1-\theta)\sigma_{k,j}^2) + v_{k+1,j+1} \left(\frac{\alpha(1-\theta)}{2} \varsigma_{k,j}^+ \right) \end{aligned}$$

with α defined above. Inserting (2.22) and rearranging (using Euler's formulas for sin and cos) yields

$$\xi(\omega) = \frac{1 - (1-\theta)\alpha\sigma_{k,j}^2(1 - \cos \omega \Delta_x) + i(1-\theta)\alpha\Delta_x \mu_{k,j} \sin \omega \Delta_x}{1 + \theta\alpha\sigma_{k,j}^2(1 - \cos \omega \Delta_x) - i\theta\alpha\Delta_x \mu_{k,j} \sin \omega \Delta_x}.$$

Note that ξ is a function of k and j , due to the non-constant PDE parameters. As discussed earlier (see also Mitchell and Griffiths [1980]), we expect the system to be stable if the criterion (2.23) holds for all k and j in the grid. Computing the modulus of ξ and requiring that it does not exceed one leads, after straightforward manipulations, to the stability criterion

$$\forall \omega : 2\alpha\sigma_{k,j}^2 + (2\theta - 1)\alpha^2 [\sigma_{k,j}^4 + \mu_{k,j}^2 \Delta_x^2 + \cos \omega \Delta_x (\mu_{k,j}^2 \Delta_x^2 - \sigma_{k,j}^4)] \geq 0.$$

As $\cos \omega \Delta_x \in [-1, 1]$, this expression can be simplified to (2.24). \square

From (2.24) we can immediately conclude that the finite difference scheme is always stable if $\frac{1}{2} \leq \theta \leq 1$, irrespective of the magnitudes of Δ_x and Δ_t . For $\frac{1}{2} \leq \theta \leq 1$, we therefore say that the theta scheme is *absolutely stable*, or simply *A-stable*. Both the fully implicit ($\theta = 1$) and the Crank-Nicolson ($\theta = \frac{1}{2}$) finite difference schemes are thus *A*-stable. For the explicit scheme ($\theta = 0$), however, stability is *conditional*, requiring

$$\frac{2}{\alpha}\sigma_{k,j}^2 \geq \sigma_{k,j}^4 + \mu_{k,j}^2 \Delta_x^2 + |\mu_{k,j}^2 \Delta_x^2 - \sigma_{k,j}^4|.$$

For small drifts, this expression amounts to the restriction $\sigma_{k,j}^2 \leq \Delta_x^2/\Delta_t$ which can be quite onerous, often requiring the (laborious) use of thousands of time steps in the finite difference grid. We shall not consider fully explicit methods any further in this book.

Returning to the case $\frac{1}{2} \leq \theta \leq 1$, let us introduce a stronger definition of stability. A time-stepping method is said to be *strongly A-stable* if the modulus of the amplification factor ξ is strictly below 1 for any value of the time step, including the limit⁸ $\Delta_t \rightarrow \infty$. From (2.24), we see that if $\Delta_t \rightarrow \infty$ (which implies $\alpha \rightarrow \infty$), then the modulus of the amplification factor could reach 1 in the special case of $\theta = 1/2$. In other words, the Crank-Nicolson scheme is *not* strongly *A*-stable. For large time steps, harmonics in the Crank-Nicolson finite difference solution will effectively not be damped from one time step to the next, opening up the possibility that unwanted high-frequency oscillations can creep into the numerical solution. In practice, this is primarily a problem if high-frequency eigenmodes have high amplification factors, as can happen if there is an outright discontinuity in the terminal value function g . The problem is especially noticeable if the discontinuity in the value function is “close” in both time and space to $t = 0$ and $x = x(0)$ (as would be the case for a short-dated option with a discontinuity close to the starting value of x). Oscillations can be prevented by setting the time step smaller than twice the maximum stable explicit time step (see Tavella and Randall [2000]), but this can often be computationally expensive. We shall deal with other methods to suppress oscillations in Section 2.5.

We conclude this section by noting a deep connection between the stability of a finite difference scheme and its convergence to the true solution

⁸If further $|\xi|$ approaches zero for $\Delta_t \rightarrow 0$, the scheme is said to be *L-stable*.

of the PDE as $\Delta_t \rightarrow 0$ and $\Delta_x \rightarrow 0$. First, we define a finite difference scheme to be *consistent* if local (Taylor) truncation errors approach zero for $\Delta_t \rightarrow 0$ and $\Delta_x \rightarrow 0$. All the schemes we have encountered so far are consistent. Further, define a finite difference scheme to be *convergent* if the difference between the numerical solution and the exact PDE solution at a fixed point in the domain converges to zero uniformly as $\Delta_t \rightarrow 0$ and $\Delta_x \rightarrow 0$ (not necessarily independently of each other). We then have

Theorem 2.3.2 (Lax Equivalence Theorem). *For a well-posed⁹ linear terminal value PDE, a consistent 2-level finite difference scheme is convergent if and only if it is stable.*

A more precise statement of the above result, as well as a proof, can be found in Mitchell and Griffiths [1980].

2.4 Non-Equidistant Discretization

In practice, we often wish to align the finite difference grid to particular dates (e.g., those on which a coupon or a dividend is paid) and particular values of x (e.g., those on which strikes and barriers are positioned). Also, for numerical reasons we may want to make certain important parts of the finite difference grid more densely spaced to concentrate computational effort on domains of particular importance to the solution of the PDE. To do so, we will now relax our earlier assumption of equidistant discretization in time and space. Doing so for the time domain is actually trivial and merely requires us to replace Δ_t in (2.18) with $\Delta_{t,i} \triangleq t_{i+1} - t_i$, where the spacing of the time grid $\{t_i\}_{i=0}^n$ is now no longer constant. The backward induction algorithm can proceed as before. We note that the ability to freely select the time grid will allow us to line up perfectly with dates that carry high significance for the product in question (e.g. dates on which cash flows take place, see Section 2.7.3) or to, say, use coarser time steps for the part of the finite difference grid that is far in the future. For an adaptive algorithm to automatically select the time-step, see d'Halluin et al. [2001].

For the spatial step, we have a number of options to induce non-equidistant spacing. One method involves a non-linear change of variables $y = h(x)$ in the PDE, followed by a regular equidistant discretization in the new variable y . This maps into a non-equidistant discretization in x which, provided that $h(\cdot)$ is chosen carefully, will have the desired geometry. Discussion of this method along with guidelines for choosing $h(\cdot)$ can be found in Chapter 5 of Tavella and Randall [2000]. We will here pursue a more direct alternative, where we simply introduce an irregular grid $\{x_j\}_{j=0}^{m+1}$

⁹Well-posed means that the PDE we are solving has a unique solution that depends continuously on the problem data (PDE coefficients, domain, boundary conditions, etc.)

and redefine the finite difference operators (2.5)–(2.6) to achieve maximum precision. For this, define

$$\Delta_{x,j}^+ \triangleq x_{j+1} - x_j, \quad \Delta_{x,j}^- \triangleq x_j - x_{j-1},$$

and set

$$\delta_x^+ V(t, x_j) = \frac{V(t, x_{j+1}) - V(t, x_j)}{\Delta_{x,j}^+}, \quad \delta_x^- V(t, x_j) = \frac{V(t, x_j) - V(t, x_{j-1})}{\Delta_{x,j}^-}.$$

By a Taylor expansion, we get

$$\begin{aligned} \delta_x^+ V(t, x_j) &= \frac{\partial V(t, x_j)}{\partial x} + \frac{1}{2} \frac{\partial^2 V(t, x_j)}{\partial x^2} \Delta_{x,j}^+ \\ &\quad + \frac{1}{6} \frac{\partial^3 V(t, x_j)}{\partial x^3} (\Delta_{x,j}^+)^2 + O((\Delta_{x,j}^+)^3), \end{aligned} \quad (2.25)$$

$$\begin{aligned} \delta_x^- V(t, x_j) &= \frac{\partial V(t, x_j)}{\partial x} - \frac{1}{2} \frac{\partial^2 V(t, x_j)}{\partial x^2} \Delta_{x,j}^- \\ &\quad + \frac{1}{6} \frac{\partial^3 V(t, x_j)}{\partial x^3} (\Delta_{x,j}^-)^2 + O((\Delta_{x,j}^-)^3). \end{aligned} \quad (2.26)$$

Maximum accuracy on the first-order derivative approximation is achieved by selecting a weighted combination of (2.25)–(2.26) such that the terms of order $O(\Delta_{x,j}^+)$ and $O(\Delta_{x,j}^-)$ cancel. That is, we set

$$\begin{aligned} \delta_x V(t, x_j) &= \frac{\Delta_{x,j}^-}{\Delta_{x,j}^+ + \Delta_{x,j}^-} \cdot \delta_x^+ V(t, x_j) + \frac{\Delta_{x,j}^+}{\Delta_{x,j}^+ + \Delta_{x,j}^-} \cdot \delta_x^- V(t, x_j) \quad (2.27) \\ &= \frac{\partial V(t, x_j)}{\partial x} + O\left(\frac{(\Delta_{x,j}^+)^2 \Delta_{x,j}^- + (\Delta_{x,j}^-)^2 \Delta_{x,j}^+}{\Delta_{x,j}^+ + \Delta_{x,j}^-}\right) \end{aligned}$$

which is second-order accurate, in the sense that reducing both $\Delta_{x,j}^+$ and $\Delta_{x,j}^-$ by a factor of k will reduce the error by a factor of k^2 . To estimate the derivative $\partial^2 V(t, x_j)/\partial x^2$ we set

$$\begin{aligned} \delta_{xx} V(t, x_j) &= \frac{\delta_x^+ V(t, x_j) - \delta_x^- V(t, x_j)}{\frac{1}{2} (\Delta_{x,j}^+ + \Delta_{x,j}^-)} \quad (2.28) \\ &= \frac{\partial^2 V(t, x_j)}{\partial x^2} + O\left(\frac{(\Delta_{x,j}^+)^2 - (\Delta_{x,j}^-)^2}{\Delta_{x,j}^+ + \Delta_{x,j}^-} + \frac{(\Delta_{x,j}^+)^3 + (\Delta_{x,j}^-)^3}{\Delta_{x,j}^+ + \Delta_{x,j}^-}\right) \end{aligned}$$

which is only first-order accurate, unless $\Delta_{x,j}^+ = \Delta_{x,j}^-$. Despite this, the global discretization error will typically remain second-order in the spatial step, even for a non-equidistant grid. A proof of this perhaps somewhat

surprising result can be found in the monograph Axelsson and Barker [1991] on finite element methods.

Development of a theta scheme around the definitions (2.27) and (2.28) proceeds in the same way as in Section 2.2. The resulting time-stepping scheme is identical to (2.18), after a modification of the matrix \mathbf{A} . Specifically, we must simply redefine the c -, u -, and l -arrays in (2.7)–(2.9) as follows:

$$c_j(t) \triangleq \frac{\Delta_{x,j}^+ - \Delta_{x,j}^-}{\Delta_{x,j}^+ \Delta_{x,j}^-} - \frac{1}{\Delta_{x,j}^- \Delta_{x,j}^+} \sigma(t, x_j)^2 - r(t, x_j), \quad (2.29)$$

$$u_j(t) \triangleq \frac{\Delta_{x,j}^-}{(\Delta_{x,j}^+ + \Delta_{x,j}^-) \Delta_{x,j}^+} \mu(t, x_j) + \frac{1}{(\Delta_{x,j}^+ + \Delta_{x,j}^-) \Delta_{x,j}^+} \sigma(t, x_j)^2, \quad (2.30)$$

$$l_j(t) \triangleq -\frac{\Delta_{x,j}^+}{(\Delta_{x,j}^+ + \Delta_{x,j}^-) \Delta_{x,j}^-} \mu(t, x_j) + \frac{1}{(\Delta_{x,j}^+ + \Delta_{x,j}^-) \Delta_{x,j}^-} \sigma(t, x_j)^2. \quad (2.31)$$

For an example where having a non-equidistant grid is essential to the numerical performance of the scheme, see Section 9.4.3.

2.5 Smoothing and Continuity Correction

2.5.1 Crank-Nicolson Oscillation Remedies

As discussed earlier, for discontinuous terminal conditions, the Crank-Nicolson scheme may exhibit localized oscillations if the time step is too coarse relative to the spatial step. Depending on the timing and spatial position of the discontinuities, these spurious oscillations may negatively affect the computed option value or, more likely, its first (“delta”) or second (“gamma”) x -derivatives. Further, in the presence of discontinuous terminal conditions, the expected $O(\Delta_t^2)$ convergence order of the Crank-Nicolson scheme may not be realized. While $O(\Delta_t^2)$ convergence is possible without spurious oscillations in some multi-level time-stepping schemes, there is evidence that these schemes are less robust than the Crank-Nicolson scheme for many financially relevant problems, see, e.g., Windcliff et al. [2001]. Fortunately, it is relatively easy to remedy the problems in the Crank-Nicolson scheme. Specifically, a theoretical result by Rannacher [1984] shows that second-order convergence can be achieved for the Crank-Nicolson scheme, provided that two simple algorithm modifications are taken:

- The discontinuous terminal payout is least-squares (L^2) projected onto the space of linear Lagrange basis functions¹⁰.

¹⁰Recall that the linear Lagrange basis functions (also called “hat” functions) are simply small triangles given by $l_j(x) = 1_{\{x_{j-1} < x \leq x_j\}} \cdot \frac{x - x_{j-1}}{x_j - x_{j-1}} + 1_{\{x_j < x \leq x_{j+1}\}} \cdot \frac{x_{j+1} - x}{x_{j+1} - x_j}$, $j = 1, \dots, m$. For an algorithm to perform the L^2 -projection, see Pooley et al. [2003].

- Two fully implicit time steps ($\theta = 1$) are taken before we switch to Crank-Nicolson ($\theta = \frac{1}{2}$) time stepping (“Rannacher stepping”).

Both techniques effectively smoothen out the discontinuity before the Crank-Nicolson scheme is applied, dampening the problematic high-frequency modes of the numerical solution. As demonstrated in Pooley et al. [2003] (see also Giles and Carter [2006]), applying either technique in isolation will typically not suffice; both are jointly required to ensure smooth second-order convergence. That said, the application of Lagrange basis function projection may conveniently be substituted with simpler smoothing techniques, with no loss of convergence order. The usefulness of such payoff smoothing extends beyond the case of discontinuous boundary conditions, so we proceed to discuss a few common techniques next.

2.5.2 Continuity Correction

By the Shannon sampling theorem, (see Shannon [1949]) if the spectrum of $g(x)$ contains frequencies higher than $1/(2\Delta_x)$ (the *Nyquist frequency*), information is lost when we sample $g(x)$ on our mesh $\{x_j\}_{j=0}^{m+1}$. In other words, whenever $g(x)$ or its derivatives are non-smooth, we will incur a *quantization* error where important features of the payout (e.g., the discontinuity of the slope of a call option at the strike) will be lost between grid points. As the grid geometry is modified, and the location of critical points (strikes, barriers, etc.) relative to x -grid changes, the computed finite difference solution will jump back and forth in erratic fashion. This so-called *odd-even effect* will result in poor convergence and an undesirably strong dependence of the solution on the grid geometry.

One straightforward way to reduce the odd-even effect (and to smooth out the high-frequency components of the payoff) is to apply a common technique from probability theory known as a *continuity correction*. Here, we simply imagine that the value of g at a grid point x_j represents the average value of the function over the interval $[x_j - (x_j - x_{j-1})/2, x_j + (x_{j+1} - x_j)/2]$. In setting the terminal boundary value $V(T, x_j)$ we thus write

$$V(T, x_j) = \frac{1}{(x_{j+1} - x_{j-1})/2} \int_{x_j - (x_j - x_{j-1})/2}^{x_j + (x_{j+1} - x_j)/2} g(x) dx. \quad (2.32)$$

We note that this implies that $V(T, x_j) \neq g(x_j)$, unless g is linear in x . The application of continuity correction to parabolic PDE solvers was first proposed in Kreiss et al. [1970].

2.5.3 Grid Shifting

Consider the effect of using (2.32) on a *digital call option*, $g(x) = 1_{\{x > H\}}$, where the level H (the digital strike) is located between nodes x_k and

x_{k+1} . For nodes x_j , $j > k + 1$, clearly $V(T, x_j) = 1$; for nodes x_j , $j < k$, $V(T, x_j) = 0$. The smoothing algorithm will have effect only at x_k or x_{k+1} , and will set either $V(T, x_k)$ or $V(T, x_{k+1})$ to a value somewhere between 0 and 1, depending on which of x_k or x_{k+1} is closest to H . If H happens to be exactly midway between x_k or x_{k+1} , the continuity correction is seen to have no effect whatsoever.

The digital option example above gives rise to a method listed in Tavella and Randall [2000] (see also Cheuk and Vorst [1996]). Here, we simply arrange the spatial grid such that the x -values where the payoff (or its derivatives) is discontinuous are exactly midway between grid nodes. If necessary, we can use a scheme with non-equidistant grid spacing to accomplish this (see Section 2.4). Our example above shows that aligning the grid in this way will, in a loose sense, make the payoff smooth.

For digital options, the grid shifting technique can be very efficient, and such “locking” of the location of strikes and barriers relative to the spatial grid can often reduce odd-even effects even better than the continuity correction discussed earlier. To demonstrate, consider the concrete task of using a finite difference grid to price a digital call option on a stock S in the Black-Scholes model. In this case, we conveniently have a theoretical option price to compare against, since it is easily shown that the time 0 value $V(0)$ must be

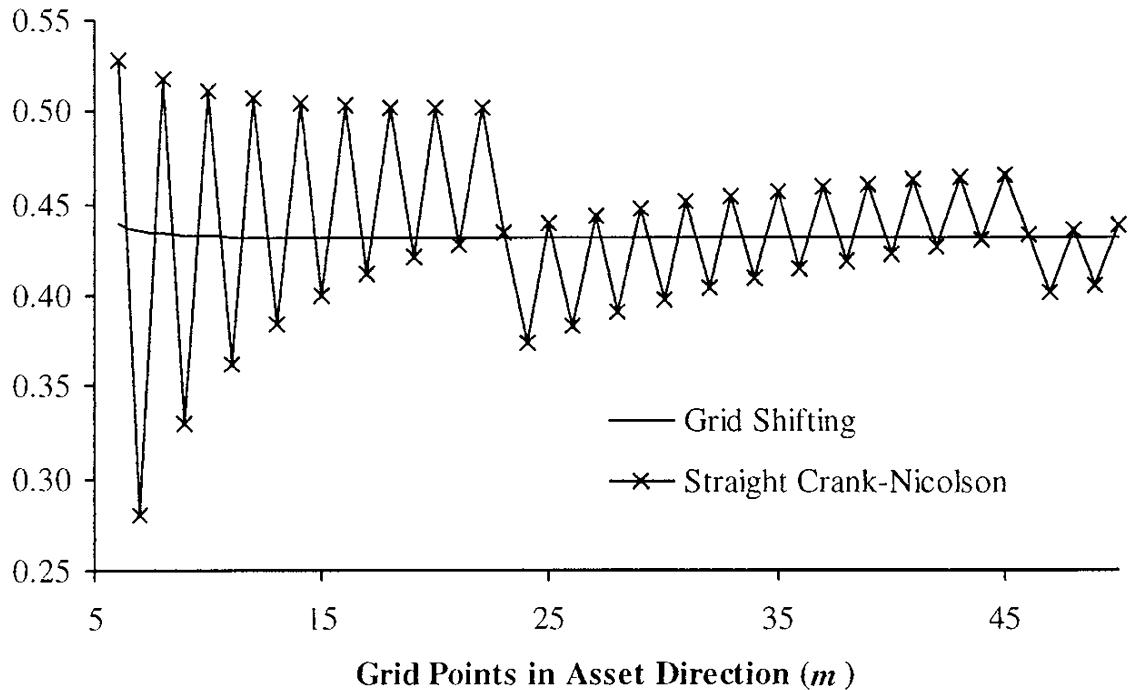
$$V(0) = e^{-rT} Q(S(T) > H) = e^{-rT} \Phi \left(\frac{\ln(S(0)/H) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} \right). \quad (2.33)$$

For our numerical work, we discretize the asset equidistantly in log-space (i.e., we work with the PDE (2.3)) and determine the spatial grid boundaries by probabilistic means using a multiplier of $\alpha = 4.5$, see Section 2.1. Spatial boundary conditions are $\partial V / \partial x = \partial^2 V / \partial x^2$, implemented as described in Section 2.2.2. In one experiment, we apply a straight Crank-Nicolson approach, with no attempt to regularize the payoff condition. In a second experiment, we combine Crank-Nicolson with Rannacher stepping and also nudge the entire spatial grid upwards until the log-barrier $\ln(H)$ is located exactly half-way between two spatial grid points. Numerical results are shown in Figure 2.1.

As Figure 2.1 shows, a naive Crank-Nicolson implementation is plagued by severe odd-even effects and very slow convergence — 100’s of spatial steps appear to be necessary before acceptable levels of the option price are reached. On the other hand, grid shifting combined with Rannacher stepping results in a perfectly smooth¹¹ convergence profile, and 5-digit price precision is here reached in less than 30 steps.

¹¹It can be verified that the convergence order in m is, as expected, close to 2 in this experiment.

Fig. 2.1. 3 Year Digital Option Price



Notes: Finite difference estimates for the Black-Scholes price of a 3 year digital option with a strike of $H = 100$. The initial asset price is $S(0) = 100$, the interest rate is $r = 0$, and the volatility is $\sigma = 20\%$. Time stepping is performed with an equidistant grid containing $n = 50$ points. Spatial discretization in log-space is equidistant, as described in the main text; the number of grid points (m) is as listed on the x -axis of the figure. The “Straight Crank-Nicolson” graph shows the convergence profile for a pure Crank-Nicolson finite difference grid. The “Grid Shifting” graph shows the convergence profile for a Crank-Nicolson finite difference grid with Rannacher stepping and a shift of the spatial grid to center $\ln(H)$ midway between two grid points. From (2.33), the theoretical value of the option is 0.4312451.

2.6 Convection-Dominated PDEs

Recall from Section 2.3 that stability of the explicit finite difference scheme requires that (omitting grid subscripts on μ and σ)

$$\frac{2\Delta_x^2}{\Delta_t} \sigma^2 \geq \sigma^4 + \mu^2 \Delta_x^2 + |\mu^2 \Delta_x^2 - \sigma^4|.$$

As discussed, this condition can be violated if Δ_t is too large relative to Δ_x . However, for fixed Δ_t and Δ_x we notice that instability can also be triggered if the absolute value of the drift μ is raised to be sufficiently large relative to the diffusion coefficient σ .

While theta schemes with $\theta \geq 1/2$ are always stable, large drifts in the PDE can nevertheless cause spurious oscillations and an overall deterioration in numerical performance of these schemes. PDEs for which this effect

occurs are said to be *convection-dominated*. To quantify matters, assume for simplicity that the finite difference grid is equidistant in the x -direction, and consider the matrix \mathbf{A} in (2.11) with tri-diagonal coefficients c , u , and l given by (2.7)–(2.9). As discussed in e.g. d’Halluin et al. [2005], spurious oscillations can occur when, for some t and some j , either $u_j(t) < 0$ or $l_j(t) < 0$. From (2.8) and (2.9), to avoid spurious oscillations we would thus need

$$\sigma(t, x_j)^2 \geq |\mu(t, x_j)|\Delta_x. \quad (2.34)$$

Intuitively, in convection-dominated systems, the central difference coefficient δ_x and δ_{xx} used to discretize the PDE can no longer fully contain the large expected up- or downward trend of the underlying process for x ; as a result, spurious oscillations can occur.

2.6.1 Upwinding

There are a number of well-established techniques to deal with convection-dominated PDEs. First, we can obviously attempt to lower Δ_x such that (2.34) is satisfied. This, however, may not be practical from a computational standpoint (and may require that Δ_t is lowered as well to avoid spurious oscillations originating from the time-stepping scheme). An alternative is to modify the first-order discrete operator δ_x such that it points in the direction of the large absolute drift. For instance, we can simply elect to use a suitably oriented one-sided difference, rather than a central difference, whenever (2.34) is violated. This procedure is known as *upstream differencing* or *upwinding*. To formalize the idea, introduce a new first-order difference operator δ_x^* given as

$$\delta_x^* V(t, x_j) = \begin{cases} \frac{1}{2} (V(t, x_{j+1}) - V(t, x_{j-1})) \Delta_x^{-1}, & |\mu(t, x_j)|\Delta_x \leq \sigma(t, x_j)^2, \\ (V(t, x_j) - V(t, x_{j-1})) \Delta_x^{-1}, & \mu(t, x_j)\Delta_x < -\sigma(t, x_j)^2, \\ (V(t, x_{j+1}) - V(t, x_j)) \Delta_x^{-1}, & \mu(t, x_j)\Delta_x > \sigma(t, x_j)^2. \end{cases}$$

Using δ_x^* instead of δ_x modifies the matrix \mathbf{A} in (2.11). Specifically, if $\mu(t, x_j)\Delta_x < -\sigma(t, x_j)^2$ we replace (2.7)–(2.9) with:

$$c_j(t) = \mu(t, x_j)\Delta_x^{-1} - \sigma(t, x_j)^2\Delta_x^{-2} - r(t, x_j), \quad (2.35)$$

$$u_j(t) = \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}, \quad (2.36)$$

$$l_j(t) = -\mu(t, x_j)\Delta_x^{-1} + \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}. \quad (2.37)$$

And when $\mu(t, x_j)\Delta_x > \sigma(t, x_j)^2$, we use

$$c_j(t) = -\mu(t, x_j)\Delta_x^{-1} - \sigma(t, x_j)^2\Delta_x^{-2} - r(t, x_j), \quad (2.38)$$

$$u_j(t) = \mu(t, x_j)\Delta_x^{-1} + \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}, \quad (2.39)$$

$$l_j(t) = \frac{1}{2}\sigma(t, x_j)^2\Delta_x^{-2}. \quad (2.40)$$

For non-equidistant grids, a similar modification to (2.29)–(2.31) is required. We omit the straightforward details.

Let us try to gain some further understanding of the upwind algorithm. Comparison of (2.35)–(2.40) with (2.7)–(2.9), shows that upwinding amounts to using a regular central difference operator δ_x on a PDE with a diffusion coefficient modified to be $\sigma(t, x) + \sqrt{|\mu(t, x)|\Delta_x}$. The numerical scheme in effect introduces enough artificial diffusion into the PDE to satisfy (2.34). Doing so, however, comes at a cost: the convergence order of the scheme will be reduced to $O(\Delta_x)$ if one-sided differencing ends up being activated in a significant part of the grid. We note that higher-order upwinding schemes are possible if the finite difference operator δ_x^* is allowed to act on more than three neighboring points. For such schemes, the matrix \mathbf{A} will no longer be tri-diagonal.

2.6.2 Other Techniques

As discussed earlier, upwinding amounts to adding numerical diffusion at nodes where the scheme is convection dominated. Alternatively, we can increase $\sigma(t, x)$ directly, to $\sigma(t, x) + \varepsilon$ where ε is chosen to be large enough for the scheme to satisfy (2.34). By solving the resulting PDE for different values of ε , it may be possible to determine how the error associated with ε scales in ε . This, in turn, will allow us to extrapolate to the limit $\varepsilon = 0$. See p. 135 of Tavella and Randall [2000] for an example.

The upwinding scheme presented in Section 2.6.1 switches abruptly from central differencing to one-sided differencing when the condition (2.34) is violated. In some schemes, the switch from central to one-sided differencing is made smooth by using a weighted average of a one-sided and a central difference operator. The weight on the central difference is close to one when $\sigma(t, x)^2 \gg |\mu(t, x)|\Delta_x$, but decreases smoothly to zero as $\sigma(t, x)^2/|\mu(t, x)|$ tends to zero. While it is unclear whether a smooth transition to upwinding is truly important (the convergence order is typically not improved over straight upwinding), Duffy [2000] suggests that the class of exponentially fitted schemes (see Duffy [2000] and Stoyan [1979]) may be quite robust in derivatives pricing applications.

In some finance applications, multi-dimensional PDEs might arise where $\sigma(t, x) = 0$ for one of the underlying variables; see for instance Section 2.7.5. While upwinding techniques still apply here, we note that specialized methods exist with better ($O(\Delta_x^2)$) convergence, should they become necessary. See, for instance, d'Halluin et al. [2005] for details on the so-called *semi-Lagrangian* methods.

2.7 Option Examples

In our discussion so far, we have assumed that options are characterized by a single terminal payoff function $g(x)$ and a set of spatial boundary

conditions determining the option price at the boundaries of the x -domain. In reality, many options are more complicated than this and may involve early exercise decisions, pre-maturity cash flows, path dependency, and more. In this section, we provide some relatively straightforward examples of such complications and how to modify the basic finite difference algorithm to deal with them. More examples will be provided later, in the context of specific fixed income securities.

2.7.1 Continuous Barrier Options

We have already touched upon the concept of an up-and-out knock-out option, an option that expires worthless if the x -process ever rises above a critical level H . As we described, we here must simply solve the PDE (2.1) on a domain $[\underline{M}, H]$, where \underline{M} represents the lowest attainable value of the process $x(t)$ on $[0, T]$. The boundary condition at the upper boundary is then dictated to be $V(t, H) = 0$, i.e. of the Dirichlet type. We can generalize this to allow both “up” and “down” type barriers, and to perhaps give a non-zero payout (a “rebate”) at the time the barrier(s) are hit (provided this happens before the option maturity). Specifically, if we have a lower barrier at \underline{H} , an upper barrier of \bar{H} , a time-dependent lower rebate function of $\underline{f}(t)$, and a time-dependent upper rebate function of $\bar{f}(t)$, we must dimension our spatial grid $\{x_j\}_{j=0}^{m+1}$ to have $x_0 = \underline{H}$, $x_{m+1} = \bar{H}$, and we then simply impose the Dirichlet boundary conditions $V(t, x_0) = \underline{f}(t)$ and $V(t, x_{m+1}) = \bar{f}(t)$. See (2.10) and the definition of Ω for the algorithm required to incorporate such Dirichlet boundary conditions into the finite difference scheme.

In practice, barrier options sometimes involve time-dependent barriers, possibly with discontinuities. For instance, *step-up* and *step-down* barrier options will have piecewise flat barriers that increase (step-up) or decrease (step-down) at discrete points in time. Extension of the finite difference algorithm to cover step-up and step-down options is relatively straightforward. As an illustration, consider a zero-rebate up-and-out single-barrier option where the (upper) barrier is flat, except for a discontinuous change at time $T^* < T$, at which point the barrier moves from a value of H^* to a value of H , with $H > H^*$. We set the x -domain of our finite difference grid to $x \in [\underline{M}, H]$, with \underline{M} a probabilistic lower limit, as defined above; accordingly, our spatial grid would be $\{x_j\}_{j=0}^{m+1}$, where $x_0 = \underline{M}$ and $x_{m+1} = H$. In preparation for the shift in barrier levels at time T^* , we make sure that one level in the spatial grid — say x_{k+1} , $k < m$, — is set exactly at the level H^* . Similarly, we make sure that one level in the time grid is set exactly to T^* . Starting at time T , we then iterate backwards in time by repeated solution of m -dimensional tri-diagonal systems of equations, at each step integrating a prescribed rebate function by supplying the Dirichlet boundary condition $V(t, x_{m+1}) = 0$. The moment we hit T^* , the PDE now only applies to the smaller region $[\underline{M}, H^*]$, covered by the reduced spatial grid $\{x_j\}_{j=0}^{k+1}$ with

$x_{k+1} = H^*$. From T^* back to time 0, the backward induction algorithm then involves only k -dimensional tri-diagonal systems of equations, with the Dirichlet boundary condition $V(t, x_{k+1}) = 0$. Spatial nodes above x_{k+1} correspond to zero option value and can be ignored¹². Modification of the algorithm outlined above to handle more than two barrier discontinuities is straightforward.

We can extend our definition of barrier options even further by making the topology of “alive” and “dead” regions more complicated. At time t , assume for instance that the PDE applies in an “alive” region of $x \in L(t)$ and a rebate function $R(t, x)$ that applies in the “dead” region $D(t) = \mathcal{B} \setminus L(t)$. Assume that we discretize the problem on a single rectangular finite difference grid spanning the spatial domain $[\underline{M}, \bar{M}]$, where \underline{M} and \bar{M} are set such that the alive regions are covered, up to probabilistic limits (if necessary). Given option values at time t_{i+1} , we then only need to run the basic matrix equation (2.18) for values in our grid $\{x_j\}$ that lie inside $L(t_i)$. This requires scaling down the dimension of the matrix \mathbf{A} as needed, and providing the relevant boundary conditions (given through $R(t_i, x_j)$) at the boundary (or boundaries) of $L(t_i)$. The parts of the spatial grid that lie outside of $L(t_i)$ can be directly filled in with values provided by the rebate function R . Notice that, if possible, the spatial grid should be set such that the boundaries of $L(t_i)$ are contained in the mesh; this will likely require us to use the techniques outlined in Section 2.4.

If the alive region has the simple form $L(t) = [\alpha(t), \beta(t)]$ for smooth deterministic functions α and β , an alternative to the scheme above is to introduce a time-dependent transformation that straightens out the barriers, allowing us to return to the standard finite difference setup where the PDE applies to a single rectangular (t, x) domain. One possible transformation involves using a spatial variable of

$$y = y(t, x) = \frac{x - \alpha(t)}{\beta(t) - \alpha(t)}, \quad (2.41)$$

which transforms the curved x -barriers $\alpha(t)$ and $\beta(t)$ into flat y -barriers at $y = 0$ and $y = 1$, respectively. The linearity of the transformation (2.41) makes it easy to work with; see Tavella and Randall [2000] for details and a discussion of extensions to multi-dimensional PDEs and to barriers with discontinuities.

¹²An obvious twist to the algorithm involves using different spatial grids over $[0, T^*]$ and $[T^*, T]$, allowing for more flexibility in node placement. In this case, values computed by backward induction must, at time T^* , be interpolated from one x -grid to another. The interpolation rule should be at least third-order accurate; see the discussion in Section 2.7.3.

2.7.2 Discrete Barrier Options

The barrier options considered in Section 2.7.1 are continuously monitored, in the sense that the barrier condition is observed for all times in a given interval. In practice, monitoring the barrier condition continuously can be impractical, and it may instead only be imposed on a discrete set of dates $T_1 < T_2 < \dots < T_K$, with $T_K \leq T$ and $T_1 > 0$. For the sake of concreteness, let us consider a discretely monitored up-and-out option with a constant barrier H . For a continuously monitored up-and-out barrier option it would suffice to solve the PDE on a domain $x \in [\underline{M}, H]$, where \underline{M} is a probabilistic lower limit. This is, however, no longer the case for a discretely monitored option where we need to allow the value function to “diffuse” above the barrier levels between dates in the monitoring set $\{T_k\}_{k=1}^K$. To allow for this, we discretize the PDE on a larger domain $x \in [\underline{M}, \bar{M}]$, $\bar{M} > H$. We can determine \bar{M} probabilistically by determining a confidence interval for how far above the barrier $x(t)$ can rise between monitoring dates. For instance, for the Black-Scholes PDE (2.3), assume that $\max_{k=2,\dots,K} (T_k - T_{k-1}) = \Delta_T$. Conditioned on $x(t) = H$, the probability that $x(t + \Delta_T)$ exceeds

$$x_\alpha = H + \left(r - \frac{1}{2}\sigma^2 \right) \Delta_T + \alpha\sigma\sqrt{\Delta_T}$$

is $\Phi(-\alpha)$. As in Section 2.1, we recommend setting $\bar{M} = x_\alpha$ for values of α somewhere between 3 and 5. To properly capture diffusion between barrier observation dates, we should also dimension the time grid of the finite difference scheme such that multiple time steps (at least two or three, say) are taken between observation dates. All observation dates $\{T_k\}_{k=1}^K$ should obviously be contained in the time grid.

Between barrier observation dates, we solve our PDE by the standard finite difference algorithm outlined in Section 2.2.4, as always imposing either an asymptotic Dirichlet condition at $x = \bar{M}$ or a condition on the x -derivatives of the value function. At each barrier observation time T_k , we must impose a *barrier jump condition*

$$V(T_k-, x) = V(T_k+, x)1_{\{x < H\}}, \quad k = 1, \dots, K, \quad (2.42)$$

where the notation $T_k \pm$ was introduced in Section 1.10.1 to denote the limit $T_k \pm \varepsilon$ for $\varepsilon \downarrow 0$. This merely states that all values $V(T_k, x)$ are zero for $x \geq H$, consistent with the definition of an up-and-out option. In our finite difference scheme, we incorporate this jump condition by simply interpreting the vector $\widehat{\mathbf{V}}(T_k)$ as found by regular backward induction as $\widehat{\mathbf{V}}(T_k+)$ and then replacing

$$\widehat{\mathbf{V}}(T_k+) = \left(\widehat{V}_1(T_k+), \dots, \widehat{V}_m(T_k+) \right)^\top$$

with

$$\widehat{\mathbf{V}}(T_k-) = \left(\widehat{V}_1(T_k+) 1_{\{x_1 < H\}}, \dots, \widehat{V}_m(T_k+) 1_{\{x_m < H\}} \right)^\top$$

before continuing the algorithm backwards from T_k .

The jump condition (2.42) will generally produce a discontinuity in V as a function of x , around the barrier level H . If we use Crank-Nicolson time-stepping, it will then be prudent to employ a fully implicit scheme for the first few backwards time steps (Rannacher stepping) past each barrier observation date T_k . As discussed in Section 2.5, ideally this should be combined with a smoothing algorithm acting on $\widehat{\mathbf{V}}(T_k-)$ or, perhaps more conveniently, a shift of the spatial grid such that H lies exactly mid-way between two spatial nodes in the grid.

We round off by noting that the discussion above for an up-and-out option easily extends to more complicated discrete barrier options, including those with time-varying barrier levels and rebates. For instance, assume that an option involves upper and lower time-varying barriers of $\overline{H}(t)$ and $\underline{H}(t)$, respectively, as well as a time- and state-dependent rebate of $R(t, x)$. In this case, we simply replace the jump condition (2.42) with

$$\begin{aligned} V(T_k-, x) &= V(T_k+, x) 1_{\{\underline{H}(T_k) < x < \overline{H}(T_k)\}} \\ &\quad + R(T_k, x) \left(1_{\{x \geq \overline{H}(T_k)\}} + 1_{\{x \leq \underline{H}(T_k)\}} \right), \end{aligned}$$

and otherwise proceed as above. We note that time-dependent barriers will typically require flexibility in setting the spatial grid, as there are now multiple critical x -levels to consider. The discretization in Section 2.4 can obviously assist with this.

2.7.3 Coupon-Paying Securities and Dividends

Many fixed-income securities are coupon-bearing and involve periodic transfer of a cash amount between the buyer and the seller. This can easily be incorporated into a finite difference grid, through a jump condition. Specifically, consider a security that pays its owner a single cash amount of $p(T^*, x)$ at time $T^* < T$, where p is a deterministic function $p : [0, T] \times \mathcal{B} \rightarrow \mathbb{R}$. We dimension our time grid such that T^* is contained in the grid, and then apply at time T^* the condition

$$V(T^*, x) = V(T^*+, x) + p(T^*, x). \quad (2.43)$$

This simply expresses that V will decrease by an amount p immediately after p is paid (and thereby no longer contained in V). In a finite difference algorithm, (2.43) is incorporated by replacing $\widehat{\mathbf{V}}(T^*+)$, as found by regular backward induction, with

$$\widehat{\mathbf{V}}(T^*-) = \left(\widehat{V}_1(T^*+) + p(T^*, x_1), \dots, \widehat{V}_m(T^*+) + p(T^*, x_m) \right)^\top$$

before continuing the algorithm backwards from T^* . Extensions to multiple coupons are trivial.

In some cases a derivative security does not itself pay coupons, but is written on a security that does. This involves no particular complications, except for the case where payments may affect the state variable underlying the PDE. For instance, consider the classical case of a stock paying a dividend: at the time of the dividend payment, the stock jumps down by an amount equal to the dividend payment. For a model that uses the stock price (or a transformation of the stock price) as the state variable x , a dividend payment at time T^* would thus be associated with a discontinuity in the state variable, $x(T^*+) = x(T^*-) - d(T^*, x(T^* -))$, where d is the magnitude of the jump¹³. As long as the dividend-payment does not come as a surprise (i.e., at a random time), it must already be contained into the option price at $T^* -$, and will have no price effect as we move forward from $T^* -$ to $T^* +$. We can express this continuity restriction through yet another jump condition

$$V(T^* -, x) = V(T^* +, x - d(T^*, x)). \quad (2.44)$$

See Wilmott et al. [1993] for more discussion. Implementation of (2.44) in a finite difference grid proceeds as follows. First, we use regular backward induction to establish

$$\begin{aligned} \widehat{\mathbf{V}}(T^* +) &= \left(\widehat{V}_1(T^* +), \dots, \widehat{V}_m(T^* +) \right)^\top \\ &= \left(\widehat{V}(T^* +, x_1), \dots, \widehat{V}(T^* +, x_m) \right)^\top. \end{aligned}$$

Then we write

$$\widehat{\mathbf{V}}(T^* -) = \left(\widehat{V}(T^* +, x_1 - d(T^*, x_1)), \dots, \widehat{V}(T^* +, x_m - d(T^*, x_m)) \right)^\top.$$

The values $\widehat{V}_j(T^* +, x_j - d)$ here can be found by interpolation in the x -direction on the $\widehat{\mathbf{V}}(T^* +)$ -array. As shown in Tavella and Randall [2000], the order of the interpolator should be strictly higher than two, to avoid inducing spurious numerical diffusion into our θ -style finite difference schemes. We note that this rules out the piecewise linear interpolation rule proposed in Wilmott et al. [1993]. A common choice is to use cubic spline interpolation; see Chapter 6 for much information on cubic splines.

2.7.4 Securities with Early Exercise

In Section 1.10 we introduced the concept of Bermudan and American securities with early exercise features. Under the assumption that exercise

¹³To prevent negative stock prices, it may be necessary to truncate the size of d locally in the finite difference grid. For simplicity, we ignore this complication here.

values are determined by a deterministic function¹⁴ $h(t, x)$, $h : [0, T] \times \mathcal{B} \rightarrow \mathbb{R}$, finite difference grids are ideal for pricing of such securities. Let us first consider a Bermudan option with exercise opportunities restricted to the finite set $\{T_k\}_{k=1}^K$. The Bellman principle (1.67) in Section 1.10 can, as shown there, be expressed as a simple jump condition

$$V(T_k-, x) = \max(V(T_k+, x), h(T_k, x)), \quad k = 1, \dots, K, \quad (2.45)$$

which can be incorporated into a finite difference solver precisely the same way as in previous sections. The condition (2.45) will result in a kink in the value function around the level of x at which we shift from the hold region into the exercise region. If Crank-Nicolson time-stepping is used, one should ideally apply smoothing on the finite difference value vector $\hat{\mathbf{V}}(T^* -)$, particularly around the kink.

If exercise can take place continuously (that is, American-style) on a given time interval, a crude way to incorporate this into a finite difference grid is by simply applying (2.45) to every point in the time grid of the finite difference scheme. By not specifically imposing the partial differential inequalities (see Section 1.10.1), this algorithm, however, will generally only be accurate to first order in the time step, even if a Crank-Nicolson scheme is used; see Carverhill and Clewlow [1990] for a proof. As American-style exercise is rarely used in fixed income markets, we shall not pursue this issue further but just point out that a number of schemes exist to restore second-order time convergence to finite difference pricing of American options, see, e.g., Forsyth and Vetzal [2002].

2.7.5 Path-Dependent Options

Finite difference methods are normally limited to Markovian problems where dynamics are characterized by SDEs and where payouts are simple deterministic functions of the underlying state variables. A number of options, however, have terminal time T payouts that depend not only on the state of x at time T , but on the entire path $\{x(t), t \in [0, T]\}$. In general, such options must be priced by Monte Carlo methods (see Chapter 3), but exceptions exist. Indeed, barrier and American options can be considered path-dependent options, yet, as we have seen, can still be priced in a finite difference grid. Even stronger path-dependence can sometimes be handled, through the introduction of new state variables to the PDE.

To give an example, consider a path-dependent contract where the terminal payout at time T can be written as

¹⁴If h represents the value of a derivative security that has no closed-form pricing formula, it may be necessary to estimate this function by backward induction in the finite difference grid itself. Such a “preprocessing” step is typically straightforward to execute.

$$V(T) = g(x(T), I(T)), \quad (2.46)$$

where I is a path integral of the type

$$I(t) = \int_0^t h(x(s)) ds, \quad (2.47)$$

for some deterministic function h . For instance, if $h(x) = x$, we say that the option is a continuously sampled *Asian option*.

For the payout (2.46) we have $V(t) = V(t, x(t), I(t))$ where $x(t)$ satisfies the SDE (2.2) and

$$dI(t) = h(x(t)) dt, \quad I(0) = 0.$$

From the backward Kolmogorov equation, it follows that $V(t, x, I)$ solves

$$\frac{\partial V}{\partial t} + \mu(t, x) \frac{\partial V}{\partial x} + \frac{1}{2} \sigma(t, x)^2 \frac{\partial^2 V}{\partial x^2} + h(x) \frac{\partial V}{\partial I} = r(t, x)V, \quad (2.48)$$

subject to the terminal condition $V(T, x, I) = g(x, I)$. There are several complications with this PDE. First, it involves *two* spatial variables, x and I , requiring the use of a two-dimensional PDE solver. Second, the PDE contains no second-order derivative in the variable I , i.e. it is convection dominated in the I -direction. We have discussed methods to handle the latter issue in Section 2.6.1 and will turn to address the former in Section 2.9. Another complication is the fact that the term $h(x)$ multiplying $\partial V / \partial I$ may be of a different order of magnitude than the other coefficients in (2.48), increasing the difficulty of solving the equation numerically. We refer to Zvan et al. [1998] for a more detailed discussion of PDEs of the type (2.48).

In practice, it is rare that a continuous-time integral such as (2.47) is used in an option payout. Instead, one normally samples the function $h(x(t))$ only on a discrete set of dates, i.e. we replace $I(T)$ with

$$I(T) = \sum_{i=1}^n h(x(T_i)) (T_i - T_{i-1}),$$

where $T_0 < T_1 < \dots < T_n$ is a discrete schedule, with $T_0 = 0$ and $T_n = T$. Informally, we now have

$$dI(t) = \delta(T_i - t) \cdot h(x(T_i)) (T_i - T_{i-1}), \quad I(0) = 0, \quad (2.49)$$

where $\delta(\cdot)$ is the Dirac delta function. In a PDE setting, we incorporate a process such as (2.49) through appropriate jump conditions, writing

$$V(T_i-, x, I) = V(T_i+, x, I + h(x)(T_i - T_{i-1})). \quad (2.50)$$

In the same fashion as for discrete dividends (Section 2.7.3), the jump condition enforces continuity of the option price across the dates where I

gets updated. The condition is applied at each date in the discrete schedule, $i = 1, \dots, n$; in between schedule dates (where now $dI(t) = 0$), we solve the PDE

$$\frac{\partial V}{\partial t} + \mu(t, x) \frac{\partial V}{\partial x} + \frac{1}{2} \sigma(t, x)^2 \frac{\partial^2 V}{\partial x^2} = r(t, x)V,$$

which has no term involving I . When the I -direction is discretized in, say, m_I different values, the solution scheme thus involves solving m_I different *one-dimensional* PDEs backward in time; the solutions of these m_I PDEs exchange information with each other at each date in the schedule, in accordance with (2.50). As was the case for cash dividends, implementation of (2.50) will normally require support from an interpolation scheme, to align the (x -dependent) jumps in I with the knots of the discretized I -grid used in the finite difference scheme. See, e.g., Zvan et al. [1999] or Wilmott et al. [1993] for further details. An application of this idea in the context of interest rate derivatives is given in Section 18.4.5.

On rare occasions — basically when the homogeneity condition $V(\eta x, \lambda I, t) = \lambda^\eta V(x, I, t)$, $\lambda, \eta > 0$, holds — it is possible to make a change of variables or a change of probability measure that will reduce (2.48) or its discrete-time version to a one-dimensional PDE; see e.g. Rogers and Shi [1995] or Andreasen [1998] for the case of various Asian options. Section 18.4.5 demonstrates one such method, sometimes called the method of *similarity reduction*, for pricing of “weakly path-dependent” securities, including certain callable interest rate derivatives where the notional accretes at a stochastic coupon rate (see Section 5.14.5 for definitions).

2.7.6 Multiple Exercise Rights

Certain financial products with early exercise rights allow the holder to exercise more than once. Such “multi-exercise” options are relatively rare, but the so-called *chooser cap* (also known as a *flexi-cap*) is occasionally traded and constitutes a good example for describing how to handle multi-exercise options in a PDE setting. Let there be given a set of L possible exercise dates, $T_1 < T_2 < \dots < T_L$, and assume that we have the right to exercise no more than l times, with $l < L$. Provided that we exercise at time T_i , in a chooser cap we are paid¹⁵ $(S(T_i) - K)^+$, where $S(\cdot)$ is some interest rate index and K is the strike. Clearly, we would never exercise at time T_i unless $S(T_i) > K$, but how much larger than K the rate $S(T_i)$ needs to be to trigger optimal exercise is not obvious, and must at least depend on i) how many of our l exercise opportunities we have already used up at time T_i ; and ii) how much value is lost by using (rather than postponing) one of the remaining exercise opportunities.

¹⁵We have ignored a day count scaling constant in the payout. Also, in most cases payment takes place at time T_{i+1} , rather than at T_i ; such a payment delay can be handled by a discount operation.

While the question of how to exercise optimally on a chooser cap may appear quite complex, it is surprisingly easy to implement in a finite difference setting by combining techniques from Sections 2.7.4 and 2.7.5 above. The key to the method is to introduce an additional state variable I to keep track of how many exercise opportunities are left. Assume that all interest rates are functions of a Markov state variable $x(\cdot)$, and let therefore $V(t, x, I)$ denote the value of the chooser cap at time t , given $x(t) = x$ and given that there are still I exercise opportunities left. Notice that the variable I can only take $l + 1$ distinct values: $0, 1, \dots, l$; notice also that $V(t, x, 0) = 0$ for all t and x , since $I = 0$ corresponds to the situation where there are no exercise opportunities left. Additionally, at the terminal time T_L we clearly have

$$V(T_L, x, I) = (S(T_L, x) - K)^+, \quad I = 1, 2, \dots, l, \quad (2.51)$$

where we have written $S(T_L) = S(T_L, x)$ to emphasize the deterministic dependence of S on the state variable x .

For given dynamics of $x(t)$, starting with the terminal conditions in (2.51), we may roll the l different value functions $V(\cdot, x, I)$, $I = 1, 2, \dots, l$, back through time in standard finite difference manner. At each time T_i , $i = 1, \dots, L - 1$, jump conditions similar to (2.45) must be applied, for all $I = 1, 2, \dots, l$:

$$V(T_{i-}, x, I) = \max(V(T_i+, x, I), V(T_i+, x, I - 1) + (S(T_i, x) - K)^+).$$

Notice that these conditions simply express that exercise is optimal only if the exercise value (the cap payout plus the value of a chooser cap with one less exercise opportunity) exceeds the hold value (the non-exercised chooser cap). Once we have rolled all the way back to $t = 0$, the chooser cap value at time $t = 0$ may be identified as $V(0) = V(0, x(0), l)$.

We should note that the “chooser” or “flexi” feature can be added to securities other than caps (and floors). For instance, in Section 19.5 we study the so-called *flexi-swap*, another security with multiple embedded exercise rights.

2.8 Special Issues

In this section, we briefly show a few techniques that may come in handy for certain applications.

2.8.1 Mesh Refinements for Multiple Events

As discussed in Section 2.1, the domain of the state variable x is often determined as an exact or approximate confidence interval for the random variable $x(T)$, where T is the final time of interest for a particular valuation problem we want to solve. Given the number of desired spatial steps in the

scheme, the discretization step in x -direction is then obtained by dividing the size of the confidence interval by the number of steps. Similarly, the discretization step in t -direction is typically obtained by dividing T by the number of desired time steps. This is a standard procedure for building a simple rectangular mesh, and it works well if the derivative we wish to value does not have any “interesting” features between the valuation time 0 and the final time T (e.g., for a simple European option). However, as should be evident from the examples in Section 2.7, many real-life derivative securities are characterized by a multitude of events during their lifetimes, all of which must be adequately captured in the PDE scheme. It is not hard to see that a grid dimensioning scheme based solely on the last event date may yield inappropriate mesh resolution at earlier dates.

To make the discussion above concrete, let us consider the example of a Bermudan option (see Section 2.7.4) with two exercise dates, T_1 and T_2 . Assume that $0 < T_1 \ll T_2$, i.e. that the first exercise date is much closer to the valuation date than the second (and last) one. Also assume that there is a decent chance that the option actually will be exercised at time T_1 , making it important to capture to good precision the value of the option expiring at T_1 . Now, if we build our mesh based only on the distribution of the state variable $x(T_2)$ at time T_2 , there would typically be too few t -points in the interval $[0, T_1]$. Also, the x -direction discretization step would be too large compared to the range of possible values of the state variable $x(T_1)$ at time T_1 , i.e. the x -grid would be too coarse for the process $x(\cdot)$ on the time interval $[0, T_1]$. Both issues would typically lead to a large discretization error in the finite difference stepping of the option over the time period $[0, T_1]$, leading to problems with accuracy in values and risk sensitivities.

The issue of the sparsity of the time grid is fairly easy to deal with, as we are free to add extra points to the time grid before time T_1 . This by itself, however, will not solve precision problems, as the space step remains large. Any proper solution should, of course, come in the form of refining both the t - and x -grids at the same time.

One possible way of refining the x -discretization is to abandon the usage of a single rectangular (t, x) -domain, and instead link together different equidistant rectangular meshes for different periods in the life of the derivative. These mesh “blocks” would generally increase in spatial width with time and would connect to each other via an interpolation scheme. To be more specific, let us assume, as in Section 2.1, that the state variable $x(\cdot)$ is the logarithm of the stock in the Black-Scholes model and is given by (2.4), with the PDE to solve given by (2.3). We extend our simple two-period example above to a derivative with K times of interest, $0 < T_1 < \dots < T_K$; these times could be specified as an additional input into valuation, or derived from the trade description (e.g. they could represent the exercise dates for a Bermudan option, or the knock-out dates for the discretely-monitored barrier option of Section 2.7.2). Suppose we are given values of m and n , and now wish to construct the mesh for the time period $[T_{k-1}, T_k]$, by using

the same time and space steps Δ_t^k , Δ_x^k as would be used in the standard scheme of Section 2.1 for a derivative security with the terminal payoff at T_k . That is, having fixed the cutoff α we would set

$$\Delta_t^k = T_k/n, \quad \Delta_x^k = 2\alpha\sigma\sqrt{T_k}/(m+1). \quad (2.52)$$

Then the rectangular, equidistant mesh for the time period $[T_{k-1}, T_k]$ is given by

$$\{t_i^k\}_{i=0}^{\lfloor(1-T_{k-1}/T_k)n\rfloor} \times \{x_j^k\}_{j=0}^{m+1}, \quad t_i^k = T_{k-1} + i\Delta_t^k, \quad x_j^k = x_{\min}^k + j\Delta_x^k, \quad (2.53)$$

where $\lfloor \cdot \rfloor$ denotes the integer part of a real number and (see (2.4))

$$x_{\min}^k = x(0) + \left(r - \frac{1}{2}\sigma^2\right)T_k - \alpha\sigma\sqrt{T_k}. \quad (2.54)$$

Note that in reality we would want to make sure that the point T_k is also in the mesh for the time period $[T_{k-1}, T_k]$, even though for simplicity of notations we did not reflect it in (2.53). It is also useful to note that the total number of time points is not going to be n , but is actually equal to

$$\sum_{k=1}^K \lfloor(1 - T_{k-1}/T_k)n\rfloor,$$

which scales linearly with n . Clearly, if exactly n points were required, a simple adjustment to the definition of the time step in (2.52) could be applied.

With a mesh as defined above, when arriving at time T_k in a backward induction scheme the solution $V(T_k, \cdot)$ would be discretized on the x -grid $\{x_j^{k+1}\}_{j=0}^{m+1}$. To solve the PDE backwards over the time period $[T_{k-1}, T_k]$, we would need to resample it on the different x -grid $\{x_j^k\}_{j=0}^{m+1}$. As with interpolation across dividends (Section 2.7.3), simple cubic interpolation would be a good choice here. Specifically, one would fit a cubic spline to the values $V(T_k, x_j^{k+1})$, $j = 0, \dots, m+1$, and then calculate $V(T_k, x_j^k)$, $j = 0, \dots, m+1$, by valuing the spline at the required grid points.

The “interpolated mesh” scheme above is rather intuitive and straightforward, but it does suffer from the need to do interpolation work that could slow down the PDE (especially in dimensions higher than 1 and/or for a large number of interface points K). Also, it is not entirely clear how interpolation will affect stability and convergence properties of the PDE. Finally, linking the interface mesh geometry to the trade specifics (such as exercise dates) may not be ideal from the point of view of designing an efficient valuation flow in a risk management system. These considerations lead us to an alternative approach that relies on *non-equidistant* discretization as developed in Section 2.4. The idea of this method is to use non-uniform

discretization to concentrate more points, both in time and space, around the initial point $t = 0$, $x = x(0)$. Clearly many ways of achieving this are possible — below we present a simple scheme we have used with good results.

We define K , the user input, to be the number of spatial refinement levels (with $K = 2$ or 3 typically used), and τ , another user input, to be a time scaling constant (typically $\tau = 4$). If T is the final horizon for valuation, we then introduce times

$$0 = T_0 < T_1 < \dots < T_K = T$$

by

$$T_k = \frac{T}{\tau^{K-k}}, \quad k = 1, \dots, K.$$

Then, the time grid for the time period $[T_{k-1}, T_k]$ is given by uniformly distributing $\tilde{n} \triangleq \lfloor n/K \rfloor$ points¹⁶ over $[T_{k-1}, T_k]$, i.e. is given by $\{t_i^k\}_{i=0}^{\tilde{n}}$ with

$$t_i^k = T_{k-1} + i \frac{T_k - T_{k-1}}{\tilde{n}} = T_{k-1} + i \frac{T_k - T_{k-1}}{\lfloor n/K \rfloor}.$$

(Note that we can use this specification with the interpolated mesh as well, instead of the time grid definition in (2.53)). The fact that the width of the intervals $[T_{k-1}, T_k]$ grow with k means that the time grid is more finely spaced in the beginning of the interval $[0, T]$ than at the end.

The x -grid we are going to define will be universal — i.e. the same for all time steps on the whole time interval $[0, T]$ — and non-uniform. To construct it, we first define a set of nested x -subdomains $[x_{\min}^k, x_{\max}^k]$, with x_{\min}^k defined by (2.54) and x_{\max}^k defined accordingly, i.e.

$$x_{\max}^k = x(0) + \left(r - \frac{1}{2}\sigma^2 \right) T_k + \alpha\sigma\sqrt{T_k},$$

for $k = 0, \dots, K$. Then we define step sizes by

$$\Delta_x^k = \frac{x_{\max}^k - x_{\min}^k}{\tilde{m} + 1}, \quad \tilde{m} = \left\lfloor \frac{m}{K} \right\rfloor.$$

The x -grid is then constructed by distributing grid points uniformly in subintervals $[x_{\min}^k, x_{\min}^{k-1}]$ and $[x_{\max}^{k-1}, x_{\max}^k]$ with the space step Δ_x^k , and is given by

$$\left(\bigcup_{k=1}^K \left\{ x_j^{\min,k} \right\}_{j=0}^{m^k+1} \right) \cup \left(\bigcup_{k=1}^K \left\{ x_j^{\max,k} \right\}_{j=0}^{m^k+1} \right),$$

where

$$x_j^{\min,k} = x_{\min}^k + j\Delta_x^k, \quad x_j^{\max,k} = x_{\max}^{k-1} + j\Delta_x^k,$$

¹⁶And, as advised earlier, adding trade event dates that fall into this period — although we do not reflect this in our notations for simplicity.

and

$$m^k = \left\lfloor \frac{x_{\min}^{k-1} - x_{\min}^k}{\Delta_x^k} \right\rfloor - 1 = \left\lfloor \frac{x_{\max}^k - x_{\max}^{k-1}}{\Delta_x^k} \right\rfloor - 1.$$

This distribution of space points results in an x -grid that is more dense around the point $x = x(0)$ than at the edges. It is worth noting that with only one refinement level $K = 1$, the standard rectangular uniform mesh sized by the terminal distribution of the state variable is recovered.

2.8.2 Analytics at the Last Time Step

In cases where the dynamics of underlying PDE variables are tractable, one naturally wonders whether finite difference methods could somehow be improved by incorporating analytical results into the scheme. Here, and in the next section, we discuss two simple ideas.

Suppose that we are faced with the problem of pricing a contingent claim with terminal boundary condition $g(x(T))$, where $x(t)$ is a Markovian process with known Arrow-Debreu state prices:

$$G(t, x; s, y) = \mathbb{E}^Q \left(\delta(x(s) - y) e^{-\int_t^s r(u, x(u)) du} | x(t) = x \right), \quad s > t.$$

Assume also that the claim in question involves a jump condition at time $0 < T^* < T$ (but no jump conditions between T^* and T). If our finite difference grid is $\{x_j\}_{j=0}^{m+1}$, we can now use a series of $m + 2$ outright convolutions to compute

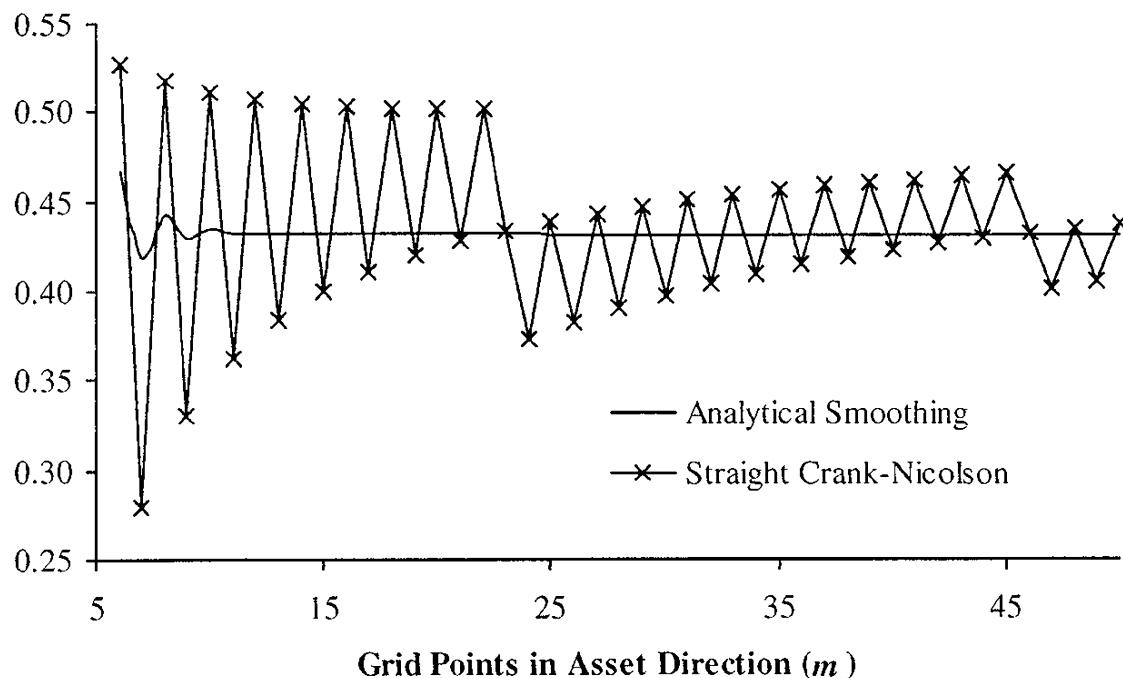
$$V(T^*, x_j) = \int_{\mathbb{R}} G(T^*, x_j; T, y) g(y) dy, \quad j = 0, 1, \dots, m + 1. \quad (2.55)$$

If we are lucky (i.e., if both g and G are sufficiently simple), then the integral on the right-hand side may be known in closed form for all values of x_j . If not, we can always perform a series of numerical integrations, the total cost of which is typically¹⁷ $O(m^2)$, i.e. more expensive than the typical $O(m)$ cost of a single time step in a finite difference method. There are several reasons why we may want to perform the numerical integrations nevertheless. First, the convolution expression (2.55) is exact, as it is based on the true transition density. Second, if the gap between T^* and T is large, an ordinary finite difference grid would need to roll back from T to T^* using multiple time steps n^* , at a total cost of $O(n^* m)$; if n^* is of the same magnitude as m , the computational effort of the convolution scheme would be comparable to that of a finite difference grid. Third, for discontinuous payouts, the integration in (2.55) will have a naturally smoothing effect, similar to (but often better than) the continuity correction method of Section 2.5.2. The smoothing

¹⁷There are exceptions. For instance, if fast Fourier transform (FFT) methods are applicable, the cost may be reduced to $O(m \ln(m))$. See Section 8.4 for details.

effect is discussed in more detail in Section 23.2.4 and is also demonstrated below, in Figure 2.2, where we have continued our investigation of the 3 year digital option considered earlier in Section 2.5.3. Since the model used in Figure 2.2 is ordinary Black-Scholes and $g(x) = 1_{\{x>H\}}$, the integrals in (2.55) can here be computed in closed form from (2.33).

Fig. 2.2. 3 Year Digital Option Price



Notes: Finite difference estimates for the Black-Scholes price of a 3 year digital option. All contract and model parameters are as in Figure 2.1. Time stepping is performed with an equidistant grid containing $n = 50$ points. Spatial discretization in log-space is equidistant, as described in the main text; the number of grid points (m) is as listed on the x -axis of the figure. The “Straight Crank-Nicolson” graph shows the convergence profile for a pure Crank-Nicolson finite difference grid. The “Analytical Smoothing” graph shows the convergence profile for a Crank-Nicolson finite difference grid starting at $T^* = 2.5$ years, with the terminal boundary condition set equal to a 0.5 year digital option price (as in (2.55)). The theoretical value of the option is 0.4312451.

In principle, we could continue rolling back from T^* (through, possibly, jump conditions at earlier times) by performing convolutions, rather than solving finite difference grids. In practice, this rarely leads to improvements over a finite difference grid, unless the densities and payoffs are quite simple¹⁸. Moreover, in many cases we may not have *exact* Arrow-Debreu

¹⁸For simple densities (especially Gaussian), special-purpose methods exist to compute convolutions rapidly, typically involving payoff approximations through piecewise polynomials or other simple functions. We do not cover these methods in

prices, only approximate ones based on, say, a small-time expansion (see, e.g., Section 13.1.9.1). In this case, a one-time convolution may be safe — especially if $T - T^*$ is small — whereas repeated convolutions may lead to unacceptable biases.

2.8.3 Analytics at the First Time Step

The idea in Section 2.8.2 of replacing the finite difference stepping with analytical integration is even easier to apply over the *first*, rather than the *last*, time step. Suppose T^* is the first “interesting” time for a given derivative security, i.e. there might be a jump condition at time T^* but none over the time interval $[0, T^*]$. Then, rather than stepping the finite difference scheme from T^* to 0, we can perform a *single* integration to calculate the value $V(0, x(0))$ of the derivative at time zero from the discretized values $\{V(T^*, x_j)\}_{j=0}^{m+1}$ of the derivative at time T^* (using the same notations as in Section 2.8.2),

$$V(0, x(0)) = \int_{\mathbb{R}} G(0, x(0); T^*, y) \tilde{V}(T^*, y) dy,$$

where $\tilde{V}(T^*, y)$ is interpolated (using cubic splines, say) from the values $\{V(T^*, x_j)\}_{j=0}^{m+1}$ on the grid. If the integral is computed numerically — as is most often the case — the numerical cost is often comparable with that of the finite difference stepping because only one value $V(0, x(0))$ is required at time 0, not the whole slice.

While there are typically no numerical cost savings that arise from using integration over the first time step, there are accuracy and stability considerations that favor this approach. We have already seen in Section 2.8.1 that the standard discretization of a PDE often leads to insufficient fidelity in resolving any features of the payoff that are close to today, and numerical integration can be of considerable help in this regard. Moreover, as we discuss in much detail later in Chapter 23, an integration scheme typically allows us to treat discontinuities in the value $V(T^*, x)$ arising from the jump condition at time T^* explicitly. If the discontinuity is introduced at the value of the state variable x^* , then the integration scheme can (and should) explicitly take this information into account. For example we would write

$$\begin{aligned} V(0, x(0)) &= \int_{-\infty}^{x^*} G(0, x(0); T^*, y) \tilde{V}^-(T^*, y) dy \\ &\quad + \int_{x^*}^{\infty} G(0, x(0); T^*, y) \tilde{V}^+(T^*, y) dy \end{aligned}$$

this book except for a brief mention in Section 11.A. For a representative example see Hu et al. [2006].

and calculate $\tilde{V}^-(T^*, y)$ by interpolating the grid values in the time interval $(-\infty, x^*)$, and $\tilde{V}^+(T^*, y)$ by interpolating the grid values in (x^*, ∞) , separately¹⁹.

The usefulness of the method is only limited by the availability of the closed-form expression for the time 0 Arrow-Debreu prices $G(0, x(0); T^*, \cdot)$. For some models this is not an issue; for most others, sufficiently close approximations could be obtained in a small-time limit (see e.g. Section 13.1.9.1 for a typical approach) that can be useful for times T^* that are not too large. By a change of measure, we see that

$$\begin{aligned} V(0, x(0)) &= \mathbb{E} \left(e^{-\int_0^{T^*} r(s) ds} V(T^*, x(T^*)) \right) \\ &= P(0, T^*) \mathbb{E}^{T^*} (V(T^*, x(T^*))), \end{aligned}$$

where \mathbb{E}^{T^*} is the expected value operator under the T^* -forward measure Q^{T^*} ; so we really only need the expression for the *density* (rather than Arrow-Debreu security prices) of $x(T^*)$ under Q^{T^*} , either exact or approximate.

Finally, we note that while the integration over the first time step can be seen to offer similar advantages to those of the methods in Section 2.8.1, the two approaches are not substitutes for each other, but are complementary. We typically recommend using direct integration over the time step $[0, T^*]$, where T^* is the smaller of the time of the first jump condition or the limit of applicability of the approximation to the density of $x(T^*)$, and then (if needed) use the methods in Section 2.8.1 over the time interval $[T^*, T]$, with T being the final maturity of the option in question.

2.9 Multi-Dimensional PDEs: Problem Formulation

We now turn our attention to the numerical solution of multi-dimensional terminal value problems. Let the spatial variable x be p -dimensional, $x = (x_1, \dots, x_p)^\top$, and consider the PDE

$$\frac{\partial V}{\partial t} + \sum_{h=1}^p \mu_h(t, x) \frac{\partial V}{\partial x_h} + \frac{1}{2} \sum_{h=1}^p \sum_{l=1}^p s_{h,l}(t, x) \frac{\partial^2 V}{\partial x_h \partial x_l} - r(t, x) = 0, \quad (2.56)$$

where $s_{h,h}(t, x) \geq 0$ and $s_{h,l}(t, x) = s_{l,h}(t, x)$ for $h, l = 1, \dots, p$. The PDE is assumed subject to the terminal value condition $V(T, x) = g(x)$, $g : \mathbb{R}^p \rightarrow \mathbb{R}$.

From the results in Chapter 1, we recognize that the PDE provides the solution to the expectation

¹⁹One of the functions $\tilde{V}^-(T^*, y), \tilde{V}^+(T^*, y)$ is often known analytically and for all values of y (rather than sampled on the grid); this is for instance the case for the Bermudan options of Section 2.7.4. The integration algorithm should obviously take advantage of this.

$$V(t, x) = \mathbb{E}_t \left(e^{-\int_0^T r(u, x) du} g(x(T)) \mid x(t) = x \right),$$

where the components of $x(t)$ satisfy risk-neutral SDEs of the type

$$dx_h(t) = \mu_h(t, x(t)) dt + \sigma_h(t, x(t)) dW(t), \quad h = 1, \dots, p. \quad (2.57)$$

Here $W(t)$ is a d -dimensional Brownian motion, $\mu_h : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}$, $h = 1, \dots, p$, are (scalar) drifts, and $\sigma_h : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}^{1 \times d}$, $h = 1, \dots, p$, are d -dimensional (row vector) diffusion coefficients. The PDE coefficients $s_{h,l}$ in (2.56) represent the instantaneous covariance matrix for the components of $x(\cdot)$, i.e., $s_{h,l}(t, x) = \sigma_h(t, x)\sigma_l(t, x)^\top$. We assume enough regularity on μ_h , σ_h , r , and g to ensure that (2.56) has a unique solution.

For the purpose of solving (2.56) numerically, we assume that the PDE is to be solved on a (finite) spatial domain in x , $x \in [\underline{M}_1, \bar{M}_1] \times \dots \times [\underline{M}_p, \bar{M}_p]$, where $\underline{M}_h, \bar{M}_h$, $h = 1, \dots, p$, are constants either dictated by the contract at hand (barrier options) or found by a suitable probabilistic truncation (see Section 2.1).

2.10 Two-Dimensional PDE with No Mixed Derivatives

To illustrate the construction of finite difference discretization of (2.56), we start out with the simple case where $p = d = 2$ and there are no mixed partial derivatives in the PDE: $s_{1,2}(t, x) = s_{2,1}(t, x) = 0$ for all t and x . Probabilistically, the absence of mixed derivatives corresponds to the case where the stochastic process increments $dx_1(t)$ and $dx_2(t)$ are independent. Defining $\gamma_h(t, x)^2 = s_{h,h}(t, x)$, $h = 1, 2$, the PDE to be solved now becomes

$$\frac{\partial V}{\partial t} + (\mathcal{L}_1 + \mathcal{L}_2) V = 0, \quad (2.58)$$

where

$$\mathcal{L}_h \triangleq \mu_h(t, x) \frac{\partial}{\partial x_h} + \frac{1}{2} \gamma_h(t, x)^2 \frac{\partial^2}{\partial x_h^2} - \frac{1}{2} r(t, x), \quad h = 1, 2.$$

Notice that we have divided the term $r(t, x)$ into equal pieces in \mathcal{L}_1 and \mathcal{L}_2 .

To discretize (2.58) in x , introduce grids $x_1 \in \{x_1^{j_1}\}_{j_1=0}^{m_1+1}$ and $x_2 \in \{x_2^{j_2}\}_{j_2=0}^{m_2+1}$. To simplify notation, assume these grids are equidistant such that $x_1^{j_1} = \underline{M}_1 + j_1 \Delta_1$ and $x_2^{j_2} = \underline{M}_2 + j_2 \Delta_2$. Let $V_{j_1, j_2}(t) \triangleq V(t, x_1^{j_1}, x_2^{j_2})$. We define discrete central difference operators as before

$$\begin{aligned} \delta_{x_1} V_{j_1, j_2}(t) &= \frac{V_{j_1+1, j_2}(t) - V_{j_1-1, j_2}(t)}{2\Delta_1}, \\ \delta_{x_2} V_{j_1, j_2}(t) &= \frac{V_{j_1, j_2+1}(t) - V_{j_1, j_2-1}(t)}{2\Delta_2}, \end{aligned}$$

and

$$\delta_{x_1 x_1} V_{j_1, j_2}(t) = \frac{V_{j_1+1, j_2}(t) - 2V_{j_1, j_2}(t) + V_{j_1-1, j_2}(t)}{\Delta_1^2},$$

$$\delta_{x_2 x_2} V_{j_1, j_2}(t) = \frac{V_{j_1, j_2+1}(t) - 2V_{j_1, j_2}(t) + V_{j_1, j_2-1}(t)}{\Delta_2^2}.$$

These operators, in turn, give rise to the discrete operators

$$\widehat{\mathcal{L}}_h \triangleq \mu_h(t, x)\delta_{x_h} + \frac{1}{2}\gamma_h(t, x)^2\delta_{x_h x_h} - \frac{1}{2}r(t, x), \quad h = 1, 2,$$

where x is constrained to take values in the spatial grid. A Taylor expansion shows that this operator is second-order accurate (compare to Lemma 2.2.1),

$$(\mathcal{L}_1 + \mathcal{L}_2)V(t, x) = (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)V(t, x) + O(\Delta_1^2 + \Delta_2^2).$$

2.10.1 Theta Method

Turning to a theta-style time discretization, consider first proceeding exactly as in Section 2.2.3. Assuming equidistant time spacing Δ_t , we get for the period $[t_i, t_{i+1}]$,

$$\begin{aligned} & \left(1 - \theta\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) V_{j_1, j_2}(t_i) \\ &= \left(1 + (1 - \theta)\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) V_{j_1, j_2}(t_{i+1}) + e_i^{i+1}, \end{aligned}$$

where

$$e_i^{i+1} = O\left(\Delta_t \left(\Delta_1^2 + \Delta_2^2 + 1_{\{\theta \neq \frac{1}{2}\}} \Delta_t + \Delta_t^2\right)\right),$$

and where it is understood that $\widehat{\mathcal{L}}_1$ and $\widehat{\mathcal{L}}_2$ are to be evaluated at $(t, x) = (t_i^{i+1}(\theta), x_1^{j_1}, x_2^{j_2})$ with $t_i^{i+1}(\theta)$ defined as in (2.14). If $\widehat{V}_{j_1, j_2}(t) \triangleq \widehat{V}(t, x_1^{j_1}, x_2^{j_2})$ is a finite difference approximation to $V_{j_1, j_2}(t)$, we thus get the scheme

$$\left(1 - \theta\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) \widehat{V}_{j_1, j_2}(t_i) = \left(1 + (1 - \theta)\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) \widehat{V}_{j_1, j_2}(t_{i+1}), \quad (2.59)$$

to be solved for the $m_1 m_2$ interior points $\widehat{V}_{j_1, j_2}(t_i)$, $j_1 = 1, \dots, m_1$, $j_2 = 1, \dots, m_2$, given the values of $\widehat{V}_{j_1, j_2}(t_{i+1})$, and given appropriate boundary conditions at $j_1 = 0$, $j_1 = m_1 + 1$, $j_2 = 0$, and $j_2 = m_2 + 1$.

The scheme (2.59) represents a system of linear equations in $m_1 m_2$ unknowns $\{\widehat{V}_{j_1, j_2}(t_i)\}$. When written out as a matrix equation (which requires us to arrange the various $\widehat{V}_{j_1, j_2}(t_i)$ in some order in a $(m_1 m_2)$ -dimensional vector), the matrix to be inverted is sparse but, unfortunately, no longer tri-diagonal. Solution of the system of equations by standard methods (e.g., Gauss-Jordan elimination or LU decomposition) is out of the question due

to the size of the matrix²⁰. We can proceed in two ways: either we use a specialized sparse-matrix solver; or we attempt to redo the discretization (2.59) to make it computationally efficient. We personally prefer the second approach and shall outline one method in the next section. As for the first approach, we simply note that a good iterative sparse solver should be able to solve (2.59) in order $O((m_1 m_2)^{5/4})$ operations. See Saad [2003] for concrete algorithms.

2.10.2 The Alternating Direction Implicit (ADI) Method

The ADI method is an example of a so-called *operator splitting* method, where the simultaneous application of two operators (here $\widehat{\mathcal{L}}_1$ and $\widehat{\mathcal{L}}_2$) is split into two *sequential* operator applications. To illustrate the idea, set $\theta = \frac{1}{2}$ (Crank-Nicolson scheme) in (2.59) and approximate

$$\left(1 - \frac{1}{2}\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) \approx \left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) \left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right), \quad (2.60)$$

$$\left(1 + \frac{1}{2}\Delta_t (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2)\right) \approx \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right). \quad (2.61)$$

It is easy to see²¹ (and to verify, by a Taylor expansion) that the operators on the right-hand sides of these approximations have the same order truncation error as do the left-hand sides, namely $O(\Delta_t(\Delta_1^2 + \Delta_2^2 + \Delta_t^2))$. To the order of our original scheme, no accuracy is gained or lost in using the right-hand sides of (2.60)–(2.61). What is gained, however, is a considerable improvement in computational efficiency, originating in the fact that the resulting scheme

$$\begin{aligned} & \left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) \left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right) \widehat{V}_{j_1,j_2}(t_i) \\ &= \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right) \widehat{V}_{j_1,j_2}(t_{i+1}) \end{aligned} \quad (2.62)$$

can be *split* into the system

$$\left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) U_{j_1,j_2} = \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right) \widehat{V}_{j_1,j_2}(t_{i+1}), \quad (2.63)$$

$$\left(1 - \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_2\right) \widehat{V}_{j_1,j_2}(t_i) = \left(1 + \frac{1}{2}\Delta_t \widehat{\mathcal{L}}_1\right) U_{j_1,j_2}, \quad (2.64)$$

²⁰Recall that the solution of a general linear system with $m_1 m_2$ unknowns is an $O(m_1^2 m_2^2)$ operation. For, say, m_1 and m_2 in the order of 100, this would involve around 1,000,000 times more work than what is required for a one-dimensional (tri-diagonal) scheme ($O(m)$).

²¹To those versed in operator notation, we notice that the right- and left-hand sides both approximate, to identical order, $\exp(\pm 0.5\Delta_t(\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2))$.

where we have introduced an *intermediate value* U_{j_1,j_2} . The advantage of this decomposition is the fact that in each of (2.63) and (2.64), there is only one operator on the left-hand side, leading to simple tri-diagonal equation systems. To formalize this, first define

$$\mathbf{U}_1^{j_2} = (U_{1,j_2}, U_{2,j_2}, \dots, U_{m_1,j_2})^\top.$$

Then, for a fixed value of j_2 we can write for the first step

$$\left(\mathbf{I} - \frac{1}{2} \Delta_t \mathbf{A}_1^{j_2} \left(\frac{t_{i+1} + t_i}{2} \right) \right) \mathbf{U}_1^{j_2} = \mathbf{M}_2^{j_2} \left(\frac{t_{i+1} + t_i}{2} \right), \quad (2.65)$$

where $\mathbf{A}_1^{j_2}$ is an $(m_1 \times m_1)$ -dimensional tri-diagonal matrix of the same form as (2.11) (to get $\mathbf{A}_1^{j_2}$, basically freeze $x_2 = x_2^{j_2}$ and substitute μ_1 and γ_1 for μ and σ in the definition of the one-dimensional matrix \mathbf{A}). The m_1 -dimensional vector $\mathbf{M}_2^{j_2}$ has components $M_{2,j_1}^{j_2}$, $j_1 = 1, \dots, m_1$, given by

$$\begin{aligned} M_{2,j_1}^{j_2} \left(\frac{t_{i+1} + t_i}{2} \right) &= \left(1 + \frac{1}{2} \Delta_t \hat{\mathcal{L}}_2 \right) \hat{V}_{j_1,j_2}(t_{i+1}) \\ &= \frac{1}{2} \varsigma_{j_1,j_2}^- \hat{V}_{j_1,j_2-1}(t_{i+1}) + \frac{1}{2} \varsigma_{j_1,j_2}^+ \hat{V}_{j_1,j_2+1}(t_{i+1}) \\ &\quad + \left(1 - \frac{1}{2} \varsigma_{j_1,j_2} \right) \hat{V}_{j_1,j_2}(t_{i+1}), \end{aligned} \quad (2.66)$$

where we have defined

$$\begin{aligned} \varsigma_{j_1,j_2}^\pm &\triangleq \frac{\Delta_t}{2\Delta_2^2} \left(\gamma_2 \left(\frac{t_{i+1} + t_i}{2}, x_1^{j_1}, x_2^{j_2} \right)^2 \pm \Delta_2 \mu_2 \left(\frac{t_{i+1} + t_i}{2}, x_1^{j_1}, x_2^{j_2} \right) \right), \\ \varsigma_{j_1,j_2} &\triangleq \frac{\Delta_t}{\Delta_2^2} \left(\gamma_2 \left(\frac{t_{i+1} + t_i}{2}, x_1^{j_1}, x_2^{j_2} \right)^2 + \frac{1}{2} \Delta_2^2 r \left(\frac{t_{i+1} + t_i}{2}, x_1^{j_1}, x_2^{j_2} \right) \right). \end{aligned}$$

For known values of $\hat{V}(t_{i+1})$, (2.65) defines a simple tri-diagonal equation system which can be solved for $\mathbf{U}_1^{j_2}$ in $O(m_1)$ operations. Repeating the procedure above for $j_2 = 1, \dots, m_2$ allows us to find U_{j_1,j_2} for all $j_1 = 1, \dots, m_1$, $j_2 = 1, \dots, m_2$, at a total computational cost of $O(m_1 m_2)$.

Turning to the second step of (2.63)–(2.64), we first fix j_1 and define

$$\hat{\mathbf{V}}_2^{j_1}(t) = (\hat{V}_{j_1,1}(t), \hat{V}_{j_1,2}(t), \dots, \hat{V}_{j_1,m_2}(t))^\top.$$

In the same fashion as earlier, we can then write

$$\left(\mathbf{I} - \frac{1}{2} \Delta_t \mathbf{A}_2^{j_1} \left(\frac{t_{i+1} + t_i}{2} \right) \right) \hat{\mathbf{V}}_2^{j_1}(t_i) = \mathbf{M}_1^{j_1} \left(\frac{t_{i+1} + t_i}{2} \right), \quad (2.67)$$

where $\mathbf{A}_2^{j_1}$ is an $(m_2 \times m_2)$ -dimensional tri-diagonal matrix and where the right-hand side vector now has components

$$M_{1,j_2}^{j_1} \left(\frac{t_{i+1} + t_i}{2} \right) = \left(1 + \frac{1}{2} \Delta_t \widehat{\mathcal{L}}_1 \right) U_{j_1,j_2}, \quad j_2 = 1, \dots, m_2.$$

For brevity we omit writing out the $M_{1,j_2}^{j_1}$ (which will be similar to (2.66)), but just notice that the right-hand side of (2.67) is known after the first step of the ADI algorithm (above) is complete. For a given value of j_1 , we can solve the tri-diagonal system (2.67) for $\widehat{\mathbf{V}}_2^{j_1}(t_i)$ in $O(m_2)$ operations. Looping over all m_1 different values of j_1 , the full matrix of time t_i values $\widehat{\mathbf{V}}_{j_1,j_2}(t_i)$, $j_1 = 1, \dots, m_1$, $j_2 = 1, \dots, m_2$, can then be found at a total computational cost of $O(m_1 m_2)$.

The scheme outlined above is known as the *Peaceman-Rachford* scheme. As is the case for all ADI schemes, the scheme works by alternating the directions that are treated fully implicitly in the finite difference grid: in the first step, the x_1 -direction is fully implicit and the x_2 -direction is fully explicit, and in the second step the order is reversed. In effect, both spatial variables end up being discretized “semi-implicitly”, i.e. similar to a Crank-Nicolson scheme, resulting in convergence order is $O(\Delta_1^2 + \Delta_2^2 + \Delta_t^2)$. We emphasize, however, that whereas a direct application of the Crank-Nicolson scheme will involve (if an efficient sparse-matrix solver is used) a computational cost of $O((m_1 m_2)^{5/4})$ per time step, the computational cost of the Peaceman-Rachford ADI scheme is only $O(m_1 m_2)$. A (tedious) von Neumann analysis reveals that the scheme is A -stable, but, like the Crank-Nicolson scheme, not strongly A -stable.

While the Peaceman-Rachford scheme is a classical example of an ADI scheme, there are many others. For instance, consider a theta-version of the *Douglas-Rachford* scheme:

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_1) U_{j_1,j_2} = (1 + (1 - \theta) \Delta_t \widehat{\mathcal{L}}_1 + \Delta_t \widehat{\mathcal{L}}_2) \widehat{\mathbf{V}}_{j_1,j_2}(t_{i+1}), \quad (2.68)$$

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_2) \widehat{\mathbf{V}}_{j_1,j_2}(t_i) = U_{j_1,j_2} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{\mathbf{V}}_{j_1,j_2}(t_{i+1}), \quad (2.69)$$

where we understand that in $\widehat{\mathcal{L}}_1$ and $\widehat{\mathcal{L}}_2$ the PDE coefficients are to be evaluated at time $t_i^{i+1}(\theta)$. Again, notice how the scheme consists of two steps, each involving the solution of tri-diagonal sets of equations along only one of the x_1 - or x_2 -directions. The computational cost thus remains at $O(m_1 m_2)$. It can be shown that the convergence order of this scheme is $O(\Delta_1^2 + \Delta_2^2 + 1_{\{\theta \neq \frac{1}{2}\}} \Delta_t + \Delta_t^2)$ and it is A -stable for $\theta \geq \frac{1}{2}$, and strongly A -stable for $\theta > \frac{1}{2}$. By elimination of U_{j_1,j_2} we note that the unsplit version of the Douglas-Rachford scheme is

$$\begin{aligned} & (1 - \theta \Delta_t \widehat{\mathcal{L}}_1) (1 - \theta \Delta_t \widehat{\mathcal{L}}_2) \widehat{\mathbf{V}}_{j_1,j_2}(t_i) \\ &= ((1 - \theta \Delta_t \widehat{\mathcal{L}}_1) (1 - \theta \Delta_t \widehat{\mathcal{L}}_2) + \Delta_t \widehat{\mathcal{L}}_1 + \Delta_t \widehat{\mathcal{L}}_2) \widehat{\mathbf{V}}_{j_1,j_2}(t_{i+1}). \end{aligned}$$

It is not difficult to see that this approximates (2.59) to second order.

2.10.3 Boundary Conditions and Other Issues

The fact that ADI schemes reduce to solving sequences of matrix systems identical to the ones arising in the one-dimensional case is convenient, in the sense that many of the issues we have encountered for one-dimensional finite difference grids (oscillations, stability, convection dominance, etc.) and their remedies (smoothing, non-equidistant discretization, upwinding, etc.) carry over to the ADI setting with only minor modifications. Consider for instance the issue of applying spatial boundary conditions along the edges of the (x_1, x_2) domain, which we have so far not discussed. As for the one-dimensional PDEs, the most convenient way to express such boundary conditions is typically by imposing conditions on derivatives, like $\partial^2 V(t, x_1^0, x_2^{j_2}) / \partial x_1^2 = \partial V(t, x_1^0, x_2^{j_2}) / \partial x_1$ and so forth. For the Peaceman-Rachford scheme, say, such conditions can be incorporated directly into (2.65) and (2.67) by altering the matrices $\mathbf{A}_1^{j_2}$ and $\mathbf{A}_2^{j_1}$, as well as the boundary elements of $\mathbf{M}_1^{j_1}$ and $\mathbf{M}_2^{j_2}$, in the manner outlined in Section 2.2.1. If instead we wish to impose Dirichlet boundary conditions, we need to add corrective terms to the tri-diagonal systems, as in (2.19). To complete the first part of the split scheme, this then requires us to establish what boundary terms are needed for the intermediate quantity U_{j_1, j_2} , i.e. we must define $U_{j_1, 0}$ and U_{j_1, m_2+1} for $j_1 = 1, \dots, m_1$, as well as U_{0, j_2} and U_{m_1+1, j_2} for $j_2 = 1, \dots, m_2$. While U_{j_1, j_2} is a purely mathematic construct, sometimes it is adequate to think of U_{j_1, j_2} as a proxy for V_{j_1, j_2} evaluated at $t_i^{i+1}(\theta)$, which obviously makes determination of boundary conditions straightforward. For maximum precision, however, we should use the ADI equations themselves to express the boundary conditions of U directly in terms of boundary conditions for $V(t_i)$ and $V(t_{i+1})$. Here, the Douglas-Rachford scheme is particularly easy to deal with, as a rearrangement of (2.69) directly relates U_{j_1, j_2} to $\widehat{V}_{j_1, j_2}(t_i)$ and $\widehat{V}_{j_1, j_2}(t_{i+1})$,

$$U_{j_1, j_2} = (1 - \theta \Delta_t \widehat{\mathcal{L}}_2) \widehat{V}_{j_1, j_2}(t_i) + \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}_{j_1, j_2}(t_{i+1}).$$

The Peaceman-Rachford scheme requires some further manipulations to express U in terms of $V(t_i)$ and $V(t_{i+1})$; see Mitchell and Griffiths [1980] for the details.

2.11 Two-Dimensional PDE with Mixed Derivatives

Consider now the case where the 2-dimensional PDE (2.58) has a mixed partial derivative,

$$\frac{\partial V}{\partial t} + (\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_{1,2}) V = 0, \quad (2.70)$$

where \mathcal{L}_1 and \mathcal{L}_2 are as in (2.58), and where

$$\mathcal{L}_{1,2} = s_{1,2}(t, x) \frac{\partial^2}{\partial x_1 \partial x_2} \triangleq \rho(t, x) \gamma_1(t, x) \gamma_2(t, x) \frac{\partial^2}{\partial x_1 \partial x_2}. \quad (2.71)$$

The quantity $\rho(t, x)$ is the instantaneous correlation between the processes $x_1(t)$ and $x_2(t)$ in (2.57), i.e. $\rho(t, x) \in [-1, 1]$.

The presence of $\mathcal{L}_{1,2}$ prevents a direct application of the ADI methods in Section 2.10.2, since the mixed operator $\mathcal{L}_{1,2}$ is not amenable to operator splitting. We shall demonstrate two ways to overcome this problem: a) orthogonalization of the PDE; and b) predictor-corrector schemes.

2.11.1 Orthogonalization of the PDE

The idea here is to introduce new variables $y_1(t, x_1, x_2)$ and $y_2(t, x_1, x_2)$ such that the PDE loses its mixed derivative term when stated in terms of these variables. To demonstrate this idea, assume first that $\rho(t, x)$, $\gamma_1(t, x)$, and $\gamma_2(t, x)$ are all functions of time only and independent of x . Then define, say,

$$y_1(t, x_1, x_2) = x_1, \quad (2.72)$$

$$y_2(t, x_1, x_2) = -\rho(t) \frac{\gamma_2(t)}{\gamma_1(t)} x_1 + x_2 \triangleq a(t)x_1 + x_2, \quad (2.73)$$

where we must assume that $\gamma_1(t) \neq 0$ for all t .

Lemma 2.11.1. *Consider the PDE (2.70) subject to the terminal value condition $V(T, x) = g(x)$. Define $y = (y_1, y_2)^\top$ and $v(t, y) = V(t, x)$. With the variable change defined in (2.72)–(2.73), v satisfies*

$$\begin{aligned} \frac{\partial v}{\partial t} + \mu_1^y(t, y) \frac{\partial v}{\partial y_1} + \mu_2^y(t, y) \frac{\partial v}{\partial y_2} + \frac{1}{2} \gamma_1(t)^2 \frac{\partial^2 v}{\partial y_1^2} \\ + \frac{1}{2} (1 - \rho(t)^2) \gamma_2(t)^2 \frac{\partial^2 v}{\partial y_2^2} - r(t, y_1, y_2 - a(t)y_1) = 0, \end{aligned} \quad (2.74)$$

where

$$\mu_1^y(t, y) \triangleq \mu_1(t, x_1, x_2) = \mu_1(t, y_1, y_2 - a(t)y_1), \quad (2.75)$$

$$\begin{aligned} \mu_2^y(t, y) &\triangleq \frac{da(t)}{dt} x_1 + a(t) \mu_1(t, x_1, x_2) + \mu_2(t, x_1, x_2) \\ &= \frac{da(t)}{dt} y_1 + a(t) \mu_1^y(t, y) + \mu_2(t, y_1, y_2 - a(t)y_1). \end{aligned} \quad (2.76)$$

The equation (2.74) is subject to the terminal value condition $v(T, y_1, y_2) = g(x_1, x_2) = g(y_1, y_2 - a(T)y_1)$.

Proof. While the result can be established by the usual mechanics of ordinary calculus, we will take the opportunity to show how stochastic calculus can

also conveniently prove results of this type. Going back to the processes underlying the PDE (see (2.57)), we write

$$dx_1(t) = \mu_1(t, x) dt + \gamma_1(t) dW_1(t), \quad (2.77)$$

$$dx_2(t) = \mu_2(t, x) dt + \gamma_2(t) \left(\rho(t) dW_1(t) + \sqrt{1 - \rho(t)^2} dW_2(t) \right), \quad (2.78)$$

for independent scalar Brownian motions $W_1(t)$ and $W_2(t)$; this is easily seen to generate the correct correlation $\rho(t)$ between x_1 and x_2 . An application of Ito's lemma then shows that the processes for y_1 and y_2 are

$$\begin{aligned} dy_1(t) &= dx_1(t) = \mu_1(t, x) dt + \gamma_1(t) dW_1(t), \\ dy_2(t) &= \frac{da(t)}{dt} x_1(t) dt + a(t) \mu_1(t, x) dt + a(t) \gamma_1(t) dW_1(t) \\ &\quad + \mu_2(t, x) dt + \gamma_2(t) \left(\rho(t) dW_1(t) + \sqrt{1 - \rho(t)^2} dW_2(t) \right) \\ &= \left(\frac{da(t)}{dt} x_1(t) + a(t) \mu_1(t, x) + \mu_2(t, x) \right) dt \\ &\quad + \gamma_2(t) \sqrt{1 - \rho(t)^2} dW_2(t). \end{aligned}$$

With the definitions (2.75)–(2.76), this becomes simply

$$dy_1(t) = \mu_1^y(t, y(t)) dt + \gamma_1(t) dW_1(t), \quad (2.79)$$

$$dy_2(t) = \mu_2^y(t, y(t)) dt + \gamma_2(t) \sqrt{1 - \rho(t)^2} dW_2(t). \quad (2.80)$$

Equations (2.79)–(2.80) define a Markov SDE in $y_1(t)$ and $y_2(t)$ where, importantly, the Brownian motions on $y_1(t)$ and $y_2(t)$ are now *independent*. Writing $V(t, x) = v(t, y)$, it then follows immediately from the backward Kolmogorov equation (see Section 1.8) that v satisfies the PDE (2.74). \square

Through the chosen transformation (2.72)–(2.73), our original PDE has now been put into a form where we can immediately apply the ADI schemes outlined in Section 2.10.2.

In performing the orthogonalization of the PDE in Lemma 2.11.1 we relied on $\rho(t, x)$, $\gamma_1(t, x)$, and $\gamma_2(t, x)$ all being independent of x . This can often be relaxed. Consider for instance the case where $\rho(t, x) = \rho(t)$, $\gamma_1(t, x) = \gamma_1(t, x_1)$, and $\gamma_2(t, x) = \gamma_2(t, x_2)$; here the correlation ρ is still assumed deterministic, but we now allow for some (though not full) x -dependence in γ_1 and γ_2 . Assuming that $\gamma_1(t, x_1) > 0$ and $\gamma_2(t, x_2) > 0$ we can introduce new variables

$$z_1(t, x_1) = \int \frac{1}{\gamma_1(t, x_1)} dx_1, \quad (2.81)$$

$$z_2(t, x_2) = \int \frac{1}{\gamma_2(t, x_2)} dx_2. \quad (2.82)$$

Applying Ito's lemma to (2.77)–(2.78) we see that

$$dz_1(t, x_1) = \left(- \int \frac{\partial \gamma_1(t, x_1)}{\partial t} \frac{1}{\gamma_1(t, x_1)^2} dx_1 + \frac{\mu_1(t, x)}{\gamma_1(t, x_1)} - \frac{1}{2} \frac{\partial \gamma_1(t, x_1)}{\partial x_1} \right) dt + dW_1(t) \quad (2.83)$$

and

$$dz_2(t, x_2) = \left(- \int \frac{\partial \gamma_2(t, x_2)}{\partial t} \frac{1}{\gamma_2(t, x_2)^2} dx_2 + \frac{\mu_2(t, x)}{\gamma_2(t, x_2)} - \frac{1}{2} \frac{\partial \gamma_2(t, x_2)}{\partial x_2} \right) dt + \rho(t) dW_1(t) + \sqrt{1 - \rho(t)^2} dW_2(t). \quad (2.84)$$

As we assumed that $\gamma_1(t, x_1) > 0$ and $\gamma_2(t, x_2) > 0$, the functions z_1 and z_2 are increasing in x_1 and x_2 , respectively, and are thereby invertible. As such, we can rewrite (2.83)–(2.84) in the more appealing form

$$\begin{aligned} dz_1(t, x_1) &= \mu_1^z(t, z_1, z_2) dt + dW_1(t), \\ dz_2(t, x_1) &= \mu_2^z(t, z_1, z_2) dt + \rho(t) dW_1(t) + \sqrt{1 - \rho(t)^2} dW_2(t). \end{aligned}$$

Through the transformation (2.81)–(2.82), we have reduced our original system to one where the coefficients on $W_1(t)$ and $W_2(t)$ are no longer state-dependent, similar to the case that lead to Lemma 2.11.1. We can now proceed with another variable transformation, as in (2.72)–(2.73), to orthogonalize the system and prepare it for an application of the ADI method.

While the orthogonalization method outlined here can be very effective on a range of practical problems, it suffers from a few drawbacks. Most obviously, the method is not completely general and requires a certain structure on the parameters of the PDE. Another drawback is that the introduction of a time-dependent transformation on one or more variables (Lemma 2.11.1) often makes the alignment of the finite difference grid along (time-independent) critical level points in x -space impossible. Also, the introduction of terms like $y_1 da(t)/dt$ in the drift of y_2 (see (2.76)) can be problematic, particularly if the functions $\gamma_1(t)$ and $\gamma_2(t)$ are not smooth. For instance, it is not unlikely that $y_1 da(t)/dt$ will locally be of such magnitude that upwinding will be necessary to prevent oscillations; see Section 2.6.1. Further, we note that inversion of the transformations (2.81)–(2.82) will not always be possible to perform analytically and may require numerical (root-search) work, complicating the scheme and potentially slowing it down. Finally, as we shall highlight in future chapters, maintaining the “continuity” of a numerical scheme with respect to input parameters is of critical importance for the smoothness of risk sensitivities. Such continuity is difficult to ensure if complicated transformations are applied to model variables. So, in the end, we recommend formulating the PDEs in terms of financially meaningful variables, avoiding excessive transformations, and relying on methods such as developed in the next section when dealing with mixed derivatives and other numerical complications.

2.11.2 Predictor-Corrector Scheme

In this section we shall consider a completely general method for handling mixed derivatives in two-dimensional PDEs. While a bit slower than the method outlined in Section 2.11.1, it does not involve any variable transformations and, by extension, does not suffer from the drawbacks associated with such transformations. As a first step, consider the discretization of the mixed derivative $\partial^2 V / \partial x_1 \partial x_2$. There are a few possibilities (see Mitchell and Griffiths [1980]), but we shall just use

$$\begin{aligned} \delta_{x_1 x_2} V_{j_1, j_2}(t) &= \delta_{x_1} \delta_{x_2} V_{j_1, j_2}(t) \\ &= \frac{V_{j_1+1, j_2+1}(t) - V_{j_1+1, j_2-1}(t) - V_{j_1-1, j_2+1}(t) + V_{j_1-1, j_2-1}(t)}{4\Delta_1 \Delta_2}. \end{aligned} \quad (2.85)$$

Extensions to non-equidistant grids follow directly from (2.27) and the relation $\delta_{x_1 x_2} V_{j_1, j_2}(t) = \delta_{x_1} \delta_{x_2} V_{j_1, j_2}(t)$. As we have not encountered mixed difference operators before, for completeness we show the following lemma.

Lemma 2.11.2. *For the discrete operator (2.85) we have*

$$\delta_{x_1 x_2} V_{j_1, j_2}(t) = \frac{\partial^2 V(t, x_1^{j_1}, x_2^{j_2})}{\partial x_1 \partial x_2} + O(\Delta_1^2 + \Delta_2^2).$$

Proof. A Taylor expansion of $V(t, x)$ around the point $x = (x_1^{j_1}, x_2^{j_2})^\top$ gives

$$\begin{aligned} V_{j_1+1, j_2 \pm 1}(t) &= V_{j_1, j_2}(t) + \Delta_1 \frac{\partial V}{\partial x_1} \pm \Delta_2 \frac{\partial V}{\partial x_2} + \frac{1}{2} \Delta_1^2 \frac{\partial^2 V}{\partial x_1^2} + \frac{1}{2} \Delta_2^2 \frac{\partial^2 V}{\partial x_2^2} \\ &\quad \pm \Delta_1 \Delta_2 \frac{\partial^2 V}{\partial x_1 \partial x_2} + \frac{1}{6} \Delta_1^3 \frac{\partial^3 V}{\partial x_1^3} \pm \frac{1}{6} \Delta_2^3 \frac{\partial^3 V}{\partial x_2^3} \\ &\quad + \frac{1}{2} \Delta_1 \Delta_2^2 \frac{\partial^3 V}{\partial x_1 \partial x_2^2} \pm \frac{1}{2} \Delta_1^2 \Delta_2 \frac{\partial^3 V}{\partial x_1^2 \partial x_2} + \dots, \\ V_{j_1-1, j_2 \pm 1}(t) &= V_{j_1, j_2}(t) - \Delta_1 \frac{\partial V}{\partial x_1} \pm \Delta_2 \frac{\partial V}{\partial x_2} + \frac{1}{2} \Delta_1^2 \frac{\partial^2 V}{\partial x_1^2} + \frac{1}{2} \Delta_2^2 \frac{\partial^2 V}{\partial x_2^2} \\ &\quad \mp \Delta_1 \Delta_2 \frac{\partial^2 V}{\partial x_1 \partial x_2} - \frac{1}{6} \Delta_1^3 \frac{\partial^3 V}{\partial x_1^3} \pm \frac{1}{6} \Delta_2^3 \frac{\partial^3 V}{\partial x_2^3} \\ &\quad - \frac{1}{2} \Delta_1 \Delta_2^2 \frac{\partial^3 V}{\partial x_1 \partial x_2^2} \pm \frac{1}{2} \Delta_1^2 \Delta_2 \frac{\partial^3 V}{\partial x_1^2 \partial x_2} + \dots. \end{aligned}$$

A little thought then shows that

$$\begin{aligned} V_{j_1+1, j_2+1}(t) - V_{j_1+1, j_2-1}(t) - V_{j_1-1, j_2+1}(t) + V_{j_1-1, j_2-1}(t) \\ = 4\Delta_1 \Delta_2 \frac{\partial^2 V}{\partial x_1 \partial x_2} + O(\Delta_1^3 \Delta_2 + \Delta_1 \Delta_2^3), \end{aligned}$$

as error terms of order Δ_1^4 , Δ_2^4 , and $\Delta_1^2 \Delta_2^2$ will cancel. The result follows.

□

Equipped with (2.85), we can approximate the operator $\mathcal{L}_{1,2}$ in (2.71) as

$$\widehat{\mathcal{L}}_{1,2} V_{j_1,j_2}(t) \triangleq \rho(t, x_1^{j_1}, x_2^{j_2}) \gamma_1(t, x_1^{j_1}, x_2^{j_2}) \gamma_2(t, x_1^{j_1}, x_2^{j_2}) \delta_{x_1 x_2} V_{j_1,j_2}(t),$$

which is accurate to order $O(\Delta_1^2 + \Delta_2^2)$. The first easy way to modify our ADI scheme to incorporate $\widehat{\mathcal{L}}_{1,2}$ is to treat the mixed derivative fully explicitly. In the Douglas-Rachford scheme (2.68)–(2.69), for instance, we thus modify the right-hand side of the first step as follows:

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_1) U_{j_1,j_2} = (1 + (1 - \theta) \Delta_t \widehat{\mathcal{L}}_1 + \Delta_t \widehat{\mathcal{L}}_2 + \Delta_t \widehat{\mathcal{L}}_{1,2}) \widehat{V}_{j_1,j_2}(t_{i+1}), \quad (2.86)$$

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_2) \widehat{V}_{j_1,j_2}(t_i) = U_{j_1,j_2} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}_{j_1,j_2}(t_{i+1}). \quad (2.87)$$

The addition of $\widehat{\mathcal{L}}_{1,2}$ this way clearly preserves the ADI structure of the scheme which will continue to involve only sequences of tri-diagonal linear equations. However, having, in effect, only a one-sided time-differencing of the mixed derivative term will lower the convergence order of the time step to $O(\Delta_t)$, irrespective of the choice of θ .

To change the time at which the mixed operator $\widehat{\mathcal{L}}_{1,2}$ is evaluated, consider using a *predictor-corrector* scheme, where the results of (2.86)–(2.87) are re-used in a one-time²² iteration. Specifically, we write, for some $\lambda \in [0, 1]$,

Predictor:

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_1) U_{j_1,j_2}^{(1)} = (1 + (1 - \theta) \Delta_t \widehat{\mathcal{L}}_1 + \Delta_t \widehat{\mathcal{L}}_2 + \Delta_t \widehat{\mathcal{L}}_{1,2}) \widehat{V}_{j_1,j_2}(t_{i+1}), \quad (2.88)$$

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_2) U_{j_1,j_2}^{(2)} = U_{j_1,j_2}^{(1)} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}_{j_1,j_2}(t_{i+1}). \quad (2.89)$$

Corrector:

$$\begin{aligned} (1 - \theta \Delta_t \widehat{\mathcal{L}}_1) Z_{j_1,j_2}^{(1)} &= (1 + (1 - \theta) \Delta_t \widehat{\mathcal{L}}_1 + \Delta_t \widehat{\mathcal{L}}_2 \\ &\quad + (1 - \lambda) \Delta_t \widehat{\mathcal{L}}_{1,2}) \widehat{V}_{j_1,j_2}(t_{i+1}) + \lambda \Delta_t \widehat{\mathcal{L}}_{1,2} U_{j_1,j_2}^{(2)}, \end{aligned} \quad (2.90)$$

$$(1 - \theta \Delta_t \widehat{\mathcal{L}}_2) \widehat{V}_{j_1,j_2}(t_i) = Z_{j_1,j_2}^{(1)} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}_{j_1,j_2}(t_{i+1}). \quad (2.91)$$

²²We can run the iteration more than once if desired, but a single iteration will normally suffice.

Notice how the Douglas-Rachford scheme is first run once, in (2.88)–(2.89), to yield a first guess (a “predictor”), $U_{j_1,j_2}^{(2)}$, for the time t_i value $V_{j_1,j_2}(t_i)$. In a second run of the Douglas-Rachford scheme, in (2.90)–(2.91), this guess is used as a “corrector” to affect the time at which $\widehat{\mathcal{L}}_{1,2}$ is evaluated, by applying this operator to $(1 - \lambda)\widehat{V}_{j_1,j_2}(t_{i+1}) + \lambda U_{j_1,j_2}^{(2)}$; when $\lambda = \frac{1}{2}$ we effectively center the time-differencing of the mixed term. The scheme now relies on three intermediate variables, $U_{j_1,j_2}^{(1)}$, $U_{j_1,j_2}^{(2)}$, and $Z_{j_1,j_2}^{(1)}$.

The combined predictor-corrector scheme above (in a slightly less general form, with $\Delta_1 = \Delta_2$) was suggested by Craig and Sneyd [1988]. It can be shown that the scheme has convergence order

$$O\left((\Delta_1 + \Delta_2)^2 + 1_{\{\theta \neq \frac{1}{2}\}}\Delta_t + 1_{\{\lambda \neq \frac{1}{2}\}}\Delta_t + \Delta_t^2\right),$$

so second order convergence in the time domain is still achievable by setting $\theta = \lambda = \frac{1}{2}$. The scheme will be A -stable for $\theta \geq \frac{1}{2}$ and $\frac{1}{2} \leq \lambda \leq \theta$. The computational cost of the predictor-corrector is clearly still $O(m_1 m_2)$ per time step, as both the predictor and corrector schemes have $O(m_1 m_2)$ cost per time-step. Even though the standard Douglas-Rachford scheme is effectively run twice, we should point out that when intelligently implemented, (2.88)–(2.91) is typically only about 30-40% slower than the Douglas-Rachford scheme, as a number of results from the predictor step can be cached and reused in the corrector step.

As for the standard ADI grids, extensions to non-equidistant grids are straightforward using the techniques in Section 2.4. Boundary conditions in the x -domain are imposed along the lines outlined in Section 2.10.3.

2.12 PDEs of Arbitrary Order

We now turn our attention back to the general p -dimensional PDE (2.56). To prepare for a numerical scheme, let us rewrite the PDE as follows:

$$\frac{\partial V}{\partial t} + \sum_{h=1}^p \mathcal{L}_h V + \sum_{h=1}^p \sum_{l=h+1}^p \mathcal{L}_{h,l} V = 0, \quad (2.92)$$

where

$$\begin{aligned} \mathcal{L}_h &= \mu_h(t, x) \frac{\partial}{\partial x_h} + \frac{1}{2} s_{h,h}(t, x) \frac{\partial^2}{\partial x_h^2} - p^{-1} r(t, x), \\ \mathcal{L}_{h,l} &= s_{h,l}(t, x) \frac{\partial^2}{\partial x_h \partial x_l}. \end{aligned}$$

The method we present here for solution of (2.92) is a p -dimensional version of the predictor-corrector scheme outlined above. The extension

is straightforward and we simply list it here without further discussion; see Craig and Sneyd [1988] for additional background. To simplify notation, we have omitted sub-indices everywhere (i.e., $\widehat{V}(t_i)$ is used instead of $\widehat{V}_{j_1, j_2, \dots, j_p}(t_i)$).

Predictor:

$$\begin{aligned} & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_1\right) U^{(1)} \\ &= \Delta_t \left(\Delta_t^{-1} + (1 - \theta) \widehat{\mathcal{L}}_1 + \sum_{h=2}^p \widehat{\mathcal{L}}_h + \sum_{h=1}^p \sum_{l=h+1}^p \widehat{\mathcal{L}}_{h,l} \right) \widehat{V}(t_{i+1}), \\ & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_2\right) U^{(2)} = U^{(1)} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}(t_{i+1}), \\ & \vdots \\ & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_p\right) U^{(p)} = U^{(p-1)} - \theta \Delta_t \widehat{\mathcal{L}}_p \widehat{V}(t_{i+1}). \end{aligned}$$

Corrector:

$$\begin{aligned} & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_1\right) Z^{(1)} \\ &= \Delta_t \left(\Delta_t^{-1} + (1 - \theta) \widehat{\mathcal{L}}_1 + \sum_{h=2}^p \widehat{\mathcal{L}}_h \right. \\ & \quad \left. + (1 - \lambda) \sum_{h=1}^p \sum_{l=h+1}^p \widehat{\mathcal{L}}_{h,l} \right) \widehat{V}(t_{i+1}) + \lambda \Delta_t \sum_{h=1}^p \sum_{l=h+1}^p \widehat{\mathcal{L}}_{h,l} U^{(p)}, \\ & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_2\right) Z^{(2)} = Z^{(1)} - \theta \Delta_t \widehat{\mathcal{L}}_2 \widehat{V}(t_{i+1}), \\ & \vdots \\ & \left(1 - \theta \Delta_t \widehat{\mathcal{L}}_p\right) \widehat{V}(t_i) = Z^{(p-1)} - \theta \Delta_t \widehat{\mathcal{L}}_p \widehat{V}(t_{i+1}). \end{aligned}$$

With m_h points in the x_h -direction, $h = 1, \dots, p$, the computational cost of the predictor-corrector scheme is $O(\prod_{h=1}^p m_h)$. For $p \leq 3$, sufficient conditions for A -stability are $\theta \geq \frac{1}{2}$ and $\frac{1}{2} \leq \lambda \leq \theta$. For $p \geq 4$, sufficient conditions are $\theta \leq \frac{1}{2}$ and

$$\frac{1}{2} \leq \lambda \leq \frac{p^{p-1}}{(p-1)^p} \theta.$$

See Craig and Sneyd [1988] for a proof. Convergence is similar to the two-dimensional case.

As a final comment, let us note that as dimensionality increases, the computational complexity of an iterative sparse solver will start approaching that of ADI. Specifically, for a p -dimensional problem, the complexity of the former is $O(m_{\text{total}})$ and for the latter $O(m_{\text{total}}^{(2p+1)/2p})$, with $m_{\text{total}} = m_1 \cdot m_2 \cdot \dots \cdot m_p$.

Monte Carlo Methods

While the finite difference method is flexible and powerful, it has a number of limitations. First, its usage is restricted to problems where the state variable dynamics are Markovian. Second, for strongly path-dependent problems, the method often does not apply. And third, it is unsuited for problems where the dimension of the underlying vector of state variables is high. To expand on the last point, recall from Section 2.9 that the (ADI) finite difference method applied to a p -dimensional problem has computational complexity $O(m^p)$ per time step, where m is the average number of spatial points per dimension. The exponential growth in p — the “curse of dimensionality” — is typical of grid-based methods and prevents the practical usage of the method for p larger than about 4 or 5.

In this chapter, we study the *Monte Carlo method*, a numerical technique where the computational effort grows only linearly in the problem dimension p . While convergence of the Monte Carlo method is relatively slow, it is nearly always the method of choice for high-dimensional pricing problems. Compared to finite difference methods, Monte Carlo methods are easy to apply to problems with non-Markovian dynamics as well as strong path-dependency in the payout. On the other hand, as Monte Carlo methods inherently run forward in time, dynamic programming techniques are challenging to implement, making Monte Carlo pricing of American and Bermudan options significantly more involved than for the naturally backward-working finite difference method.

3.1 Fundamentals

Consider a European-style derivative V with time T payout $V(T) = g(T)$, where $g(T)$ is an \mathcal{F}_T -measurable (and integrable) random variable. Where finite difference methods start with a PDE representation of the price of a contingent claim at times $t < T$, the starting point for the Monte Carlo method is the basic martingale relation (see (1.15))

$$V(t) = N(t) \mathbb{E}_t^{Q^N} (g(T)/N(T)), \quad (3.1)$$

where $N(\cdot)$ is a numeraire and Q^N is the measure induced by $N(\cdot)$. To evaluate this expression numerically, we need a numerical technique to compute expectations of a random variable. For this, we turn to the law of large numbers:

Theorem 3.1.1 (Strong Law of Large Numbers). *Let Y_1, Y_2, \dots be a sequence of independent identically distributed (i.i.d.) random variables with expectation $\mu < \infty$. Define the sample mean*

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (3.2)$$

Then

$$\lim_{n \rightarrow \infty} \bar{Y}_n = \mu, \quad a.s.$$

This result forms the basis for the *Monte Carlo method*, which computes the expectation in (3.1) by simply i) generating independent realizations of $g(T)/N(T)$ under Q^N ; and ii) forming their average. Specifically, let $g_1/N_1, \dots, g_n/N_n$ denote n independent samples from the distribution of $g(T)/N(T)$, conditional on \mathcal{F}_t . Then our Monte Carlo estimator for $V(t)$ is the sample mean

$$\bar{V}(t) = N(t) \frac{1}{n} \sum_{i=1}^n g_i/N_i. \quad (3.3)$$

We shall delve into how to generate samples from the distribution of $g(T)/N(T)$ shortly, but before doing so let us consider the expected convergence rate of the Monte Carlo method as n is increased. The key result is here the central limit theorem:

Theorem 3.1.2 (Central Limit Theorem). *Let Y_1, Y_2, \dots be a sequence of i.i.d. random variables with expectation μ and standard deviation $\sigma < \infty$. Let the sample mean be defined as in (3.2). Then, for $n \rightarrow \infty$,*

$$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\mathcal{N}(0, 1)$ is a standard Gaussian distribution and \xrightarrow{d} denotes convergence in distribution¹. Further, if we define

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2},$$

¹Recall that a sequence of variables X_n with cumulative distribution functions F_n converge in distribution to a random variable X with distribution F if $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$ at which $F(x)$ is continuous.

then also

$$\frac{\bar{Y}_n - \mu}{s_n/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Define the Gaussian percentile u_γ as $\Phi(u_\gamma) = 1 - \gamma$, where Φ is the Gaussian cumulative distribution function. From Theorem 3.1.2, and from the definition of convergence in distribution (see footnote 1), the probability that the confidence interval

$$[\bar{V}(t) - u_{\gamma/2} \cdot s_n/\sqrt{n}, \bar{V}(t) + u_{\gamma/2} \cdot s_n/\sqrt{n}] \quad (3.4)$$

fails to include the true value $V(t)$ approaches γ for large n . Here

$$s_n \triangleq \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{g_i N(t)}{N_i} - \bar{V}(t) \right)^2},$$

with the quantity s_n/\sqrt{n} known as the *standard error*. For given γ , the rate at which the confidence interval for $V(t)$ contracts is $O(n^{-\frac{1}{2}})$. This is relatively slow: to reduce the width of the interval by a factor of 2, n must increase by a factor of 4. On the other hand, we notice that the (asymptotic) convergence rate only depends on n , not on the specifics of the g_i 's. In particular, if $g(T) = g(X(T))$ where X is p -dimensional, the asymptotic convergence rate is independent of p . As we shall see shortly, in most applications the work required to generate samples of $g(X(T))$ is (at most) linear in p .

3.1.1 Generation of Random Samples

At the most basic level, the Monte Carlo method requires the ability to draw independent realizations of a scalar random variable Z with a specified cumulative distribution function $F(z) = P(Z \leq z)$, where P is a probability measure. On a computer, the starting point for this exercise is normally a *pseudo-random number generator*, a software program that will generate a sequence of numbers uniformly distributed on $[0, 1]$ (i.e. from $\mathcal{U}(0, 1)$). Press et al. [1992] list a number of generators producing sequences of uniform numbers u_1, u_2, \dots from iterative relationships of the form

$$\begin{aligned} I_{i+1} &= (aI_i + c) \bmod(m), \\ u_{i+1} &= I_{i+1}/m. \end{aligned}$$

The externally specified starting point I_0 is the *seed* of the random number generator. In this so-called *general linear congruential generator*, the choice of the *multiplier* a , the *modulus* m , and the *increment* c must be done

with great care to ensure that the period length of the generator is large² and that the resulting algorithm is efficient on a computer. The latter, for instance, can be accomplished by setting m to be a power of 2 such that the modulo operation can be done by bit-shifting. For detailed discussion and a number of concrete algorithms (including computer code), we refer to Press et al. [1992]. The algorithms in Press et al. [1992] should suffice for most fixed income applications, but we should note the existence of more sophisticated methods that (theoretically, at least) have better performance than linear congruential generators. For instance, the so-called *Mersenne twister* proposed in Matsumoto and Nishimura [1998] has become popular, especially the specific variant MT19937 which has a period of $2^{19937} - 1$. For an extensive survey of pseudo-random number generators, see L'Ecuyer [1994].

So far we have only discussed techniques to generate $\mathcal{U}(0, 1)$ numbers, but many methods exist to convert uniformly distributed numbers into draws from the distribution F of Z . We cover a few important techniques next.

3.1.1.1 Inverse Transform Method

The idea of the inverse transform method is straightforward. Let U be a random variable uniformly distributed on $[0, 1]$, and consider setting

$$Z = F^{-1}(U), \quad (3.5)$$

where we assume that F^{-1} is well-defined, for all but a finite number of points³. As desired,

$$\mathbb{P}(Z \leq z) = \mathbb{P}(F^{-1}(U) \leq z) = \mathbb{P}(U \leq F(z)) = F(z),$$

where the last equality follows from the property of uniformly distributed random variables. The inverse transform method (3.5) is quite general, but its practical usefulness hinges on being able to compute F^{-1} fast. Many distributions allow for closed-form inversion; this includes the *exponential distribution* where $F(z) = 1 - e^{-z\lambda}$ for some positive constant λ , and the *Cauchy distribution* where $F(z) = 1/2 + (1/\pi) \arctan((z-t)/s)$ for constants t and $s > 0$.

For the important case of the Gaussian distribution, no closed-form expression for the inverse distribution exists. Nevertheless, the inverse transform

²Note that if a number $I_k = I_i$, the sequences starting from I_k and I_i are identical. In practice, we would want the generator to have *full period*, in the sense that the sequence would produce $m - 1$ distinct values before repeating the sequence.

³For discrete random variables, the distribution function is discontinuous around each of the possible (discrete) outcomes of Z . We can handle this by simply defining $F^{-1}(u) = \inf\{z : F(z) \geq u\}$.

method can still be applied as fast and extremely accurate approximations for Φ^{-1} exist. For instance, Beasley and Springer [1977] suggest the rational approximation

$$\Phi^{-1}(x) \approx \frac{\sum_{i=0}^3 a_i (x - \frac{1}{2})^{2i+1}}{1 + \sum_{i=0}^3 b_i (x - \frac{1}{2})^{2i}}, \quad 0.5 \leq x \leq 0.92, \quad (3.6)$$

for constants $a_i, b_i, i = 0, \dots, 3$, listed in Appendix 3.A. For values of x greater than 0.92, Moro [1995] proposes the approximation

$$\Phi^{-1}(x) \approx \sum_{i=1}^8 c_i [\ln(-\ln(1-x))]^i, \quad 0.92 \leq x < 1, \quad (3.7)$$

for constants $c_i, i = 0, \dots, 8$, given in Appendix 3.A. Taken together, (3.6) and (3.7) provide an approximation valid for $0.5 \leq x < 1$; when $0 < x < 0.5$ we can compute $\Phi^{-1}(x)$ by symmetry: $\Phi^{-1}(1-x) = -\Phi^{-1}(x)$. The precision of (3.6)–(3.7) is excellent⁴, with the error less than 3×10^{-9} for x in the range $x \in [\Phi(-7), \Phi(7)]$. For alternative algorithms, see for instance Acklam [2003] and Wichura [1988].

Well-known alternative methods for sampling in the Gaussian distribution include the *Box-Muller method* and the related *Marsaglia polar method* (see Press et al. [1992]).

3.1.1.2 Acceptance-Rejection Method

In cases where F^{-1} is cumbersome to compute, the so-called *acceptance-rejection method* may be preferable. To describe the method, suppose that we want to sample from a density $f(z) = dF(z)/dz$, and further suppose that we have a good method to sample from a density $e(z)$, where

$$e(z)c \geq f(z), \quad z \in \mathbb{R}, \quad (3.8)$$

for some positive constant c . By necessity, $c \geq 1$ as both e and f integrate to 1. In the acceptance-rejection method, we

1. Draw a sample Z from $e(z)$.
2. Draw an independent uniform variable U , $U \sim \mathcal{U}(0, 1)$.
3. Accept the sample Z if $U \leq f(Z)/(ce(Z))$; otherwise discard it.

⁴If even higher precision is required, we can use (3.6)–(3.7) as a guess for the root y in the equation $\Phi(y) = x$. Any number of numerical root search routines (e.g. Newton-Raphson) can then be applied to improve the precision of the solution further. Typically only one or two iterations will be required to get the solution to within machine precision on a PC.

The proof of why this algorithm works is straightforward and we omit it. Note that the third step of the acceptance-rejection method can be wasteful if too many samples need rejection. The key to the numerical efficiency of the acceptance-rejection method is thus evidently the ability to identify densities $e(z)$ that are “close” to $f(z)$, in the sense that c is close to 1 for all x . Indeed, it can easily be shown that the probability of rejecting a sample is $1/c$. Press et al. [1992] list good choices for $e(z)$ for a number of standard densities $f(z)$.

To demonstrate the mechanics of setting up an acceptance-rejection scheme for a particular distribution, let us consider sampling of a variable χ_ν^2 from a *chi-square distribution* with ν degrees of freedom. This distribution arises in a number of interest rate applications and is characterized by the cumulative distribution function

$$P(\chi_\nu^2 \leq z) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_0^z e^{-y/2} y^{(\nu/2)-1} dy, \quad \nu > 0, z \geq 0,$$

where Γ is the gamma function. For reasonably large degrees of freedom ν , the chi-square density is typically bell-shaped. The chi-square distribution is a special case of the *gamma distribution* with density

$$f(z; a, b) = \frac{a(az)^{b-1} e^{-az}}{\Gamma(b)}, \quad a, b > 0, z \geq 0. \quad (3.9)$$

The chi-square distribution corresponds to $a = \frac{1}{2}$ and $b = \frac{\nu}{2}$. Rather than considering how to simulate a chi-square distribution, we will consider the more general question of how to draw from (3.9). We note that if a variable X has gamma density $f(z; 1, b)$, then aX , $a > 0$, has gamma density $f(z; a, b)$, so, in fact, it suffices to consider a simulation algorithm for the unit-scale density

$$f(z) = \frac{z^{b-1} e^{-z}}{\Gamma(b)},$$

where we assume that $b \geq 1$. One simple choice of “comparison” density for an acceptance-rejection algorithm is the exponential density

$$e(z) = \lambda e^{-\lambda z},$$

which, as mentioned earlier, can easily be simulated by inverse transform techniques. Note that

$$\frac{f(z)}{e(z)} = \frac{1}{\lambda \Gamma(b)} z^{b-1} e^{(\lambda-1)z},$$

which can be checked to have a maximum value of

$$\sup \left(\frac{f(z)}{e(z)} \right) = \frac{1}{\lambda \Gamma(b)} \left(\frac{b-1}{e(1-\lambda)} \right)^{b-1}, \quad (3.10)$$

where we must assume that $\lambda < 1$. To satisfy (3.8) we take $c = \sup(f(z)/e(z))$ and now search for the value of λ that minimizes c , thereby optimizing computational speed. It is easy to see that (3.10) is minimized for $\lambda = 1/b$, corresponding to $c = b^b e^{1-b}/\Gamma(b)$. Note that

$$\frac{f(z)}{ce(z)} = \frac{z^{b-1}}{\lambda} e^{b-1+(\lambda-1)z} b^{-b},$$

with the third step of the acceptance-rejection algorithm best done in logarithms.

The algorithm outlined above was proposed by Fishman [1976] and works best for moderate values of b . For larger values, the Gamma distribution starts looking like a bell-shaped Gaussian distribution and is no longer well-approximated by an exponential distribution. Indeed, we notice that the probability of rejection ($1/c$) is approximately $e\sqrt{b/(2\pi)}$, so of order $O(\sqrt{b})$. Modifications to the basic Fishman algorithm to accelerate sampling can be found in Cheng and Feast [1980]. Another common idea is to set $e(z)$ to the *Cauchy density*

$$e(z) = \frac{1}{s\pi \left(1 + ((z-t)/s)^2 \right)},$$

where $s > 0$ and t are constants. This distribution is bell-shaped and, as discussed earlier, can be simulated by the inverse transform method. Press et al. [1992] list computer code and references for this case. For values $b \in [0, 1]$, the acceptance-rejection technique of Ahrens and Dieter [1974] can also be used.

3.1.1.3 Composition

A third and final method to generate random variables from a given distribution function exploits known functional relationships that map variables sampled from one or more distributions to variables sampled from a target distribution. This technique is known as *composition*. A classical example of composition is the *log-normal distribution* $\mathcal{LN}(\mu, \sigma^2)$ which, as we saw earlier in Chapter 1, is defined through the relation

$$X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow e^X \sim \mathcal{LN}(\mu, \sigma^2),$$

where \sim denotes “distributed as”, and where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2 . In other words, a sample Z from $\mathcal{LN}(\mu, \sigma^2)$ can be generated by drawing (by the inverse transformation method, say) a $\mathcal{N}(0, 1)$ variable X , and then setting $Z = e^{\mu+\sigma X}$.

Another classical example of a functional map is the *Student's t-distribution*, where samples can be generated by multiplying independent

samples from a standard Gaussian and a chi-square distribution; see Andersen et al. [2003] for a financial application of this. While we earlier demonstrated that the chi-square and gamma distributions can be generated by acceptance-rejection techniques, in fact we can also use composition for this. For instance, it is known that if X_1, X_2, \dots, X_ν are independent standard Gaussian variables, then

$$Z = \sum_{i=1}^{\nu} X_i^2 \quad (3.11)$$

is distributed chi-square with ν degrees of freedom. Also, if U_1, \dots, U_b are independent uniformly distributed variables, then

$$Z = -a \sum_{i=1}^b \ln U_i \quad (3.12)$$

is gamma distributed with density (3.9). For small integer-valued distribution parameters b or ν , (3.11) or (3.12) often define a faster simulation scheme than acceptance-rejection methods.

For later use, we note that the relationship (3.11) can be generalized to

$$\tilde{\chi}_{\nu}^2(\lambda) = \sum_{i=1}^{\nu} (X_i + a_i)^2$$

for a series of constants a_i , $i = 1, \dots, \nu$. The random variable $\tilde{\chi}_{\nu}^2(\lambda)$ follows a so-called *non-central chi-square distribution* with ν degrees of freedom and *non-centrality parameter* $\lambda = \sum_i a_i^2$. The distribution function is given by

$$P(\tilde{\chi}_{\nu}^2(\lambda) \leq z) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\lambda\right)^j}{j! \Gamma\left(\frac{\nu}{2} + j\right) 2^{(\nu/2)+j}} \int_0^z y^{\nu/2+j-1} e^{-y/2} dy, \quad (3.13)$$

an expression that also holds for non-integer ν . If $\nu > 1$, samples from a non-central chi-squared distribution can be generated by composition, using the relation

$$\tilde{\chi}_{\nu}^2(\lambda) = (Z + \sqrt{\lambda})^2 + \chi_{\nu-1}^2,$$

where Z is a standard Gaussian random variable independent of $\chi_{\nu-1}^2$. To handle the case $\nu \leq 1$, one can observe from the expression (3.13) that a non-central chi-square variable can be expressed as a regular chi-square variable $\chi_{\nu+2N}^2$, where N is an independent *Poisson-distributed* discrete variable with intensity $\lambda/2$,

$$P(N = j) = e^{-\lambda/2} \frac{(\lambda/2)^j}{j!}, \quad j = 0, 1, \dots$$

This suggests a composition rule for arbitrary ν : draw Poisson variables N (by the inverse transformation method, say) and then draw $\chi_{\nu+2N}^2$ using the methods in Section 3.1.1.2.

3.1.2 Correlated Gaussian Samples

The previous section dealt with the generation of scalar random variables. In applications, however, we may face the task of generating *vectors* of random variables, drawn from a joint multi-variate distribution. Of primary importance in financial applications is the multi-variate Gaussian distribution, so we devote this section to issues surrounding the generation of correlated Gaussian samples.

Recall that a p -dimensional Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ is characterized by a p -dimensional vector-valued mean μ and a $p \times p$ symmetric, positive semi-definite⁵ covariance matrix Σ . The joint density is

$$\phi_p(z; \mu, \Sigma) = \frac{1}{(2\pi)^{p/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(z - \mu)^\top \Sigma^{-1}(z - \mu)\right), \quad z \in \mathbb{R}^p.$$

The following result is useful:

Lemma 3.1.3 (Linear Transformation). *Let $Z \sim \mathcal{N}(\mu, \Sigma)$ be p -dimensional. Given a $d \times p$ matrix A and a d -dimensional vector B , then*

$$AZ + B \sim \mathcal{N}(A\mu + B, A\Sigma A^\top).$$

We can use this lemma as follows. Suppose that we generate p independent standard (that is, $\mathcal{N}(0, 1)$) Gaussian samples and collect them in a p -dimensional vector X . This can be accomplished using the techniques in Section 3.1.1. Clearly $X \sim \mathcal{N}(0, I)$, where I is the p -dimensional identity matrix. Define a $(p \times p)$ -dimensional matrix C satisfying

$$CC^\top = \Sigma. \tag{3.14}$$

Then

$$Z = \mu + CX$$

is distributed $\mathcal{N}(\mu, \Sigma)$.

It remains to determine a matrix C that satisfies (3.14). While there is generally an infinite number of such matrices, two particular choices are of primary importance. We discuss these below.

3.1.2.1 Cholesky Decomposition

In the Cholesky decomposition, we impose the constraint that the matrix C be lower triangular (that is, having all zeros above the diagonal), thereby conveniently reducing the number of multiplications required to compute CX to $p(1 + (p - 1)/2)$, rather than p^2 . Assuming that the matrix is positive definite (not only positive semi-definite), the Cholesky decomposition is well-defined, and given by

⁵That is, all eigenvalues of Σ are non-negative.

$$C_{i,i} = \sqrt{\Sigma_{i,i} - \sum_{k=1}^{i-1} C_{i,k}^2}, \quad i = 1, \dots, p,$$

$$C_{i,j} = \frac{1}{C_{j,j}} \left(\Sigma_{i,j} - \sum_{k=1}^{j-1} C_{i,k} C_{j,k} \right), \quad j = 1, \dots, p-1, \quad j < i.$$

For instance, if

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $\rho \in [-1, 1]$ and $\sigma_1, \sigma_2 > 0$, then

$$C = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix},$$

a result that we have already used in Section 2.11. Press et al. [1992], among others, list computer code implementing the relations above.

If the matrix Σ is only positive semi-definite (but not positive definite), the Cholesky decomposition will fail. In this case, linear algebra tells us that the matrix Σ is rank-deficient, with rank $r < p$. As such, we must be able to set $Z = \mu + MY$, where M is a $p \times r$ matrix and $Y \sim \mathcal{N}(0, \Sigma_Y)$ is r -dimensional, with the covariance matrix having full rank r . Using Cholesky composition instead on Σ_Y , we can find a lower diagonal matrix C_Y satisfying $C_Y C_Y^\top = \Sigma_Y$. Thus, in this case

$$Z = \mu + MC_Y X$$

where X is a vector of r (not p) independent standard Gaussian samples. The matrix M can be found by the singular value decomposition (SVD) algorithm, see Press et al. [1992], or the algorithm in the next section.

3.1.2.2 Eigenvalue Decomposition

As an alternative to Cholesky decomposition, we can also consider diagonalizing Σ through an eigenvalue decomposition. Here, we write

$$\Sigma = E\Lambda E^\top, \tag{3.15}$$

where Λ is a diagonal matrix of eigenvalues λ_i , $i = 1, \dots, p$, and the columns of E contain the orthonormal eigenvectors of Σ . Some eigenvalues may be zero, if Σ is rank-deficient (positive semi-definite). Comparison with (3.14) implies that one choice of C is

$$C = E\sqrt{\Lambda} = E \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{\lambda_p} \end{pmatrix}. \tag{3.16}$$

The eigenvalue decomposition (3.15) is relatively straightforward, at least as eigenvalue problems go, due to the fact that Σ is symmetric and positive semi-definite; see Press et al. [1992] for an algorithm. While both Cholesky decomposition and eigenvalue decompositions have computational complexity $O(n^3)$, in practice the Cholesky method is often much faster than the eigenvalue method, making the Cholesky method preferable in practice. Nevertheless, decompositions of the type (3.16) have certain appealing theoretical properties that shall be useful later, so the next section explores (3.16) further.

3.1.3 Principal Components Analysis (PCA)

Consider a p -dimensional Gaussian variable Z with a given covariance matrix Σ . Assume, with no loss of generality, that the mean of Z is 0 and that Σ has full rank (positive definite). Consider now writing, as an approximation,

$$Z \approx DX, \quad (3.17)$$

where X is an r -dimensional vector of independent standard Gaussian variables, $r \leq p$, and D is a $(p \times r)$ -dimensional matrix. How should we choose D in an optimal way?

First, we obviously need to define what constitutes an “optimal” approximation in (3.17). We here have in mind L^2 closeness of the covariance matrix DD^\top to Σ (see Lemma 3.1.3), so let us define the optimal D^* as the matrix that minimizes the norm

$$f(D) = \text{tr} \left((\Sigma - DD^\top) (\Sigma - DD^\top)^\top \right).$$

This is just the matrix representation of the usual Frobenius norm on the squared differences between Σ and DD^\top . The value of D that minimizes $f(D)$ can be shown to be

$$D^* = E_r \sqrt{\Lambda_r}, \quad (3.18)$$

where Λ_r is an $r \times r$ diagonal matrix containing the *largest* r eigenvalues of Σ , and E_r is a $p \times r$ matrix of r p -dimensional eigenvectors corresponding to the eigenvalues in Λ_r .

Equipped with the optimal D , we now go back to the approximation (3.17) and write

$$Z \approx \tilde{Z} \triangleq E_r \sqrt{\Lambda_r} X = \sqrt{\lambda_1} e_1 X_1 + \sqrt{\lambda_2} e_2 X_2 + \dots + \sqrt{\lambda_r} e_r X_r, \quad (3.19)$$

where e_i denotes the i -th column of E_r and the λ_i 's are the eigenvalues, sorted in decreasing order of magnitude. The (deterministic) vector e_i is known as the i -th *principal component* of Z , and the (random) variable $\sqrt{\lambda_i} X_i$ as the i -th *principal factor*. With (3.19), we have $\text{tr}(\text{Cov}(Z, Z)) = E(Z^\top Z) = \sum_{i=1}^p \lambda_i$ and $\text{tr}(\text{Cov}(\tilde{Z}, \tilde{Z})) = E(\tilde{Z}^\top \tilde{Z}) = \sum_{i=1}^r \lambda_i$, i.e. the first r terms in the decomposition (3.19) explain a fraction

$$\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^p \lambda_i}$$

of the sum of the diagonal elements of the covariance matrix of Z . Principal components decomposition will thus result in a loss of total variance, unless the covariance matrix is either rank-deficient (i.e. has eigenvalues that are strictly zero), or we use a full set of principal components ($p = r$). In many cases of interest to us here, the loss of variance can be small, even if r is a modest number, e.g. 2 or 3. We notice that the covariance matrix for Z , as approximated by (3.19), will be *rank-deficient*, as the number r of non-zero eigenvalues is less than p .

While we have used a setting with Gaussian variables to motivate our treatment of principal components analysis (PCA), it is, in fact, a generically useful tool for uncovering the structure of large-dimensional random vectors, and replacing them with more manageable, lower-dimensional variables; see, e.g., Theil [1971] for more details and an application to empirical non-Gaussian data. Also, PCA identifies which directions of a multi-dimensional random variable are “important”, potentially allowing us to allocate computational resources in an intelligent manner. One example of this is shown later in this chapter, in Section 3.2.10.

3.2 Generation of Sample Paths

So far, we have assumed that random variables are characterized by a known distribution function. In most of our applications, however, the random variables $g(T)/N(T)$ used in the basic pricing equation (3.1) are specified through an SDE or, more generally, an Ito process. In this section, we shall discuss Monte Carlo simulation of such processes. We start out with a motivating example, set in the Black-Scholes-Merton economy.

3.2.1 Example: Asian Basket Options in Black-Scholes Economy

Consider a dividend-free stock S , with Black-Scholes dynamics

$$dS(t)/S(t) = r dt + \sigma dW(t), \quad (3.20)$$

where $W(t)$ is a Brownian motion in the risk-neutral measure \mathbb{Q} , r is a constant interest rate, and σ is a constant volatility. Let there be given an increasing set of observation times $\{t_1, t_2, \dots, t_m\}$, with $t_m = T$, and define the \mathcal{F}_T -measurable (discretely observed) stock average

$$A(T) = \frac{1}{m} \sum_{i=1}^m S(t_i). \quad (3.21)$$

An *Asian* (or *average rate*) call option with strike K is defined by the terminal payout

$$g(T) = (A(T) - K)^+; \quad (3.22)$$

we wish to price this option by Monte Carlo simulation.

As discussed earlier (see (1.39)), the geometric Brownian motion process (3.20) allows us to express S directly in terms of the Brownian motion,

$$S(t) = S(0)e^{rt - \frac{1}{2}\sigma^2 t + \sigma W(t)}, \quad t > 0,$$

whereby, with $\Delta_i \triangleq t_i - t_{i-1}$ and $t_0 = 0$,

$$S(t_i) = S(t_{i-1}) \exp \left(\left[r - \frac{1}{2}\sigma^2 \right] \Delta_i + \sigma [W(t_i) - W(t_{i-1})] \right),$$

$i = 1, \dots, m$. By the properties of Brownian motion, the increments $W(t_i) - W(t_{i-1})$ are independent Gaussian variables distributed as $\mathcal{N}(0, \Delta_i)$. For the purposes of Monte Carlo simulation, we can therefore write

$$S(t_i) = S(t_{i-1}) \exp \left(\left(r - \frac{1}{2}\sigma^2 \right) \Delta_i \right) \exp \left(\sigma \sqrt{\Delta_i} Z_i \right), \quad i = 1, \dots, m, \quad (3.23)$$

where the Z_i are independent standard $\mathcal{N}(0, 1)$ Gaussian random variables. To produce a single sample draw of $g(T)$, we thus

1. Draw independent standard Gaussian samples $Z_i, i = 1, \dots, m$ (see Section 3.1.1).
2. Starting from $S(0)$, generate $S(t_i), i = 1, \dots, m$, from the iteration (3.23).
3. Compute $g(T)$ from (3.21)–(3.22).

Repeating this procedure n times (with Gaussian samples independent from one path to the next), we can generate n random samples g_1, g_2, \dots, g_n of $g(T)$. Our estimate of the time 0 price of the Asian option is then, from (3.3) with $N(t) = e^{rt}$ and non-random,

$$\bar{V}(0) = e^{-rT} \frac{1}{n} \sum_{j=1}^n g_j.$$

Asymptotic confidence intervals can be computed from (3.4). The pricing algorithm involves drawing mn Gaussian variables, so the computational cost of the pricing algorithm is $O(mn)$.

Increasing the complexity, let us now consider an Asian option on a p -dimensional basket of stocks S_1, S_2, \dots, S_p , each following geometric Brownian motion,

$$dS_k(t)/S_k(t) = r dt + \sigma_k dW_k(t), \quad k = 1, \dots, p.$$

The Brownian motions W_k and W_j are assumed correlated with constant correlation coefficient $\rho_{k,j}$, $j, k = 1, \dots, p$, $j \neq k$. Define a unit-weighted basket price as

$$B(t) = \sum_{k=1}^p S_k(t),$$

and set the terminal Asian option payout to be

$$g(T) = \left(\frac{1}{m} \sum_{i=1}^m B(t_i) - K \right)^+, \quad (3.24)$$

where the time line $\{t_i\}$ is as before. Equivalent to (3.23), we draw sample paths for each asset according to the prescription

$$S_k(t_i) = S_k(t_{i-1}) \exp \left(\left(r - \frac{1}{2} \sigma_k^2 \right) \Delta_i + \sigma_k \sqrt{\Delta_i} Z_{k,i} \right), \quad (3.25)$$

$$i = 1, \dots, m, \quad k = 1, \dots, p,$$

where the $Z_{k,i}$ are Gaussian samples, independently drawn at each time step but correlated across k 's. Let C be the Cholesky decomposition of the correlation matrix $\{\rho_{k,j}\}$ (see Section 3.1.2.1), in which case we can generate the correlated sample vectors $Z_i = (Z_{1,i}, Z_{2,i}, \dots, Z_{p,i})^\top$ as

$$Z_i = CX_i$$

for a p -dimensional vector X_i of independent Gaussian samples. Given joint sample paths of all basket component assets S_k , $k = 1, \dots, p$, pricing of the Asian basket option proceeds as above, substituting (3.24) for (3.22).

Completion of (3.25) requires pm samples to complete a full path of all p assets, making the total computational effort of an n -sample Monte Carlo scheme $O(nmp)$, with the (probabilistic) convergence order $O(n^{-1/2})$ and dependent only on n . As mentioned earlier, the linearity of computational cost on the dimension of the asset vector p compares favorably to the exponential growth in p of finite difference schemes. Notice also the ease with which the Monte Carlo scheme is able to incorporate path-dependence.

3.2.2 Discretization Schemes, Convergence, and Stability

At the heart of the example in Section 3.2.1 was an iterative scheme for the production of a sample path for a vector-valued SDE; see (3.25). For the simple Black-Scholes model, SDE state variables (stock prices) could be expressed analytically in terms of independent increments of a Brownian motion, making path generation straightforward. In practice, however, we are often working with SDEs that do not permit closed-form solution. In such cases, we need to *time-discretize* the SDE, much the same way as we did for the numerical solution of PDEs.

In the next few sections, we shall consider a few important SDE discretization schemes. Before moving on to this, it is useful to discuss the sense in which we consider a discretization scheme to converge to the true SDE solution. For this, consider a vector-valued SDE

$$dX(t) = \mu(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad (3.26)$$

where $X(t)$ is p -dimensional, W is a d -dimensional vector of independent Brownian motions, and $\mu : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ and $\sigma : [0, T] \times \mathbb{R}^p \rightarrow \mathbb{R}^{p \times d}$ satisfy the usual regularity conditions. Consider an equidistant⁶ time grid $\{0, \Delta, 2\Delta, \dots, m\Delta\}$, the number of references and let \widehat{X} be an approximation to X , based on some kind of time-discretization scheme on the grid $\{i\Delta\}$. For simplicity of notation, set $\widehat{X}_i \triangleq \widehat{X}(i\Delta)$. We say that the underlying approximation is *weakly consistent* if there exists a function $c(\Delta)$ with

$$\lim_{\Delta \downarrow 0} c(\Delta) = 0$$

such that (dropping the measure superscript on the expectation operator)

$$\mathbb{E} \left(\left| \mathbb{E} \left(\Delta^{-1} (\widehat{X}_{i+1} - \widehat{X}_i) \middle| \mathcal{F}_{i\Delta} \right) - \mu(i\Delta, \widehat{X}_i) \right|^2 \right) \leq c(\Delta), \quad (3.27)$$

and

$$\begin{aligned} \mathbb{E} \left(\left| \mathbb{E} \left(\Delta^{-1} (\widehat{X}_{i+1} - \widehat{X}_i) (\widehat{X}_{i+1} - \widehat{X}_i)^\top \middle| \mathcal{F}_{i\Delta} \right) \right. \right. \\ \left. \left. - \sigma(i\Delta, \widehat{X}_i) \sigma(i\Delta, \widehat{X}_i)^\top \right|^2 \right) \leq c(\Delta), \end{aligned} \quad (3.28)$$

for all $i = 0, \dots, m-1$. The notion of weak consistency⁷ thus amounts to requiring that the mean and variance of the increments of the approximating process be close to those of the true SDE solution.

A concept related to consistency is the notion of *weak convergence*. We say that an approximate solution converges weakly to X at time $T = m\Delta$ with respect to a class \mathcal{C} of test functions $g : \mathbb{R}^p \rightarrow \mathbb{R}$ if

$$\lim_{\Delta \downarrow 0} \left| \mathbb{E}(g(X(T))) - \mathbb{E}(g(\widehat{X}(T))) \right| = 0, \quad (3.29)$$

for all $g \in \mathcal{C}$. Notice that the limit necessarily involves $m \rightarrow \infty$.

⁶To keep notation manageable, we use a constant time step Δ in most of this chapter. All results are, however, easily extendable to non-equidistant grids.

⁷*Strong consistency* (which is of little use to us in this book) requires that (3.27) is satisfied, and that the variance of the difference between increments of the true process and the approximation vanish. The second requirement is stronger than (3.28).

The class of test functions used in (3.29) is normally always in the set \mathcal{C}_P^l of functions with polynomially bounded⁸ derivatives of order $0, 1, \dots, l$ with maximum power l . We say that a scheme converges with *weak order* β if, for all $g \in C_P^{2(\beta+1)}$, (3.29) can be strengthened to

$$\left| \mathbb{E}(g(X(T))) - \mathbb{E}\left(g(\hat{X}(T))\right) \right| \leq c\Delta^\beta, \quad (3.30)$$

for all $\Delta \in (0, \Delta_0)$, where Δ_0 and c are constants and c does not depend on Δ (but may depend on g).

One would generally expect that a weakly consistent scheme is weakly convergent. Indeed, this can be established to be the case under certain additional regularity conditions. We will not list the exact result here, but refer to Kloeden and Platen [2000], Theorem 9.7.4.

Finally, a brief word on stability of a time-discretized SDE. A commonly used definition of *A*-stability focuses on the behavior of a discretized test SDE of the type

$$dX(t) = \lambda X(t) dt + dW(t), \quad (3.31)$$

where λ is a complex-valued constant with real part $\text{Re}(\lambda) < 0$. We suppose that a discretization scheme can be represented as

$$\hat{X}_{i+1} = \hat{X}_i G(\lambda\Delta) + Z_i^\Delta, \quad i = 0, 1, \dots, m-1, \quad (3.32)$$

where G is a mapping of the complex plane onto itself and the Z_i^Δ 's are random variables independent of the \hat{X}_i 's. In this case, the *region of stability* for a scheme is the set of $\lambda\Delta$ for which $\text{Re}(\lambda) < 0$ and

$$|G(\lambda\Delta)| < 1. \quad (3.33)$$

Similar to the definition used for finite difference scheme discretizations, we say that an SDE time-discretization scheme is *A-stable*, if the region of stability includes all values of λ with $\text{Re}(\lambda) < 0$ and all $\Delta > 0$.

3.2.3 The Euler Scheme

An obvious first scheme to discretize (3.26) treats both dt and $dW(t)$ fully explicitly, evaluating all SDE coefficients on time step $[i\Delta, i\Delta + \Delta]$ at the left interval point $i\Delta$. In other words, we write, starting from $\hat{X}_0 = X(0)$,

$$\begin{aligned} \hat{X}_{i+1} &= \hat{X}_i + \mu(i\Delta, \hat{X}_i) \Delta + \sigma(i\Delta, \hat{X}_i) (W(i\Delta + \Delta) - W(i\Delta)), \\ i &= 0, 1, \dots, m-1. \end{aligned} \quad (3.34)$$

⁸A function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ is polynomially bounded if $|f(x)| \leq k(1 + |x|^q)$, $x \in \mathbb{R}^p$, for constants k and q .

With this scheme, Monte Carlo generation of paths is straightforward and involves, as in Section 3.2.1, replacing the increments $W(i\Delta + \Delta) - W(i\Delta)$ with $Z_i\sqrt{\Delta}$, for a d -dimensional vector of independent standard Gaussian samples Z_i .

The discretization scheme (3.34) is known as the *Euler scheme*, sometimes also called the *Euler-Maruyama* scheme. The Euler scheme is easy to implement and is a true workhorse that we will often use in this book. We note that the scheme is weakly consistent, as

$$\mathbb{E} \left(\left| \mathbb{E} \left(\Delta^{-1} (\widehat{X}_{i+1} - \widehat{X}_i) \mid \mathcal{F}_{i\Delta} \right) - \mu(i\Delta, \widehat{X}_i) \right|^2 \right) = 0,$$

and

$$\begin{aligned} \mathbb{E} \left(\left| \mathbb{E} \left(\Delta^{-1} (\widehat{X}_{i+1} - \widehat{X}_i) (\widehat{X}_{i+1} - \widehat{X}_i)^\top \mid \mathcal{F}_{i\Delta} \right) \right. \right. \\ \left. \left. - \sigma(i\Delta, \widehat{X}_i) \sigma(i\Delta, \widehat{X}_i)^\top \right|^2 \right) = O(\Delta^2). \end{aligned}$$

While one might believe that the explicit discretization of the diffusion term — which is only accurate to order $O(\sqrt{\Delta})$ — would give the scheme weak convergence order⁹ $1/2$, in fact we typically have that the Euler scheme has *weak convergence order* $\beta = 1$. We note that for this result to hold, however, regularity conditions on μ and σ stronger than those of the existence and uniqueness results (Theorem 1.6.1) are needed. For instance, in the case where μ and σ are functions of X alone, Theorem 9.7.6 in Kloeden and Platen [2000] requires that μ and σ be four times continuously differentiable with polynomial growth and uniformly bounded derivatives. See also their Theorem 15.4.2 for a more general result.

Given that the Euler scheme is fully explicit, our experience from finite difference methods suggests that the scheme may have stability problems. To investigate, we follow Section 3.2.2 and consider the test SDE

$$dX(t) = \lambda X(t)dt + dW(t),$$

which is discretized as

$$\widehat{X}_{i+1} = \widehat{X}_i (1 + \lambda\Delta) + \sqrt{\Delta} Z_i, \quad (3.35)$$

where Z_i 's are standard Gaussian. Comparison to (3.32) and (3.33) shows that the region of stability for the Euler scheme is

$$|(1 + \lambda\Delta)| < 1, \quad \text{Re}(\lambda) < 0,$$

which is the unit disc in the complex plane centered at $\lambda\Delta = -1$. For a given λ , there are thus restrictions on how big a time step Δ can be used.

⁹The so-called *strong convergence order* of the Euler scheme is in fact only $1/2$. The concept of strong convergence order is defined in Kloeden and Platen [2000] and is of little importance to applications in this book.

3.2.3.1 Linear-Drift SDEs

The restricted stability region of the Euler scheme can be a practical concern. For instance, SDEs of the important type

$$dX(t) = \kappa(\theta(t) - X(t)) dt + \sigma(t, X(t)) dW(t) \quad (3.36)$$

arise quite frequently in fixed income modeling, and in cases where κ is big (which is often the case for, say, stochastic volatility models such as those covered in Chapters 8, 9 and 13) the Euler scheme can become unstable and return meaningless results. One way to solve the problem is to switch to an implicit scheme (see next section), but in the case (3.36) we can use the fact that the drift term can be removed by a simple change of variable. For instance, for the case where $X(t)$ is scalar we can set

$$Y(t) = e^{\kappa t} X(t) - \kappa \int_0^t e^{\kappa u} \theta(u) du,$$

such that, from Ito's lemma,

$$\begin{aligned} dY(t) &= e^{\kappa t} \sigma(t, X(t)) dW(t) \\ &= e^{\kappa t} \sigma \left(t, e^{-\kappa t} \left(Y(t) + \kappa \int_0^t e^{\kappa u} \theta(u) du \right) \right) dW(t). \end{aligned}$$

Euler simulation of the process for $Y(t)$, rather than for $X(t)$, will center X around its analytically known mean

$$\mathbb{E}(X((i+1)\Delta) | X(i\Delta)) = e^{-\kappa\Delta} X(i\Delta) + \kappa \int_{i\Delta}^{i\Delta+\Delta} e^{-\kappa((i+1)\Delta-u)} \theta(u) du$$

and will often alleviate any stability problems.

3.2.3.2 Log-Euler Scheme

One potential problem with the pure Euler scheme (3.34) is the fact that all increments are locally Gaussian, thereby implying a non-zero probability of \hat{X} crossing zero and becoming negative. Many SDEs, however, are known to produce only non-negative solutions, and the functions μ and σ may not allow for negative arguments. This, for instance, is the case for the square-root process

$$dX(t) = \sqrt{X(t)} dW(t), \quad X(0) > 0,$$

where the Euler scheme cannot be directly applied. Some authors (e.g., Kloeden and Platen [2000]) suggest heuristic modifications of the Euler scheme, such as

$$\widehat{X}_{i+1} = \widehat{X}_i + \sqrt{|\widehat{X}_i|} (W(i\Delta + \Delta) - W(i\Delta)),$$

but ultimately this is not very satisfying and the resulting scheme will often have large errors¹⁰. An alternative is to introduce an invertible transformation $X(t) = f(Y(t))$, with $f : \mathbb{R} \rightarrow \mathbb{R}_+$, and then apply the Euler scheme to Y , at each step recovering X as $f(Y)$. In finance applications, where many processes are based on SDEs that bear some resemblance to geometric Brownian motion, an often-used choice for f is $f(y) = e^y$. The resulting scheme is known as the *log-Euler scheme*.

Consider the SDE (3.26) and assume for simplicity that X is scalar (if X is vector valued, the log-transform can be applied to all, or a few selected, components of X). Set $X(t) = \exp(Y(t))$, such that $Y(t) = \ln(X(t))$. The process for Y then follows from Ito's lemma:

$$dY(t) = \left(\frac{\mu(t, X(t))}{X(t)} - \frac{1}{2} \frac{\sigma(t, X(t))^2}{X(t)^2} \right) dt + \frac{\sigma(t, X(t))}{X(t)} dW(t), \quad X(t) = e^{Y(t)}.$$

Writing out a standard Euler scheme for Y and making the transformation $\widehat{X}_i = \exp(\widehat{Y}_i)$ gives us the (scalar) log-Euler scheme for X :

$$\widehat{X}_{i+1} = \widehat{X}_i \exp \left(\left(\frac{\mu(t, \widehat{X}_i)}{\widehat{X}_i} - \frac{1}{2} \frac{\sigma(t, \widehat{X}_i)^2}{\widehat{X}_i^2} \right) \Delta + \frac{\sigma(t, \widehat{X}_i)}{\widehat{X}_i} Z_i \sqrt{\Delta} \right),$$

where $Z_i \sim \mathcal{N}(0, 1)$. Generalizations of the technique above to situations where the valid range of X is some general set \mathcal{C} are obvious and involve identifying an invertible mapping function $f : \mathbb{R} \rightarrow \mathcal{C}$, preferably one that can be inverted analytically. For instance, if $\mathcal{C} = [a, \infty)$, we could use $f(y) = a + e^y$.

3.2.4 The Implicit Euler Scheme

The implicit Euler scheme for the vector-valued SDE (3.26) takes the form

$$\widehat{X}_{i+1} = \widehat{X}_i + \mu(i\Delta + \Delta, \widehat{X}_{i+1}) \Delta + \sigma(i\Delta, \widehat{X}_i) (W(i\Delta + \Delta) - W(i\Delta)), \quad (3.37)$$

for $i = 0, 1, \dots, m - 1$. We highlight the fact that the drift coefficient μ is now evaluated at time $i\Delta + \Delta$, rather than at time $i\Delta$. It is easy to show that the implicit Euler scheme is consistent. Under regularity conditions, it can also be shown that the weak convergence order is $\beta = 1$, just as was the case for the explicit Euler scheme.

The main advantage of the implicit Euler scheme over the explicit Euler scheme is numerical stability. To examine the region of stability for the implicit Euler scheme, consider again the test SDE

¹⁰For a dedicated treatment of the rather delicate problem of simulating square-root process, see Chapter 9.

$$dX(t) = \lambda X(t) dt + dW(t).$$

It will now be discretized as (compare to (3.35))

$$\widehat{X}_{i+1} = \widehat{X}_i + \widehat{X}_{i+1} \lambda \Delta + \sqrt{\Delta} Z_i,$$

or

$$\widehat{X}_{i+1} (1 - \lambda \Delta) = \widehat{X}_i + \sqrt{\Delta} Z_i.$$

Comparison to (3.32) and (3.33) shows that now

$$G(\lambda \Delta) = \frac{1}{1 - \lambda \Delta}$$

such that the stability criterion $|G(\lambda \Delta)| < 1$ is satisfied for any value of $\lambda \Delta$ where $\text{Re}(\lambda) < 0$. In other words, the implicit scheme is A -stable.

3.2.4.1 Implicit Diffusion Term

The reader may at this point wonder why the implicit scheme (3.37) only discretized the drift term (μ) implicitly, and not the diffusion term (σ). The answer lies in the differences between a regular Riemann integral and the stochastic integral. Recall in particular that the stochastic Ito integral is defined to be non-anticipative, in the sense that the integrand is always evaluated “to the left” on any partitions of the Brownian motion. As a consequence, if $\sigma(i\Delta, \widehat{X}_i)$ were replaced with $\sigma(i\Delta + \Delta, \widehat{X}_{i+1})$ in (3.37), the resulting scheme would not be weakly consistent, in the sense defined earlier. To illustrate this point, just consider the simple scalar process

$$dX(t) = \sigma X(t) dW(t),$$

which we contemplate discretizing as

$$\widehat{X}_{i+1} = \widehat{X}_i + \sigma \widehat{X}_{i+1} (W(i\Delta + \Delta) - W(i\Delta)),$$

or

$$\widehat{X}_{i+1} (1 - \sigma Z_i \sqrt{\Delta}) = \widehat{X}_i, \quad i = 0, \dots, m-1. \quad (3.38)$$

Here, a first difficulty arises: the term $(1 - \sigma Z_i \sqrt{\Delta})$ may become 0 (or very close to zero) if Z_i is an (unbounded) Gaussian variable. For fully implicit discretization schemes, it becomes necessary to use a bounded approximation to the Brownian motion. As discussed in Kloeden and Platen [2000], weak convergence order is preserved if in (3.38) we set the Z_i to be independent binomial variables with

$$P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}.$$

We assume that $1 - \sigma \sqrt{\Delta} > 0$. Rearranging and Taylor-expanding, we get

$$\begin{aligned}
\frac{\widehat{X}_{i+1} - \widehat{X}_i}{\Delta} &= \frac{\widehat{X}_i}{\Delta} \left(\frac{1}{1 - \sigma Z_i \sqrt{\Delta}} - 1 \right) \\
&= \frac{\widehat{X}_i}{\Delta} \left(1 + \sigma Z_i \sqrt{\Delta} + \sigma^2 Z_i^2 \Delta + O(\sigma^3 Z_i^3 \Delta^{3/2}) - 1 \right) \\
&= \widehat{X}_i \left(\sigma Z_i \Delta^{-1/2} + \sigma^2 Z_i^2 + O(\sigma^3 Z_i^3 \Delta^{1/2}) \right)
\end{aligned}$$

such that

$$\mathbb{E} \left(\frac{\widehat{X}_{i+1} - \widehat{X}_i}{\Delta} \middle| \widehat{X}_i \right) = \widehat{X}_i \left(\sigma^2 + O(\Delta^{1/2}) \right).$$

Clearly, this will cause a violation of the consistency condition (3.27).

In the example above, we notice that consistency can be restored if the drift of the original SDE is changed from 0 to $-\sigma^2 X(t)$ before the “doubly” implicit Euler discretization is employed. More generally, it is not difficult to show that (3.37) can be modified to treat the diffusion term implicitly, provided that the drift of the original vector-valued SDE (3.26) is first changed from μ to

$$\bar{\mu} = \mu - \sum_{j=1}^d \sum_{k=1}^p (\sigma_{X_k})_{\cdot,j} \sigma_{k,j}$$

where the p -dimensional vector $(\sigma_{X_k})_{\cdot,j}$ is the j -th column of the $(p \times d)$ -dimensional matrix $\sigma_{X_k} = \{\partial \sigma_{i,j} / \partial X_k\}$. Inspired by the theta methods of Chapter 2, we can, in fact, introduce a family of discretizations

$$\begin{aligned}
\widehat{X}_{i+1} &= \widehat{X}_i + \left[(1 - \theta) \bar{\mu}_\eta \left(i\Delta, \widehat{X}_i \right) + \theta \bar{\mu}_\eta \left(i\Delta + \Delta, \widehat{X}_{i+1} \right) \right] \Delta \\
&\quad + \left[(1 - \eta) \sigma \left(i\Delta, \widehat{X}_i \right) + \eta \sigma \left(i\Delta + \Delta, \widehat{X}_{i+1} \right) \right] Z_i \sqrt{\Delta}, \quad (3.39)
\end{aligned}$$

where the Z_i are binomially distributed variables, $\theta, \eta \in [0, 1]$ are parameters, and

$$\bar{\mu}_\eta = \mu - \eta \sum_{j=1}^d \sum_{k=1}^p (\sigma_{X_k})_{\cdot,j} \sigma_{k,j}. \quad (3.40)$$

As it turns out, all these schemes theoretically have identical convergence order $\beta = 1$, but in practice some choices of θ, η may turn out to work better than others. We shall discuss methods to raise the theoretical convergence order in Section 3.2.6. The scheme (3.39) can be verified to be A -stable for $\theta \in [1/2, 1]$.

3.2.5 Predictor-Corrector Schemes

A closer examination of the implicit Euler scheme (3.37) demonstrates the need to recover $\widehat{X}(i\Delta + \Delta)$ as the vector-valued root of a possibly non-linear equation. In general, this must be done numerically (using, say, the

Newton-Raphson algorithm), causing a severe deterioration of computational performance. An alternative is to use the explicit Euler scheme as a *predictor* and the implicit scheme as a *corrector*, much the same way we used explicit finite difference approximations as predictors in the Craig-Sneyd algorithm of Section 2.11. Moving straight to the general implicit discretization family (3.39), we write the predictor-corrector as

$$\bar{X}_{i+1} = \hat{X}_i + \mu(i\Delta, \hat{X}_i) \Delta + \sigma(i\Delta, \hat{X}_i) (W(i\Delta + \Delta) - W(i\Delta)), \quad (3.41)$$

$$\begin{aligned} \hat{X}_{i+1} &= \hat{X}_i + [(1-\theta)\bar{\mu}_\eta(i\Delta, \hat{X}_i) + \theta\bar{\mu}_\eta(i\Delta + \Delta, \bar{X}_{i+1})] \Delta \\ &\quad + [(1-\eta)\sigma(i\Delta, \hat{X}_i) + \eta\sigma(i\Delta + \Delta, \bar{X}_{i+1})] (W(i\Delta + \Delta) - W(i\Delta)), \end{aligned} \quad (3.42)$$

where $\theta, \eta \in [0, 1]$, and $\bar{\mu}_\eta$ is as given in (3.40). It is understood that the Brownian motion increments in (3.41) and (3.42) are to be identical.

For sufficiently smooth coefficients, it can be shown that the predictor-corrector scheme (3.41)–(3.42) converges weakly with order $\beta = 1$, independent of the choice of θ and η . As for stability, discretization of (3.31) leads to

$$\begin{aligned} \bar{X}_{i+1} &= \hat{X}_i (1 + \lambda\Delta) + W(i\Delta + \Delta) - W(i\Delta), \\ \hat{X}_{i+1} &= \hat{X}_i + [(1-\theta)\lambda\hat{X}_i + \theta\lambda\bar{X}_{i+1}] \Delta + W(i\Delta + \Delta) - W(i\Delta) \\ &= \hat{X}_i (1 + \lambda\Delta(1 + \theta\lambda\Delta)) + (W(i\Delta + \Delta) - W(i\Delta))(1 + \theta\lambda\Delta). \end{aligned}$$

The region of stability can be verified to be

$$|1 + \lambda\Delta(1 + \theta\lambda\Delta)| < 1, \quad \text{Re}(\lambda) < 0.$$

For $\theta = \frac{1}{2}$, the stability criterion above is identical to that of the classical *Heun scheme* (or *modified trapezoidal scheme*) used for ordinary differential equations. Indeed, the predictor-corrector scheme above can be seen as an adaptation of this scheme for SDEs. We note that SDE adaptations of more sophisticated ODE solvers (such as Runge-Kutta) are also possible, but this goes beyond the scope of this text.

3.2.6 Ito-Taylor Expansions and Higher-Order Schemes

Despite our various efforts at centering derivatives, none of the schemes listed above theoretically attain second-order weak convergence. To develop such schemes, we need to delve further into adapting classical Taylor expansions to the rules of stochastic (Ito) calculus. As we shall ultimately not have much use for higher-order schemes, we keep the treatment informal and limit ourselves to the scalar case where $p = d = 1$ in (3.26).

3.2.6.1 Ordinary Taylor Expansion of ODEs

To gain intuition, start by setting $\sigma = 0$ in (3.26), such that we first deal with an ordinary ODE

$$dX(t) = \mu(t, X(t)) dt. \quad (3.43)$$

For a given value of t , we can use Taylor's theorem to write

$$X(t + \Delta) = X(t) + \frac{dX(t)}{dt} \Delta + \frac{1}{2} \frac{d^2 X(t)}{dt^2} \Delta^2 + O(\Delta^3),$$

where we stop at order $O(\Delta^3)$. We notice that

$$\frac{dX(t)}{dt} = \mu(t, X(t))$$

and

$$\begin{aligned} \frac{d^2 X(t)}{dt^2} &= \frac{\partial}{\partial t} \mu(t, X(t)) + \frac{\partial}{\partial x} \mu(t, X(t)) \cdot \frac{dX(t)}{dt} \\ &= \left(\frac{\partial}{\partial t} + \mu(t, X(t)) \frac{\partial}{\partial x} \right) \mu(t, X(t)). \end{aligned}$$

Setting

$$\mathcal{L} \triangleq \frac{\partial}{\partial t} + \mu \frac{\partial}{\partial x},$$

we thus have

$$X(t + \Delta) = X(t) + \mu(t, X(t)) \Delta + \frac{1}{2} \mathcal{L} \mu(t, X(t)) \Delta^2 + O(\Delta^3). \quad (3.44)$$

Another way to develop (3.44) proceeds by iteration on the integral representation

$$X(t + \Delta) = X(t) + \int_t^{t+\Delta} \mu(u, X(u)) du. \quad (3.45)$$

First we recognize that (as seen above)

$$d\mu(t, X(t)) = \mathcal{L} \mu(t, X(t)) dt$$

such that

$$\mu(u, X(u)) = \mu(t, X(t)) + \int_t^u \mathcal{L} \mu(s, X(s)) ds, \quad u > t. \quad (3.46)$$

Inserting this into (3.45) gives

$$X(t + \Delta) = X(t) + \mu(t, X(t)) \int_t^{t+\Delta} du + \int_t^{t+\Delta} \int_t^u \mathcal{L} \mu(s, X(s)) ds du.$$

Applied to $\mathcal{L}\mu(s, X(s))$ the steps that lead to (3.46) yield

$$\mathcal{L}\mu(s, X(s)) = \mathcal{L}\mu(t, X(t)) + \int_t^s \mathcal{L}^2\mu(v, X(v)) dv, \quad s > t,$$

such that

$$\begin{aligned} X(t + \Delta) &= X(t) + \mu(t, X(t)) \int_t^{t+\Delta} du + \mathcal{L}\mu(t, X(t)) \int_t^{t+\Delta} \int_t^u ds du \\ &\quad + \int_t^{t+\Delta} \int_t^u \int_t^s \mathcal{L}^2\mu(v, X(v)) dv ds du \\ &= X(t) + \mu(t, X(t)) \Delta + \frac{1}{2} \mathcal{L}\mu(t, X(t)) \Delta^2 + O(\Delta^3), \end{aligned} \quad (3.47)$$

which is just (3.44). We can continue the iteration to arbitrary high order.

3.2.6.2 Ito-Taylor Expansions

One may wonder why in the previous section we bothered with the integral representation of Taylor's theorem when the usual (differential) Taylor expansion lead to the correct result. The reason is that the integral approach can be extended to SDEs, leading to stochastic *Ito-Taylor expansions*. To give a flavor of these, reintroduce a diffusion term to (3.43), and start out with the integral representation

$$X(t + \Delta) = X(t) + \int_t^{t+\Delta} \mu(u, X(u)) du + \int_t^{t+\Delta} \sigma(u, X(u)) dW(u). \quad (3.48)$$

Applying Ito's lemma to μ gives (compare to (3.46))

$$\mu(u, X(u)) = \mu(t, X(t)) + \int_t^u \mathcal{L}_0\mu(s, X(s)) ds + \int_t^u \mathcal{L}_1\mu(s, X(s)) dW(s), \quad (3.49)$$

where

$$\mathcal{L}_0 \triangleq \frac{\partial}{\partial t} + \mu \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2}, \quad \mathcal{L}_1 \triangleq \sigma \frac{\partial}{\partial x}.$$

Similarly,

$$\sigma(u, X(u)) = \sigma(t, X(t)) + \int_t^u \mathcal{L}_0\sigma(s, X(s)) ds + \int_t^u \mathcal{L}_1\sigma(s, X(s)) dW(s). \quad (3.50)$$

Plugging (3.49) and (3.50) into (3.48) yields

$$\begin{aligned} X(t + \Delta) &= X(t) + \mu(t, X(t)) \int_t^{t+\Delta} du + \sigma(t, X(t)) \int_t^{t+\Delta} dW(u) + R_1 \\ &= X(t) + \mu(t, X(t)) \Delta + \sigma(t, X(t)) (W(t + \Delta) - W(t)) + R_1, \end{aligned} \quad (3.51)$$

where the remainder R_1 is

$$\begin{aligned} R_1 = & \int_t^{t+\Delta} \int_t^u \mathcal{L}_0 \mu(s, X(s)) \, ds \, du \\ & + \int_t^{t+\Delta} \int_t^u \mathcal{L}_1 \mu(s, X(s)) \, dW(s) \, du \\ & + \int_t^{t+\Delta} \int_t^u \mathcal{L}_0 \sigma(s, X(s)) \, ds \, dW(u) \\ & + \int_t^{t+\Delta} \int_t^u \mathcal{L}_1 \sigma(s, X(s)) \, dW(s) \, dW(u). \end{aligned}$$

As for the ODE example above, we can repeat this procedure arbitrarily many times. Going just one step further, we arrive at

$$\begin{aligned} X(t + \Delta) = & X(t) + \mu(t, X(t)) \Delta + \sigma(t, X(t)) (W(t + \Delta) - W(t)) \\ & + \mathcal{L}_0 \mu(t, X(t)) \frac{1}{2} \Delta^2 \\ & + \mathcal{L}_1 \mu(t, X(t)) \int_t^{t+\Delta} \int_t^u dW(s) \, du \\ & + \mathcal{L}_0 \sigma(t, X(t)) \int_t^{t+\Delta} \int_t^u ds \, dW(u) \\ & + \mathcal{L}_1 \sigma(t, X(t)) \int_t^{t+\Delta} \int_t^u dW(s) \, dW(u) + R_2, \end{aligned} \quad (3.52)$$

where R_2 contains triple integrals over t and W .

Stochastic Taylor expansions can be continued to arbitrary order, but we shall not go any further.

3.2.6.3 Milstein Second-Order Discretization Scheme

Discarding the remainder R_1 in the one-step iteration (3.51) is seen to lead to the Euler scheme (see Section 3.2.3), known to have weak convergence order $\beta = 1$. Under additional regularity (see Talay [1984]) of μ and σ , discarding the remainder R_2 in the higher-order expansion (3.52) can form the basis of a discretization scheme with weak order $\beta = 2$. For us to implement such a scheme, however, we need to concern ourselves with the simulation of the three stochastic double integrals figuring in (3.52). We go through the integrals in order below.

First,

$$\begin{aligned} I_{(1,1)} &\triangleq \int_t^{t+\Delta} \int_t^u dW(s) \, dW(u) \\ &= \int_t^{t+\Delta} (W(u) - W(t)) \, dW(u) = \frac{1}{2} (W(t + \Delta) - W(t))^2 - \frac{1}{2} \Delta, \end{aligned} \quad (3.53)$$

where we have used the fact that

$$\int_0^t W(u) dW(u) = \frac{1}{2} W(t)^2 - \frac{1}{2} t,$$

as can be verified by Ito's lemma. Second,

$$\begin{aligned} I_{(0,1)} &\triangleq \int_t^{t+\Delta} \int_t^u ds dW(u) = \int_t^{t+\Delta} (u - t) dW(u) \\ &= \Delta (W(t + \Delta) - W(t)) - \int_t^{t+\Delta} (W(u) - W(t)) du \\ &\triangleq \Delta (W(t + \Delta) - W(t)) - I_{(1,0)}, \end{aligned} \quad (3.54)$$

where we have used the integration-by-parts formula

$$\int_0^t u dW(u) = tW(t) - \int_0^t W(u) du,$$

which follows from applying Ito's lemma to $tW(t)$. In (3.54), the remaining integral $I_{(1,0)}$ on the right-hand-side is the same as the final double integral in (3.52), namely

$$I_{(1,0)} \triangleq \int_t^{t+\Delta} \int_t^u dW(s) du = \int_t^{t+\Delta} (W(u) - W(t)) du.$$

Reversing the order of integration, we get

$$\begin{aligned} I_{(1,0)} &= \int_t^{t+\Delta} \int_t^u dW(s) du \\ &= \int_t^{t+\Delta} \int_u^{t+\Delta} ds dW(u) = \int_t^{t+\Delta} (t + \Delta - u) dW(u) \end{aligned}$$

so we see, from Theorem 1.1.3 and the discussion in Section 1.6 on linear SDEs, that $I_{(1,0)}$ is Gaussian with mean 0 and variance

$$\text{Var}(I_{(1,0)}) = \int_t^{t+\Delta} (t + \Delta - u)^2 du = \frac{1}{3} \Delta^3.$$

The covariance between $I_{(1,0)}$ and $W(t + \Delta) - W(t)$ can be computed as

$$\text{Cov}(I_{(1,0)}, W(t + \Delta) - W(t)) = \int_t^{t+\Delta} (t + \Delta - u) du = \frac{1}{2} \Delta^2.$$

With the results above, we can cast the Taylor expansion (3.52) in the form of a simulation scheme (μ , σ , and their derivatives are to be evaluated at $t = i\Delta$ and $X = \widehat{X}(i\Delta)$),

$$\begin{aligned}\widehat{X}_{i+1} = & \widehat{X}_i + \mu\Delta + \sigma Z_{i,1}\sqrt{\Delta} + \mathcal{L}_0\mu\frac{1}{2}\Delta^2 + \mathcal{L}_1\mu Z_{i,2}\sqrt{\frac{1}{3}\Delta^3} \\ & + \mathcal{L}_0\sigma\left[\Delta Z_{i,1}\sqrt{\Delta} - Z_{i,2}\sqrt{\frac{1}{3}\Delta^3}\right] + \mathcal{L}_1\sigma\left(\frac{1}{2}Z_{i,1}^2\Delta - \frac{1}{2}\Delta\right),\end{aligned}\quad (3.55)$$

where $Z_{i,1}$ and $Z_{i,2}$ are sequences of $\mathcal{N}(0, 1)$ Gaussian variables with pairwise correlation

$$\rho(Z_{i,1}, Z_{i,2}) = \frac{\frac{1}{2}\Delta^2}{\sqrt{\frac{1}{3}\Delta^3}\sqrt{\Delta}} = \sqrt{\frac{3}{4}}.$$

The scheme above is known as the *Milstein scheme*. As mentioned earlier, the scheme has weak convergence order 2 under fairly strong regularity assumptions on μ and σ . We note that in the literature on SDE simulation, the Milstein scheme is often presented in a simplified form with the integral $I_{(1,0)}$ simulated as

$$\mathbb{E}\left(\int_t^{t+\Delta} \int_t^u dW(s) du \middle| W(t), W(t+\Delta)\right) = \frac{1}{2}\Delta(W(t+\Delta) - W(t)),$$

which corresponds to replacing $Z_{i,2}\sqrt{\Delta^3/3}$ with $\frac{1}{2}\Delta Z_{i,1}$ in (3.55). See Kloeden and Platen [2000] for a discussion of why this type of simplification does not affect the weak convergence order. The same source also contains a full discussion of how to extend the Milstein scheme to multi-dimensional SDEs.

3.2.7 Other Second-Order Schemes

The need to explicitly compute derivatives of the functions μ and σ often makes the Milstein scheme inconvenient to apply. High-order simulation schemes that substitute finite difference approximations for derivatives exist, and retain second (or higher) order weak convergence, are surveyed in Kloeden and Platen [2000]. To give an example of such a scheme, consider the scalar case $d = p = 1$ and assume that SDE coefficient functions μ and σ are function of x only. A derivative-free scheme that achieves second-order weak convergence is (from Kloeden and Platen [2000], Chapter 15)

$$\begin{aligned}\widehat{X}_{i+1} = & \widehat{X}_i + \frac{1}{2}\left(\mu(\overline{X}) + \mu(\widehat{X}_i)\right)\Delta \\ & + \frac{1}{4}\left(\sigma(\overline{X}^+) + \sigma(\overline{X}^-) + 2\sigma(\widehat{X}_i)\right)Z_i\sqrt{\Delta} \\ & + \frac{1}{4}\left(\sigma(\overline{X}^+) - \sigma(\overline{X}^-)\right)(Z_i^2\Delta - \Delta)\Delta^{-1/2},\end{aligned}\quad (3.56)$$

where the Z_i 's are a sequence of $\mathcal{N}(0, 1)$ Gaussian variables, and

$$\begin{aligned}\bar{X} &= \hat{X}_i + \mu(\hat{X}_i) \Delta + \sigma(\hat{X}_i) Z_i \sqrt{\Delta}, \\ \bar{X}^\pm &= \hat{X}_i + \mu(\hat{X}_i) \Delta \pm \sigma(\hat{X}_i) \sqrt{\Delta}.\end{aligned}$$

Comparison of (3.56) with the simplified Milstein scheme in the previous section shows that (3.56) avoids derivatives by using additional supporting values \bar{X} and \bar{X}^\pm .

Another, quite different, approach to avoid explicit derivatives applies the classical idea of *Richardson extrapolation* to the Euler scheme. This idea was proposed by Talay and Tubaro [1990] and takes advantage of the fact that, under additional regularity conditions, the error of the Euler scheme can be sharpened beyond (3.30) (with $\beta = 1$) to

$$E\left(g\left(\hat{X}(T)\right)\right) = E(g(X(T))) + c\Delta + O(\Delta^2), \quad (3.57)$$

for a constant c . Let \hat{X}_Δ and $\hat{X}_{2\Delta}$ be estimates of X based on Euler discretizations with time steps of Δ and 2Δ , respectively. Provided that (3.57) holds, we can write

$$2E\left(g\left(\hat{X}_\Delta(T)\right)\right) - E\left(g\left(\hat{X}_{2\Delta}(T)\right)\right) = E(g(X(T))) + O(\Delta^2), \quad (3.58)$$

which is our second-order extrapolation formula. As the Euler scheme is simple to set up, the extrapolation scheme is an attractive alternative to other second-order techniques. In practice, however, the convergence of the Euler scheme may not always be smooth enough to make (3.58) work well. Numerical experiments will nearly always be necessary (as is also the case of the Ito-Taylor schemes, for that matter).

A final word about generation of \hat{X}_Δ and $\hat{X}_{2\Delta}$ in the Richardson extrapolation scheme. To avoid duplication of work, we discretize time in buckets of Δ and generate both \hat{X}_Δ and $\hat{X}_{2\Delta}$ simultaneously, combining time steps in pairs for the purpose of generating $\hat{X}_{2\Delta}$. That is, if we use Gaussian increments of $Z_1\sqrt{\Delta}, Z_2\sqrt{\Delta}, \dots$ for \hat{X}_Δ , we use $(Z_1 + Z_2)\sqrt{\Delta}, (Z_3 + Z_4)\sqrt{\Delta}, \dots$ for $\hat{X}_{2\Delta}$. Not only do we save work by re-using Gaussian draws, we most likely also reduce the statistical error of our Monte Carlo estimate of the difference $2E(g(\hat{X}_\Delta(T))) - E(g(\hat{X}_{2\Delta}(T)))$ by raising correlation between $g(\hat{X}_\Delta(T))$ and $g(\hat{X}_{2\Delta}(T))$. We shall return to this idea in Section 3.3.1.

3.2.8 Bias vs. Monte Carlo Error

When we use an m -step discretization scheme in an n -path Monte Carlo run, we are exposed to two types of errors on the expectation we are trying to evaluate: i) the statistical Monte Carlo error e_s (the standard error); and ii) a bias e_b , originating from the discretization scheme. Raising n will reduce the standard error, but will not affect the bias which can only be

reduced by increasing the number of steps m in the time discretization scheme. Raising m and/or n obviously involves a computational cost, so let us briefly consider explicitly the trade-offs involved in simultaneously reducing bias and standard error.

Assume first that the discretization scheme has weak order β . Proceeding informally, we interpret this as implying

$$e_b = c_b \Delta^\beta,$$

for some constant c_b . Also, we know that the variance of e_s is

$$\text{Var}(e_s) = \frac{c_s}{n},$$

for a constant c_s . The total computing time τ is reasonably assumed to be proportional to nm or, using the fact that $\Delta = T/m$,

$$\tau = n \frac{c_\tau}{\Delta} \quad (3.59)$$

for some constant c_τ . For a given computing budget τ , consider minimizing the total mean-square error (MSE) $c_b^2 \Delta^{2\beta} + \frac{c_s}{n}$. Using (3.59) to eliminate a variable, the optimization problem is

$$\min_{\Delta} \left(c_b^2 \Delta^{2\beta} + \frac{c_s c_\tau}{\tau \Delta} \right).$$

Let Δ^* be the value of Δ at which the minimum MSE is attained. Δ^* is seen to satisfy

$$\Delta^* = C \tau^{-\frac{1}{2\beta+1}}, \quad C \triangleq \left(\frac{c_s c_\tau}{2\beta c_b^2} \right)^{\frac{1}{2\beta+1}}, \quad (3.60)$$

such that the minimum MSE becomes

$$C' \tau^{-\frac{2\beta}{2\beta+1}} \quad (3.61)$$

for yet another constant C' .

Equations (3.60) and (3.61) reveal a number of structural characteristics of Monte Carlo simulation of discretized SDEs. For instance, according to (3.61), the optimal root-mean-square (RMS) error behaves with the computing time τ as

$$\text{RMS} \propto \tau^{-\frac{\beta}{2\beta+1}}. \quad (3.62)$$

The computational cost of working with SDEs that are not explicitly solvable are quantified by (3.62). For an unbiased (that is, exact) SDE simulation scheme, $\beta = \infty$ and the optimal RMS error converges at the rate of $\tau^{-\frac{1}{2}}$, consistent with the results of Section 3.1. However, for an Euler scheme ($\beta = 1$) the RMS error convergence rate is lowered to $\tau^{-\frac{1}{3}}$.

Equation (3.60) in principle tells us how to optimally allocate resources between the competing objectives of a lower bias and a lower standard

error. Let m^* be defined through $\Delta^* = T/m^*$ and let n^* be defined through $\tau = c_\tau n^* m^*/T$. After a few rearrangements, we find the intuitive result

$$\sqrt{n^*} = C''(m^*)^\beta,$$

where C'' is a constant independent of τ . When we increase or decrease our computing budget, it is thus reasonable to allocate resources in such a way that we keep the factor $n^{1/2}m^{-\beta}$ constant. More detailed discussion, as well as asymptotic limit results, can be found in Duffie and Glynn [1995].

3.2.9 Sampling of Continuous Process Extremes

We round off our discussion of path simulation schemes by considering the pricing of options that depend on continuously or high-frequency sampled extremes of an SDE. We focus on the scalar case, with our SDE given as

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

where both X and W are 1-dimensional (i.e., $p = d = 1$). We also assume that the SDE is Euler-discretized according to

$$\widehat{X}_{i+1} = \widehat{X}_i + \mu(i\Delta, \widehat{X}_i) \Delta + \sigma(i\Delta, \widehat{X}_i) \sqrt{\Delta} Z_i, \quad i = 0, 1, \dots, m-1,$$

with $m\Delta = T$.

On the interval $[0, T]$, let the maximum and minimum values of $X(t)$ be denoted $M_{[0,T]}$ and $m_{[0,T]}$, respectively. That is,

$$M_{[0,T]} \triangleq \max_{0 \leq t \leq T} X(t); \quad m_{[0,T]} \triangleq \min_{0 \leq t \leq T} X(t).$$

To give examples of options that depend on $M_{[0,T]}$ and $m_{[0,T]}$, consider for instance the up-and-out call option we encountered in Section 2.7. With a knock-out barrier of H and a terminal strike of K , the terminal maturity payout can be written as¹¹

$$g(T) = 1_{\{M_{[0,T]} < H\}} (X(T) - K)^+.$$

A *double-barrier knock-out call option* with an upper barrier of H and a lower barrier of h pays

$$g(T) = 1_{\{m_{[0,T]} > h\}} 1_{\{M_{[0,T]} < H\}} (X(T) - K)^+.$$

Finally, a so-called *lookback call option* (see Section 2.7) pays

$$g(T) = (M_{[0,T]} - K)^+.$$

¹¹If $X(t)$ is the logarithm of the asset price, we replace this expression by $g(T) = 1_{\{M_{[0,T]} \leq e^H\}} (e^{X(T)} - K)^+$, and similarly for the other payouts considered.

To price options such as those above, we must provide pathwise estimates of $M_{[0,T]}$ and $m_{[0,T]}$. Given our discretization schemes, natural estimators are

$$\widehat{M}_{[0,T]} = \max(X(0), \widehat{X}_1, \dots, \widehat{X}_m), \quad (3.63)$$

$$\widehat{m}_{[0,T]} = \min(X(0), \widehat{X}_1, \dots, \widehat{X}_m). \quad (3.64)$$

Even in cases where the discretization scheme itself is perfectly unbiased, it is clear that these estimators will underestimate the range of the extremes of $X(t)$, by consistently failing to account for the movement (the “overshoot” and “undershoot”) of X between sample points $i\Delta$, $i = 0, 1, \dots, m$. As a consequence, for each simulated path in an otherwise unbiased discretization scheme, almost surely

$$\widehat{M}_{[0,T]} < M_{[0,T]}, \quad \widehat{m}_{[0,T]} > m_{[0,T]}.$$

As shown in Andersen and Brotherton-Ratcliffe [1996], the bias introduced can be very significant, even if Δ is quite small. For instance, for a 1 year lookback option, Andersen and Brotherton-Ratcliffe [1996] report that even daily sampling produces a 6% price error.

To improve the price estimates of options that depend on continuously sampled extremes, we should alter (3.63) and (3.64) to take into consideration movements between sample dates. This can be accomplished by the Brownian bridge technique introduced in Andersen and Brotherton-Ratcliffe [1996] (see also Broadie et al. [1997]). Let us focus on a particular bucket $[i\Delta, (i+1)\Delta]$ and assume, consistent with the Euler scheme, that $\widehat{X}(t)$, $t \in [i\Delta, (i+1)\Delta]$, is a Gaussian process with conditional moments

$$\begin{aligned} \mathbb{E}(\widehat{X}(t) - \widehat{X}_i \mid \widehat{X}_i) &= \mu(i\Delta, \widehat{X}_i)(t - i\Delta), \quad t \in [i\Delta, (i+1)\Delta]; \\ \text{Var}(\widehat{X}(t) - \widehat{X}_i \mid \widehat{X}_i) &= \sigma(i\Delta, \widehat{X}_i)^2(t - i\Delta), \quad t \in [i\Delta, (i+1)\Delta]. \end{aligned}$$

Assume that we have already simulated \widehat{X}_i and $\widehat{X}((i+1)\Delta)$ by the Euler scheme above. Conditional on *both* \widehat{X}_i and \widehat{X}_{i+1} , the process for $\widehat{X}(t)$, $t \in [i\Delta, (i+1)\Delta]$ is a Gaussian process “pinned” at the levels \widehat{X}_i and \widehat{X}_{i+1} . The resulting process is known as a *Brownian bridge* with diffusion coefficient $\sigma(i\Delta, \widehat{X}_i)$. Let \widehat{M}_i^c (\widehat{m}_i^c) be defined as the continuously sampled maximum (minimum) of $\widehat{X}(t)$ on $[i\Delta, (i+1)\Delta]$. The following lemma is a special case of a result in Andersen and Brotherton-Ratcliffe [1996]:

Lemma 3.2.1.

$$\begin{aligned} \mathbb{P}(\widehat{M}_i^c \leq s \mid \widehat{X}_i, \widehat{X}_{i+1}) &= 1 - \xi_i(s), \quad s > \max(\widehat{X}_i, \widehat{X}_{i+1}), \\ \mathbb{P}(\widehat{m}_i^c \leq s \mid \widehat{X}_i, \widehat{X}_{i+1}) &= \xi_i(s), \quad s < \min(\widehat{X}_i, \widehat{X}_{i+1}), \end{aligned}$$

where

$$\xi_i(s) \triangleq \exp\left(\frac{2(s - \hat{X}_i)(\hat{X}_{i+1} - s)}{\sigma(i\Delta, \hat{X}_i)^2 \Delta}\right).$$

We can use the result of the lemma in a number of ways. Most obviously, we can apply it to sample \widehat{M}_i^c and \widehat{m}_i^c directly, by the inverse transform method (see Section 3.1.1.1). To illustrate, consider for instance sampling \widehat{M}_i^c . Having first drawn \hat{X}_i and \hat{X}_{i+1} by usual means, we draw an additional independent $\mathcal{U}(0, 1)$ uniform variable U_i , and set

$$1 - \xi_i(\widehat{M}_i^c) = U_i$$

or, after a few rearrangements,

$$\widehat{M}_i^c = \frac{1}{2} (\hat{X}_{i+1} + \hat{X}_i) + \frac{1}{2} \sqrt{(\hat{X}_{i+1} - \hat{X}_i)^2 - 2\sigma(i\Delta, \hat{X}_i)^2 \Delta \ln(1 - U_i)}.$$

This procedure can be repeated for $i = 0, 1, \dots, m-1$, giving us the improved estimator for the maximum of X over $[0, T]$,

$$\widehat{M}_{[0, T]}^c = \max(\widehat{M}_0^c, \dots, \widehat{M}_{m-1}^c).$$

For options depending on both the minimum and maximum (such as double barrier options and the double lookback options in He et al. [1998]), the necessary extensions required for joint sampling of minimum and maximum are developed in Andersen [1998].

For barrier options, we note that locating \widehat{M}_i^c and \widehat{m}_i^c directly is typically not necessary, as it suffices to check locally whether the barrier is breached. For an up-and-out knock-out option with barrier H , for each interval it thus suffices to check whether $\widehat{M}_i^c > H$ which, conditional on \hat{X}_i and \hat{X}_{i+1} , happens with likelihood $\xi_i(H)$. So, provided that \hat{X}_i and \hat{X}_{i+1} are both below H , determining whether the barrier was nevertheless breached in $[i\Delta, (i+1)\Delta]$ is a matter of drawing a uniform variable U_i and setting

$$1_{\{\widehat{M}_i^c \geq H\}} = 1_{\{U_i \leq \xi_i(H)\}}. \quad (3.65)$$

This scheme is easily extended to time-dependent barriers and to cases where there are rebates¹². As pointed out in Glasserman and Staum [2001], for

¹²To get the timing of rebate payments right, the exact time that the barrier is breached must, in principle, be located. Andersen and Brotherton-Ratcliffe [1996] list analytical Brownian bridge hitting time results that can be used for this purpose. For reasonably fine discretizations, it will often suffice to set the hitting time to, say, the mid-point of the time bucket where the barrier is known to be breached.

Markov processes and the special case of barrier options with no rebates, one can in fact avoid drawing U_i 's altogether, as

$$\begin{aligned} \mathbb{E} \left(g(\widehat{X}_m) \mathbf{1}_{\{\widehat{M}_{[0,T]}^c < H\}} \middle| \widehat{X}_0, \widehat{X}_1, \dots, \widehat{X}_m \right) \\ = \mathbb{E} \left(g(\widehat{X}_m) \prod_{i=0}^{m-1} \mathbf{1}_{\{\widehat{M}_i^c < H\}} \middle| \widehat{X}_0, \widehat{X}_1, \dots, \widehat{X}_m \right) \\ = g(\widehat{X}_m) \prod_{i=0}^{m-1} \mathbb{E} \left(\mathbf{1}_{\{\widehat{M}_i^c < H\}} \middle| \widehat{X}_i, \widehat{X}_{i+1} \right). \end{aligned}$$

Here,

$$\mathbb{E} \left(\mathbf{1}_{\{\widehat{M}_i^c < H\}} \middle| \widehat{X}_i, \widehat{X}_{i+1} \right) = \begin{cases} 0, & \widehat{X}_i \geq H \text{ or } \widehat{X}_{i+1} \geq H, \\ 1 - \xi_i(H), & \widehat{X}_i < H \text{ and } \widehat{X}_{i+1} < H. \end{cases}$$

In other words, rather than explicitly simulating the indicator functions (3.65), it suffices to adjust the terminal payout by the product of conditional survival probabilities along the path. This scheme is an example of *conditional Monte Carlo*, a variance-reduction scheme discussed in more detail in Section 25.2 and in Boyle et al. [1997]. One potential drawback of the scheme is the fact that we typically need to continue the paths for a longer period of time before a barrier crossing is detected and the path can be stopped.

We round off this section with a few comments. First, we note that the schemes above assume that X is well approximated by a Gaussian process. In some applications the geometric Brownian motion, say, may be a more appropriate model. In Lemma 3.2.1, we can easily accommodate this by simply replacing \widehat{M}_i^c , \widehat{m}_i^c , s , \widehat{X}_i , and \widehat{X}_{i+1} with $\ln \widehat{M}_i^c$, $\ln \widehat{m}_i^c$, $\ln s$, $\ln \widehat{X}_i$, and $\ln \widehat{X}_{i+1}$. Other transformations are handled the same way.

Secondly, it should be pointed out that many real options are, in fact, not sampled continuously but rather at some finite but high frequency, often daily. Running an Euler scheme with daily discretization is obviously computationally inefficient. Fortunately, we can often use our scheme above as part of a Richardson-type interpolation idea. Indeed, it can often be established (see Andersen and Brotherton-Ratcliffe [1996]) that options on process extremes converge as $O(\sqrt{\Delta})$. If we first compute a price estimate \widehat{V}^Δ based on a relatively coarse value of Δ , we can then write

$$\widehat{V}^{\Delta^*} \approx \widehat{V}^c + (\widehat{V}^\Delta - \widehat{V}^c)\Delta^*/\Delta, \quad \Delta^* < \Delta,$$

where \widehat{V}^c is the continuously monitored price computed by the scheme outlined above. We have here implicitly made the assumption that the regular Euler bias is small relative to the bias induced by using the wrong sampling frequency. The idea above is developed further in Chapter V of Andersen [1996], where a number of numerical results can also be found.

And finally, one may wonder whether it is possible to deal with a continuously monitored barrier option in a discrete-time simulation by adjusting the *barrier*, rather than the underlying *process*. As it turns out, this is indeed possible. Specifically, in the Black-Scholes-Merton model with volatility σ , let $V^c(H)$, $V^\Delta(H)$ be the values of a continuously and discretely sampled barrier options with barrier H , respectively. Assuming that the discrete sampling happens on a time grid with spacing Δ , we have the following result from Broadie et al. [1997]:

Theorem 3.2.2. *The following holds,*

$$V^\Delta(H) = V\left(He^{\pm\beta\sigma\sqrt{\Delta}}\right) + o\left(\sqrt{\Delta}\right),$$

where $+$ applies if $H > X(0)$, $-$ applies if $H < X(0)$, and $\beta = \zeta(1/2)/\sqrt{2\pi} \approx 0.5826$, with $\zeta(\cdot)$ being the Riemann zeta function.

According to this result we can price a continuous barrier option by evaluating a discrete barrier option instead (e.g., one where the barrier monitoring takes place only on the simulation dates of the Monte Carlo scheme used), but with the discrete barrier level shifted according to the theorem. Theorem 3.2.2 can also be used to save computation time by, say, turning a barrier option with daily observations into an option with quarterly observations, as the theorem shows that

$$V^{\Delta^*}(H) \approx V^\Delta\left(He^{\pm\beta\sigma(\sqrt{\Delta^*}-\sqrt{\Delta})}\right).$$

While the result of Theorem 3.2.2 is only proved for the log-normal process, practical experience shows that it is robust across a wide variety of models. A similar approach exists for lookback options, see Broadie et al. [1999].

3.2.10 PCA and Bridge Construction of Brownian Motion Paths

3.2.10.1 Brownian Bridge and Quasi-Random Sequences

To close out the section on sample path simulation, let us address alternative ways of generating sample paths of Brownian motion. So far, to produce a sample of the vector $\mathbf{W} = (W(\Delta), W(2\Delta), \dots, W(m\Delta))^\top$, we have relied exclusively on the forward recursion

$$W(i\Delta + \Delta) = W(i\Delta) + Z_i\sqrt{\Delta}, \quad W(0) = 0, \quad (3.66)$$

where Z_0, Z_1, \dots, Z_{m-1} is a sequence of independent standard Gaussian variables. Rather than filling out the elements of \mathbf{W} in order, we may, for instance, rely on a *Brownian bridge (BB) construction* where we first

sample the end-point $W(m\Delta)$, then sample the mid-point¹³ $W(\lfloor m/2 \rfloor \Delta)$ *conditional* on $W(m\Delta)$, and so forth. In executing this scheme, we can use the easily proven result below.

Lemma 3.2.3. *Let $\underline{t} < t < \bar{t}$. Conditional on $W(\underline{t})$ and $W(\bar{t})$, $W(t)$ is Gaussian with moments*

$$\begin{aligned}\mathbb{E}(W(t)|W(\underline{t}) = \underline{w}, W(\bar{t}) = \bar{w}) &= \underline{w} \times \frac{\bar{t} - t}{\bar{t} - \underline{t}} + \bar{w} \times \frac{t - \underline{t}}{\bar{t} - \underline{t}}, \\ \text{Var}(W(t)|W(\underline{t}) = \underline{w}, W(\bar{t}) = \bar{w}) &= \frac{(\bar{t} - t)(t - \underline{t})}{\bar{t} - \underline{t}}.\end{aligned}$$

The BB scheme for construction of \mathbf{W} relies on repeated application of the result in Lemma 3.2.3 to progressively fill in \mathbf{W} in the “bisection” manner described above; consult any Monte Carlo textbook (e.g. Jäckel [2002] or Glasserman [2004]) if further details are required. As is the case for the standard scheme (3.66), a total of m standard Gaussian random variables are needed to construct a single sample of \mathbf{W} by the Brownian bridge scheme, so the latter offers no computational advantage over the former. Why then use the Brownian bridge construction?

One important distinction between the BB construction and (3.66) is the fact that the Brownian bridge assigns different importance to the random numbers used to produce \mathbf{W} . For instance, the very first Gaussian number drawn in the BB technique *alone* determines the end-point $W(m\Delta)$ of the Brownian motion — and thereby establishes a significant part of the overall coarse structure of the path of W . Subsequent random number draws contribute by filling in the details of the W -path, with late draws adding only to the fine-structure of the path. In contrast, with (3.66) the end-point $W(m\Delta)$ is affected equally by the m random numbers Z_0, Z_1, \dots, Z_{m-1} . In most financial problems the coarse shape of the path of W is more critical than finer details, so ultimately the BB technique allows us to identify and isolate the important features of the Brownian motion path. In some variance reduction techniques this can be important, as it allows us to focus computational effort on the random numbers that matter the most. Also, some variance reduction techniques that are known to work particularly well on low-dimensional problems can now be applied to the (low-dimensional) random numbers that contribute most to the sample path.

One relevant technique is *quasi-random sequences* (also known as *low-discrepancy sequences*), a method of generating points on the hypercube that are as “dispersed” as possible. A good survey of the underlying ideas and theory can be found in Jäckel [2002] or Glasserman [2004], with source code available (for the special case of *Sobol sequences*) in Press et al. [1992]; suffice to say that quasi-random sequences can, under some circumstances,

¹³ $\lfloor x \rfloor$ denotes the integer part of a real variable x .

accelerate Monte Carlo convergence substantially¹⁴. It is well-known, however, that the efficacy of quasi-random sequences depend strongly on the problem dimension (here: m , the number of random numbers needed per path), and that the sequences deteriorate in higher dimensions. When quasi-random sequences are combined with BB simulation of the path, however, the (low-dimensional) points of the sequences that are well-distributed can be applied to generate — by the methods in Section 3.1.1 — the Gaussian samples that determine the coarse structure of the paths, whereas the poorly distributed (high-dimensional) parts of the sequence can be relegated to the generation of less important fine-structure details¹⁵. A full account of this idea can be found in Moskowitz and Caflisch [1996].

3.2.10.2 PC Construction

With the Brownian Bridge (BB) construction of Brownian motion, much of the variance of the sample paths \mathbf{W} is explained by the values of the first few (Gaussian) random variables drawn in the path simulation. We recall from Section 3.1.3, however, that the *optimal* way to project the variation of a Gaussian vector onto a low-dimensional set of random variables is done through a principal component (PC) construction, rather than the Brownian bridge. To demonstrate how a PC construction of $\mathbf{W} = (W(\Delta), W(2\Delta), \dots, W(m\Delta))^\top$ would proceed, we first notice that the $m \times m$ variance-covariance matrix Σ of \mathbf{W} has elements

$$\Sigma_{i,j} = \mathbb{E}(W(i\Delta) W(j\Delta)) = \sqrt{\frac{\min(i, j)}{\max(i, j)}}, \quad i, j = 1, 2, \dots, m.$$

As shown in Åkesson and Lehoczy [1998], the eigenvalues of Σ can be found analytically to be

$$\lambda_i = \frac{\Delta}{4} \sin\left(\frac{\pi}{2} \cdot \frac{2i-1}{2m+1}\right)^{-2}, \quad i = 1, \dots, m,$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_m$. Let e_i be the eigenvector associated with λ_i , then it is also known that $e_i = (e_{i,1}, e_{i,2}, \dots, e_{i,m})^\top$, where

$$e_{i,j} = \frac{2}{\sqrt{2m+1}} \sin\left(j\pi \cdot \frac{2i-1}{2m+1}\right), \quad j = 1, \dots, m.$$

From the results in Section 3.1.3, we know that we can write

¹⁴Theoretically from $O(1/\sqrt{N})$ to (nearly) $O(1/N)$. Comparative tests on actual finance problems can be found in, for instance, Brotherton-Ratcliffe [1994], Paskov and Traub [1995], and Joy et al. [1996].

¹⁵Alternatively we can use regular *pseudo-random numbers* for this.

$$\mathbf{W} = \sum_{i=1}^m Z_i \sqrt{\lambda_i} e_i, \quad (3.67)$$

where Z_1, Z_2, \dots, Z_m is a sequence of independent standard Gaussian random variables. This equation constitutes the principal components construction of the Brownian path, and it is characterized by the fact that for any $k \leq m$, the first k terms of (3.67) (that is, $\sum_{i=1}^k Z_i \sqrt{\lambda_i} e_i$) explain as much of the variance of \mathbf{W} as is possible with k Gaussian variables. Even more so than for the Brownian bridge, the PC construction of a Brownian motion thus connects the overall shape of the Brownian path to a few of the Gaussian random variables Z_i , with the remaining random variables contributing only high-frequency details. As explained above, this can be useful in certain variance reduction techniques by allowing us to focus our attention and resources on just a few of the m random variables needed to simulate \mathbf{W} . We note that the PC construction is more expensive to compute than the BB technique (the latter is $O(m)$ whereas the former can be seen from (3.67) to be $O(m^2)$), so the optimality of the PC approach may, in some applications, be outweighed by its lack of speed.

While we developed the BB and PC constructions exclusively on an equidistant time grid, they easily extend to non-equidistant grids. When the grid is non-equidistant, the variance-covariance matrix of \mathbf{W} has elements

$$\Sigma_{i,j} = \mathbb{E}(W(t_i)W(t_j)) = \sqrt{\frac{\min(t_i, t_j)}{\max(t_i, t_j)}}, \quad i, j = 1, 2, \dots, m,$$

and eigenvectors and eigenvalues must then be found numerically, rather than through the analytical results listed earlier. Also, both the BB and PC techniques can easily be extended to the case of multi-variate Brownian motions, see Jäckel [2002].

Finally, for those interested in such matters, we note that in the limit $m \rightarrow \infty$, the PC construction of Brownian motion is known as the *Karhunen-Loeve decomposition*. In the continuous-time limit, the BB representation is sometimes known as a *Haar function* decomposition of Brownian motion.

3.3 Sensitivity Computations

In most finance applications, the fact that options must be dynamically hedged and risk managed requires us not only to produce an estimate of an option price, but also to compute reliable estimates of the sensitivity of the price with respect to the underlying state variables, as well as various other model parameters. In this section, we will present a number of methods for sensitivity computations by Monte Carlo methods. For each method, we use the problem of estimating the stock price delta of options in the Black-Scholes economy as a motivating example. We shall spend much more

time on sensitivity computations (by PDE and Monte Carlo methods) in Part V of this book, often in the context of particular interest rate products. Here we just give a flavor of things to come.

3.3.1 Finite Difference Estimates

3.3.1.1 Black-Scholes Delta

Consider a T -maturity European option on a dividend-free stock S in the Black-Scholes economy. Let the payout function be $g(S(T))$, and assume that the continuously compounded interest rate is a constant r . With $V(S_0)$ denoting the time 0 price of the option given $S(0) = S_0$, we are interested in computing

$$\frac{dV}{dS_0} = \lim_{h \rightarrow 0} \frac{V(S_0 + h) - V(S_0)}{h}. \quad (3.68)$$

In a Monte Carlo setting, we can approximate this derivative (“delta”) by finite difference techniques as follows. First, for some fixed number ε draw random standard Gaussian variables Z and Z_ε , and set (see Section 3.2.1)

$$S(T) = S_0 \exp \left(\left(r - \frac{1}{2}\sigma^2 \right) T + \sigma\sqrt{T}Z \right),$$

$$S_\varepsilon(T) = (S_0 + \varepsilon) \exp \left(\left(r - \frac{1}{2}\sigma^2 \right) T + \sigma\sqrt{T}Z_\varepsilon \right),$$

where σ as always denotes the constant volatility of the stock. We then form the difference

$$\delta = e^{-rT}\varepsilon^{-1} (g(S_\varepsilon(T)) - g(S(T))),$$

such that δ constitutes a single-sample estimate for dV/dS_0 . By generating n independent replications of δ and forming the sample average, we will obtain in the limit $n \rightarrow \infty$ the finite difference ratio

$$\frac{V(S_0 + \varepsilon) - V(S_0)}{\varepsilon}. \quad (3.69)$$

We know from Chapter 2 that this estimate will be biased relative to the true derivative dV/dS_0 by an amount of order $O(\varepsilon^2)$.

We have so far not mentioned whether the standard Gaussian variables Z and Z_ε should be independent or not. To analyze this, we need to consider the variance of the Monte Carlo estimator of (3.69). From Theorem 3.1.2, we know that for a finite number of trials n , the variance of our sample average will decrease as v_ε/n , where

$$\begin{aligned} v_\varepsilon &= \varepsilon^{-2} e^{-2rT} \text{Var}(g(S_\varepsilon(T)) - g(S(T))) \\ &= \varepsilon^{-2} e^{-2rT} [\text{Var}(g(S(T))) + \text{Var}(g(S_\varepsilon(T))) \\ &\quad - 2\text{Cov}(g(S_\varepsilon(T)), g(S(T)))]. \end{aligned}$$

If the random numbers Z and Z_ε are independent, $\text{Cov}(g(S_\varepsilon(T)), g(S(T)))$ will be zero and

$$v_\varepsilon \approx 2\varepsilon^{-2} e^{-2rT} \text{Var}(g(S(T))).$$

Making ε approach zero — as is needed to reduce the bias of the finite difference approximation (3.69) — will cause v_ε grow at a rate of $O(\varepsilon^{-2})$. This is obviously not ideal as our Monte Carlo estimate will be swamped by noise if ε is picked too small. On the other hand, if we set Z and Z_ε to be *identical*, we see that

$$S_\varepsilon(T) = (S_0 + \varepsilon) S(T)/S_0$$

and would expect

$$\text{Cov}(g(S_\varepsilon(T)), g(S(T))) > 0,$$

which would reduce v_ε relative to the independent case. For smooth g , a Taylor expansion in ε shows that

$$\begin{aligned} g(S_\varepsilon(T)) &= g(S(T)) + (S_\varepsilon(T) - S(T)) g'(S(T)) + \dots \\ &= g(S(T)) + \varepsilon S(T)/S_0 \cdot g'(S(T)) + \dots \end{aligned}$$

If derivatives of g are bounded

$$\text{Cov}(g(S_\varepsilon(T)), g(S(T))) = \text{Var}(g(S(T))) + O(\varepsilon^2),$$

and similarly for $\text{Var}(g(S_\varepsilon(T)))$. In other words,

$$v_\varepsilon = e^{-2rT} O(1) \tag{3.70}$$

which is a clear improvement over the earlier $O(\varepsilon^{-2})$ result.

The result (3.70) hinged on the payout function having bounded derivatives. We can, in fact, relax this considerably, to functions that are essentially just continuous in the stock price; see Section 3.3.2.2 for a discussion. For discontinuous payouts, however, (3.70) will not hold. To demonstrate, consider a digital option paying

$$g(S(T)) = 1_{\{S(T) > K\}},$$

for some strike K . With $Z = Z_\varepsilon$ we get (assuming $\varepsilon > 0$ and that the probability measure is P)

$$\begin{aligned} \mathbb{E}\left([g(S_\varepsilon(T)) - g(S(T))]^2\right) &= P(S(T) \leq K < S_\varepsilon(T)) \\ &= P(S(T) \leq K < (1 + \varepsilon/S_0) S(T)) = O(\varepsilon), \end{aligned}$$

compared with the $O(\varepsilon^2)$ result for smooth g .

3.3.1.2 General Case

To generalize the problem considered in the previous section, we consider a setting where a random variable Y depends on a parameter $\alpha \in \mathbb{R}$, in the sense that each value of α uniquely determines a scheme for the generation of Y . The random variable Y will typically represent a (discounted) option payout, and α is typically an initial value of an asset price (as in Section 3.3.1.1) or a parameter in the (vector) equations determining the dynamics of the underlying model. Let

$$V(\alpha) = \mathbb{E}(Y(\alpha)),$$

and consider the problem of determining $dV/d\alpha$.

In the basic finite difference Monte Carlo approximation to $dV/d\alpha$, we use the sample average of one-sided difference coefficients,

$$\bar{\delta}_n = \frac{\bar{Y}_n(\alpha + \varepsilon) - \bar{Y}_n(\alpha)}{\varepsilon},$$

where $\bar{Y}_n(\alpha)$ is the sample average of n realizations of $Y(\alpha)$. In the limit,

$$\lim_{n \rightarrow \infty} \bar{\delta}_n = \frac{V(\alpha + \varepsilon) - V(\alpha)}{\varepsilon} = dV/d\alpha + O(\varepsilon^2).$$

If we instead wish to use a central estimator

$$\bar{\delta}_n^c = \frac{\bar{Y}_n(\alpha + \varepsilon) - \bar{Y}_n(\alpha - \varepsilon)}{2\varepsilon},$$

we get

$$\lim_{n \rightarrow \infty} \bar{\delta}_n^c = \frac{V(\alpha + \varepsilon) - V(\alpha - \varepsilon)}{2\varepsilon} = dV/d\alpha + O(\varepsilon^3),$$

but now need to simulate an extra random variable (that is, $Y(\alpha - \varepsilon)$), increasing the computational cost.

In the generation of $\bar{Y}_n(\alpha \pm \varepsilon)$ and $\bar{Y}_n(\alpha)$, the individual samples of $Y(\alpha + \varepsilon)$, $Y(\alpha - \varepsilon)$, and $Y(\alpha)$ would typically be based on a series of draws of vector-valued support variables Z , with $Y(\alpha) = Y(Z; \alpha)$, and so forth. For instance, in an m -step Euler simulation of an SDE with d Brownian motions, each SDE path (and each outcome of Y) would involve $d \cdot m$ i.i.d. standard Gaussian variables $Z_1, \dots, Z_{d \cdot m}$. The observations in the previous section tell us that to minimize variance we should use the same Z for $Y(\alpha + \varepsilon)$, $Y(\alpha - \varepsilon)$, and $Y(\alpha)$. In practice, this is often easiest to accomplish by simply using the same random number seed (see Section 3.1.1) in otherwise separate computations of each of the quantities $\bar{Y}_n(\alpha + \varepsilon)$, $\bar{Y}_n(\alpha - \varepsilon)$, and $\bar{Y}_n(\alpha)$. Assuming this so-called *common random number scheme* is followed, the variance analysis in Section 3.3.1.1 can be generalized to our setting, and we would expect that either i) $\text{Var}(\bar{\delta}_n) = \text{Var}(\bar{\delta}_n^c) = O(\varepsilon^{-1}n^{-1})$; or ii)

$\text{Var}(\bar{\delta}_n) = \text{Var}(\bar{\delta}_n^c) = O(n^{-1})$. Case ii) essentially requires a.s. continuity¹⁶ of $Y(\alpha)$ with respect to α , as would be the case when Y represents a continuous option payout function. Case i) generally applies when Y represents a discontinuous option payout, such as the digital option considered in Section 3.3.1.1.

If case ii) above applies, the estimator variance is independent of ε , and ε should be picked as small as possible (a matter of machine precision) to minimize the $O(\varepsilon^2)$ and $O(\varepsilon^3)$ biases of $\bar{\delta}_n$ and $\bar{\delta}_n^c$. If the overhead of evaluating $Y(Z; \alpha)$ for given Z is small relative to the cost of generating Z , the central estimator $\bar{\delta}_n^c$ will dominate. For complicated payout functions, however, there may be situations when $\bar{\delta}_n$ is preferable, despite its slower convergence rate in ε . For case i), we must weigh bias against variance in a manner quite similar to the discussion in Section 3.2.8: if ε is small the difference coefficient bias will be small, but the variance of the estimators $\bar{\delta}_n$ and $\bar{\delta}_n^c$ will be high. An RMS minimization similar to the one in Section 3.2.8 is possible; in the interest of brevity, we leave this as an exercise to the reader (see also Glasserman [2004], Chapter 7).

3.3.2 Pathwise Estimate

3.3.2.1 Black-Scholes Delta

Reverting back to the setting of Section 3.3.1.1, let us take another look at the delta definition (3.68):

$$\begin{aligned}\frac{dV}{dS_0} &= \lim_{h \rightarrow 0} \frac{V(S_0 + h) - V(S_0)}{h} \\ &= e^{-rT} \lim_{h \rightarrow 0} \mathbb{E} \left(\frac{g(S_h(T)) - g(S(T))}{h} \right),\end{aligned}\tag{3.71}$$

where we have used the same notation as earlier:

$$S_h(T) = (S_0 + h) S(T)/S_0.\tag{3.72}$$

Under sufficient regularity on g , we can interchange expectation and limit in (3.71), such that simply

$$\begin{aligned}\frac{dV}{dS_0} &= e^{-rT} \mathbb{E} \left(\lim_{h \rightarrow 0} \frac{g(S_h(T)) - g(S(T))}{h} \right) \\ &= e^{-rT} \mathbb{E} \left(g'(S(T)) \frac{dS(T)}{dS_0} \right) \\ &= e^{-rT} \mathbb{E} \left(g'(S(T)) \frac{S(T)}{S_0} \right),\end{aligned}\tag{3.73}$$

¹⁶More precisely, we need uniform integrability in the difference coefficients $[Y(\alpha + \varepsilon) - Y(\alpha)]\varepsilon^{-1}$ and $\frac{1}{2}[Y(\alpha + \varepsilon) - Y(\alpha - \varepsilon)]\varepsilon^{-1}$. See Section 3.3.2.2 for more precise conditions.

where $g'(x) \triangleq dg/dx$, and the last equality follows by the linearity of (3.72).

We can implement the result (3.73) directly in a Monte Carlo trial, by generating samples of $S(T)$ and recording the sample averages of $g'(S(T))S(T)/S_0$. The resulting estimate for dV/dS_0 is a direct and unbiased estimate of the true derivative; it is known as a *pathwise estimate*.

For (3.73) to hold, g should be continuous, but does not necessarily need to be differentiable everywhere (it suffices that g is Lipschitz continuous, as discussed below). A regular call option payout $g(x) = (x - K)^+$, for instance, is non-differentiable at $x = K$, but we simply write

$$g'(x) = 1_{\{x>K\}}$$

and proceed directly with (3.73). For discontinuous payouts¹⁷, however, care must be taken as a direct application of (3.73) will introduce a bias. To demonstrate, consider the case $g(x) = 1_{\{x>K\}}$. Proceeding informally, a literal application of (3.73) results in

$$\frac{dV}{dS_0} = e^{-rT} \mathbb{E} \left(\delta(S(T) - K) \frac{S(T)}{S_0} \right),$$

where $\delta(x)$ is the Dirac delta function. While correct, this result is unsuited for Monte Carlo simulation: no matter how many samples n we draw of $S(T)$, the likelihood of $\delta(S(T) - K)$ being non-zero is zero, and the derivative would almost surely be estimated as 0. The correct result, however, is

$$\begin{aligned} e^{-rT} \mathbb{E} \left(\delta(S(T) - K) \frac{S(T)}{S_0} \right) &= e^{-rT} \mathbb{E} \left(\delta(S(T) - K) \frac{K}{S_0} \right) \\ &= e^{-rT} \frac{K}{S_0} \mathbb{E} (\delta(S(T) - K)) \\ &= e^{-rT} \frac{K}{S_0} \varphi_S(K), \end{aligned}$$

where $\varphi_S(\cdot)$ is the density of $S(T)$.

3.3.2.2 General Case

In a general setting, the technique employed in Section 3.3.2.1 above is known as *infinitesimal perturbation analysis*. A broad overview of the technique can be found in Glasserman [2004], with applications to finance covered in Broadie and Glasserman [1996]. Our treatment follows the latter closely.

Borrowing the notation of Section 3.3.1.2, we again consider estimating $dV/d\alpha$, where $V(\alpha) = \mathbb{E}(Y(\alpha))$. The basic idea of the pathwise derivative estimate is to write

¹⁷Or for the evaluation of, say, the second derivative (gamma) of a call payout.

$$\frac{dV}{d\alpha} = \frac{d}{d\alpha} \mathbb{E}(Y(\alpha)) = \mathbb{E}\left(\frac{d}{d\alpha} Y(\alpha)\right). \quad (3.74)$$

The exchange of expectation and differentiation requires certain regularity conditions to be valid. In practice, the most interesting situation arises when Y represents a (discounted) payout function, such that

$$Y(\alpha) = g(X(\alpha)),$$

where $X(\alpha) = (X_1(\alpha), \dots, X_q(\alpha))^{\top}$ is a q -dimensional random vector of observations (possibly at different dates) of asset prices. In this case, we have the following result, from Broadie and Glasserman [1996]:

Proposition 3.3.1. *For all α in some open interval \mathcal{A} assume that $dX_i/d\alpha$ exists almost surely for all $i = 1, \dots, q$. Suppose that the function g is almost surely differentiable¹⁸ and is Lipschitz, such that*

$$|g(x) - g(y)| \leq k|x - y|$$

for some constant k . Finally, assume that there exists finite-mean random variables β_i , $i = 1, \dots, q$, such that for all $\alpha_1, \alpha_2 \in \mathcal{A}$

$$|X_i(\alpha_2) - X_i(\alpha_1)| \leq \beta_i |\alpha_2 - \alpha_1|.$$

In this case, (3.74) holds.

The first two assumptions of Proposition ensure that the random variable $dY(\alpha)/d\alpha$ exists almost surely, with its value given by the chain rule

$$\frac{d}{d\alpha} Y(\alpha) = \sum_{i=1}^q \frac{\partial g}{\partial X_i} \frac{dX_i}{d\alpha}. \quad (3.75)$$

As we saw earlier, in Section 3.3.2.1, almost sure existence of $dY(\alpha)/d\alpha$ is not sufficient for the pathwise method to yield an unbiased estimator, we also need, roughly speaking, for g to be continuous at the points at which differentiability fails. The last two conditions ensure this, and together imply that Y is almost surely Lipschitz in α :

$$|Y(\alpha_2) - Y(\alpha_1)| \leq \beta_Y |\alpha_2 - \alpha_1|, \quad \beta_Y = k \sum_{i=1}^q \beta_i.$$

As $\alpha^{-1}|Y(h + \alpha) - Y(h)|$ is then bounded by β_Y , where $\mathbb{E}(\beta_Y) < \infty$, the result of Proposition 3.3.1 follows from the dominated convergence theorem. See Broadie and Glasserman [1996] for further details.

¹⁸That is, differentiable everywhere except on some set \mathcal{X} where $P(X(\alpha) \in \mathcal{X}) = 0$.

Remark 3.3.2. If in Proposition 3.3.1 we further assume that $E(\beta_i^2) < \infty$, it follows that $E(Y(h + \varepsilon) - Y(h)) \leq \beta_Y^2 \varepsilon^2$, such that

$$\text{Var}(Y(h + \varepsilon) - Y(h)) = O(\varepsilon^2).$$

We recognize this as case ii) from Section 3.3.1.2 on finite difference estimates, for which we have now made the regularity conditions more precise. In practice, the Lipschitz continuity of g is the critical condition.

Remark 3.3.3. For discontinuous payouts, the pathwise method will yield a biased estimator. As we saw in Section 3.3.2.1, however, for a simple process where the transition density is known, the bias can often be accounted for. We shall see an example of this in Chapter 24. Notice that if the transition density is known, another method for sensitivity simulation — the likelihood ratio method — also applies. See Section 3.3.3 below for details about this method.

3.3.2.3 Sensitivity Path Generation

In Section 3.3.2.1, simulation of $dY(\alpha)/d\alpha$ was straightforward due to the simplicity of the Black-Scholes dynamics. In general, we see from (3.75) that generation of the random variables $dY(\alpha)/d\alpha$ will require us to compute the partial derivatives of the payout with respect to the underlying assets ($\partial g/\partial X_i$), as well as the sensitivities of the assets with respect to the perturbation parameter α ($dX_i/d\alpha$). The latter is normally the most difficult, and we shall outline a general approach here.

For illustration, consider a scalar SDE of the usual form

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

where X and W are one-dimensional. Let α be a parameter on which $X(t)$ depends (such as $X(0)$ or some parameter of μ or σ). Let $D_\alpha(t)$ denote $dX(t)/d\alpha$. Formally differentiating the SDE with respect to α , we get

$$dD_\alpha(t) = \mu'(t, X(t)) D_\alpha(t) dt + \sigma'(t, X(t)) D_\alpha(t) dW(t), \quad (3.76)$$

where $\mu'(t, x) = \partial\mu(t, x)/\partial x$, and similar for σ' . This SDE can be discretized and simulated in parallel with the simulation of the SDE for $X(t)$ itself. In general, the work associated with this will obviously be more substantial than for the Black-Scholes delta, where we saw that $D_\alpha(t)$ could be recovered as the simple fraction $X(t)/X(0)$ (see equation (3.73)).

A few notes on the technique above. First, some regularity is obviously needed for (3.76) to be meaningful; see Kunita [1990] for some relevant results. Second, extensions to multi-dimensional SDEs are straightforward, although the dimension of the total scheme can be large. For instance, if X is p -dimensional and we wish to compute sensitivities with respect to $X_i(0)$,

$i = 1, \dots, p$, a $p \times p$ system of SDEs for quantities $dX_i(t)/dX_j(0)$ will be required. We shall discuss approximative methods to improve efficiency of such high-dimensional matrix SDEs later, in Chapter 24 (see also Glasserman and Zhao [1999]).

3.3.3 Likelihood Ratio Method

As discussed above, the pathwise derivative method typically applies only to options with sufficiently smooth payouts¹⁹ and can be cumbersome for multi-dimensional SDEs. For processes with explicitly known transition densities, the alternative *likelihood ratio method* can be used. This method applies to discontinuous payout functions and, unlike the pathwise method, requires little knowledge of the payout function and its derivatives, making it convenient for general implementation on a computer. When both methods apply, however, the pathwise derivative method generally is more efficient.

3.3.3.1 Black-Scholes Delta

In the notation of Sections 3.3.1.1 and 3.3.2.1, the Black-Scholes price of a call option can be written

$$\begin{aligned} V(S_0) &= e^{-rT} \mathbb{E} \left((S(T) - K)^+ \mid S(0) = S_0 \right) \\ &= e^{-rT} \mathbb{E} \left(\left(e^{\ln S_0 + (r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z} - K \right)^+ \right) \\ &= e^{-rT} \mathbb{E} \left(\left(e^{Y(T)} - K \right)^+ \right), \end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$ and

$$Y(T) \sim \mathcal{N} \left(\ln S_0 + \left(r - \frac{1}{2}\sigma^2 \right) T, \sigma^2 T \right).$$

The density of Y is thereby a function of S_0 :

$$\begin{aligned} P(Y \in dy) &= \frac{dy}{\sqrt{2\pi}\sigma\sqrt{T}} \exp \left(-\frac{1}{2} \left(\frac{y - \ln S_0 - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \right)^2 \right) \\ &\triangleq \varphi(y; S_0) dy. \end{aligned}$$

Thereby

$$V(S_0) = e^{-rT} \int_{-\infty}^{\infty} (e^y - K)^+ \varphi(y; S_0) dy, \quad (3.77)$$

such that

¹⁹But see Remark 3.3.3.

$$\begin{aligned}
\frac{dV(S_0)}{dS_0} &= e^{-rT} \int_{-\infty}^{\infty} (e^y - K)^+ \frac{\partial \varphi(y; S_0)}{\partial S_0} dy \\
&= e^{-rT} \int_{-\infty}^{\infty} (e^y - K)^+ \frac{\partial \varphi(y; S_0)}{\partial S_0} \frac{\varphi(y; S_0)}{\varphi(y; S_0)} dy \\
&= e^{-rT} \int_{-\infty}^{\infty} (e^y - K)^+ \frac{\partial \ln \varphi(y; S_0)}{\partial S_0} \varphi(y; S_0) dy. \tag{3.78}
\end{aligned}$$

Comparison of (3.78) with (3.77) demonstrates that we can effectively compute the Black-Scholes delta as the price of a security that pays out at time T the amount

$$(e^{Y(T)} - K)^+ l(Y(T)), \tag{3.79}$$

where l is the so-called *log-likelihood ratio* (also known as the *score function*)

$$l(Y(T)) = \frac{\partial \ln \varphi(Y(T); S_0)}{\partial S_0} = \frac{Y(T) - \ln S_0 - (r - \frac{1}{2}\sigma^2)T}{S_0 \sigma^2 T} = \frac{Z}{S_0 \sigma \sqrt{T}}. \tag{3.80}$$

By differentiating the density, rather than the payout itself, the likelihood ratio technique applies to even discontinuous payouts, requiring only that the density is smooth (which is clearly the case here). Notice in particular that the log-likelihood ratio is independent of the payout, allowing us to use the same function $l(Y(T))$ for all European style payout functions $V(T) = g(S(T))$.

3.3.3.2 General Case

As in Section 3.3.2.2, consider now the general case where a random variable $Y(\alpha)$ represents a (deflated) payout function g applied to a vector of random variables $X(\alpha) = (X_1(\alpha), \dots, X_q(\alpha))^{\top}$. Again, α is a parameter with respect to which we wish to compute sensitivities. Let the joint density of $X(\alpha)$ be denoted $f(x; \alpha)$, $x \in \mathbb{R}^q$. We then have

$$V(\alpha) = \int_{\mathbb{R}^q} g(x) f(x; \alpha) dx.$$

Making the reasonable assumption that density $f(x; \alpha)$ is a smooth function of α , we interchange integration and differentiation, such that

$$\frac{\partial V(\alpha)}{\partial \alpha} = \int_{\mathbb{R}^q} g(x) \frac{\partial f(x; \alpha)}{\partial \alpha} dx = \int_{\mathbb{R}^q} g(x) \frac{\partial \ln f(x; \alpha)}{\partial \alpha} f(x; \alpha) dx.$$

As for the Black-Scholes case above, the derivative $\partial V(\alpha)/\partial \alpha$ can thus be computed as the expectation of the payout modified by a log-likelihood ratio:

$$g(x)l(x), \quad l(x) = \frac{\partial \ln f(x; \alpha)}{\partial \alpha}.$$

3.3.3.3 Euler Schemes

In practice, the reliance on explicit knowledge of a transition density can be a considerable obstacle, and may rule out the application of the likelihood ratio method for many complex models. In cases where process dynamics are simulated through a simple time-discretization scheme, the situation is, however, salvageable, as we shall now demonstrate.

For illustration, consider an asset $X(t)$ that follows an SDE of the type

$$dX(t) = \mu(t, X(t); \alpha) dt + \sigma(t, X(t); \alpha) dW(t),$$

where μ and σ are smooth functions, and where α is a parameter. In general, we do not know the exact transition density for $X(t)$. However, suppose now that we use an Euler scheme to simulate on some grid $\{t_i\}_{i=1}^m$, i.e.

$$\widehat{X}(t_{i+1}) = \widehat{X}(t_i) + \mu(t_i, \widehat{X}(t_i); \alpha) (t_{i+1} - t_i) + \sigma(t_i, \widehat{X}(t_i); \alpha) \sqrt{t_{i+1} - t_i} Z_i,$$

where $t_0 = 0$ and Z_0, Z_1, \dots, Z_{m-1} is a sequence of i.i.d. standard Gaussian random variables. Clearly, the transition density for the $\widehat{X}(t_{i+1})$ is now Gaussian,

$$\widehat{X}(t_{i+1}) | \widehat{X}(t_i) \sim \mathcal{N}\left(m_i(\widehat{X}(t_i); \alpha), s_i(\widehat{X}(t_i); \alpha)\right),$$

where, for $i = 0, \dots, m-1$,

$$\begin{aligned} m_i(\widehat{X}(t_i); \alpha) &= \widehat{X}(t_i) + \mu(t_i, \widehat{X}(t_i); \alpha) (t_{i+1} - t_i), \\ s_i(\widehat{X}(t_i); \alpha) &= \sigma(t_i, \widehat{X}(t_i); \alpha)^2 (t_{i+1} - t_i). \end{aligned}$$

Set $\widehat{X} = (\widehat{X}(t_1), \dots, \widehat{X}(t_m))^\top$; the density of this vector is (where $x_0 = X(0)$)

$$f(x_1, \dots, x_m; \alpha) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi} \sqrt{s_{i-1}(x_{i-1}; \alpha)}} \exp\left(-\frac{(x_i - m_{i-1}(x_{i-1}; \alpha))^2}{2s_{i-1}(x_{i-1}; \alpha)}\right). \quad (3.81)$$

Consider some (potentially path-dependent) security V with payout function $g(\widehat{X})$ and time 0 price of $V(\alpha)$. Equipped with (3.81), we can estimate the parameter sensitivity by the likelihood ratio method as

$$\frac{\partial V(\alpha)}{\partial \alpha} = E\left(g(\widehat{X}) \frac{d}{d\alpha} \ln(f(\widehat{X}; \alpha))\right),$$

where

$$\begin{aligned} &\frac{d}{d\alpha} \ln(f(\widehat{X}; \alpha)) \\ &= \frac{d}{d\alpha} \sum_{i=1}^m \left(-\frac{1}{2} \ln(s_{i-1}(x_{i-1}; \alpha)) - \frac{(x_i - m_{i-1}(x_{i-1}; \alpha))^2}{2s_{i-1}(x_{i-1}; \alpha)}\right). \quad (3.82) \end{aligned}$$

This derivative can typically be computed in closed form; if not, one can estimate it by finite differences (as in Su and Randall [2008]). Notice that when $\alpha = X(0)$ (i.e. we are trying to estimate the delta), only s_0 and m_0 will depend on α , simplifying computations.

The idea used above extends easily to vector-valued $X(t)$. In principle, higher-order schemes (e.g. the Milstein scheme) can also be used, although the complexity increases considerably.

3.3.3.4 Some Remarks

The main advantage of the likelihood ratio is the fact that it applies to classes of payouts for which other methods (pathwise methods, finite difference method) do not work well. Moreover, the method is easy and efficient to implement, as the log-likelihood ratio l is independent of the payout and does not — unlike the general pathwise method — require simulation of any quantities other than the vector X itself. As discussed above, the primary drawback of the method is its reliance on explicit knowledge of the process density, ruling out many of the more advanced models (although, as Section 3.3.3.3 demonstrated, there may sometimes be ways around this). Further, the variance of the likelihood ratio method can often be quite big, particularly if the parameter α simultaneously affects multiple stochastic variables. A fuller discussion of this issue, as well as the related issue of absolute continuity, can be found in Glasserman [2004]. We note in passing that the likelihood ratio method is a special case of a body of methods that have emerged from the so-called *Malliavin calculus*; see Fournie et al. [1999] for a survey. Most Malliavin methods other than the basic likelihood ratio method are, however, not particularly attractive due to computational issues²⁰.

We round off this section by noting that the various methods for derivative estimates can often be successfully combined. For instance, while it is common to use either the pathwise method or the likelihood ratio method to compute first order sensitivities (such as delta), second-order sensitivities (such as gamma) are often done by the finite difference method applied to first-order sensitivities, often using fairly sizable shifts of the underlying variables. By combining the pathwise method with the likelihood ratio method, we can also address the fact that the first derivative of many kinked option payouts is discontinuous, allowing us to produce a bias-free estimate of the second derivative. See Fournie et al. [1999] for other examples of combining the pathwise method with the likelihood ratio method.

²⁰Besides, the Malliavin calculus itself, a very technical area of mathematics even for specialists, can be avoided altogether, as Chen and Glasserman [2007b] demonstrate.

3.4 Variance Reduction Techniques

As discussed earlier, the convergence of the Monte Carlo method is quite slow, of order $O(n^{-1/2})$ where n is the number of Monte Carlo samples. While there is little that can be done to improve²¹ the $n^{-1/2}$ order itself, the constant multiplying $n^{-1/2}$ can be affected by a careful choice of the Monte Carlo estimator. Methods to improve numerical efficiency this way are known collectively as *variance reduction techniques*, and constitute a major area of research in the theory of Monte Carlo methods. Our introduction of the topic is limited to a few basic examples. More details are provided later in the book for concrete models and products (see e.g. Chapter 25), and more information can be found in the standard Monte Carlo literature, including Hammersley and Handscomb [1965] and the survey article Boyle et al. [1997].

3.4.1 Variance Reduction and Efficiency

We recall that the goal of the Monte Carlo method is to estimate some quantity μ (e.g., the price of a financial contract) as the sample mean of n i.i.d. random variables Y_1, \dots, Y_n , where each Y_i has expectation $E(Y_i) = \mu$ and variance $\text{Var}(Y_i) = \sigma^2$. From Section 3.1, we know that for large n the standard error of the sample mean $n^{-1} \sum Y_i$ is $\sigma n^{-1/2}$, with the probabilistic error bounds on the estimate of μ being proportional to the standard error.

Suppose now that we have available two sets of i.i.d. sequences $Y_{1,i}$ and $Y_{2,i}$, $i = 1, \dots, n$, where $E(Y_{1,i}) = E(Y_{2,i}) = \mu$, but $\text{Var}(Y_{1,i}) = \sigma_1^2$ and $\text{Var}(Y_{2,i}) = \sigma_2^2$, with $\sigma_1 \neq \sigma_2$. Also suppose that the time it takes on a computer to generate individual samples $Y_{1,i}$ and $Y_{2,i}$ is τ_1 and τ_2 , respectively. Which of the two estimators $n^{-1} \sum Y_{1,i}$ and $n^{-1} \sum Y_{2,i}$ is preferable? To answer this question, assume that we have a large fixed computing time budget τ . The number of replications of $Y_{1,i}$ and $Y_{2,i}$ that can be executed are thus (the integer parts of) τ/τ_1 and τ/τ_2 , respectively. To this correspond sample mean standard errors of

$$\frac{\sigma_1}{\sqrt{\tau/\tau_1}} \text{ and } \frac{\sigma_2}{\sqrt{\tau/\tau_2}},$$

respectively. It follows that, for large τ , the estimator based on the sequence $Y_{1,i}$, $i = 1, \dots, n$, is preferable, if

$$\frac{\sigma_1}{\sqrt{\tau/\tau_1}} < \frac{\sigma_2}{\sqrt{\tau/\tau_2}},$$

or equivalently

²¹As we discussed earlier, the quasi-random Monte Carlo method can theoretically achieve better convergence order than $O(n^{-1/2})$.

$$\sigma_1^2 \tau_1 < \sigma_2^2 \tau_2.$$

For obvious reasons, the product of variance and per-sample computing time is known as the *efficiency* of a Monte Carlo estimator. In devising methods to improve Monte Carlo performance, efficiency should always constitute the measure of comparison. For instance, a high-variance estimator may, in fact, be preferable to a low-variance estimator, provided that the former takes less time to compute than the latter.

Duffie and Glynn [1995] discuss Monte Carlo efficiency in more depth, with additional analysis of the effects of bias (see also Section 3.2.8) and cases where τ_1 and τ_2 are random.

3.4.2 Antithetic Variates

3.4.2.1 The Gaussian Case

A simple and easily implemented variance reduction technique is the method of *antithetic variates*. Assume that we are interested in estimating the expected value of a random variable $Y = G(Z)$, where G is a real-valued function and Z is a q -dimensional vector of independent standard Gaussian random variables. This problem routinely arises in determining the expected value of a function of assets driven by a vector SDE; Z then represents the aggregation of all independent standard Gaussian variables used to produce Brownian motion increments; see for instance Section 3.2.3. For n independent realizations of Z , Z_1, \dots, Z_n , rather than using the regular sample average estimator for $E(Y)$, consider instead using

$$\bar{Y}_n^a = n^{-1} \sum_{i=1}^n \frac{G(Z_i) + G(-Z_i)}{2}.$$

In other words, in addition to the set Z_1, \dots, Z_n of Gaussian samples, we also effectively include the set $-Z_1, \dots, -Z_n$ in the Monte Carlo trial. As $-Z_1, \dots, -Z_n$ itself is a sequence of n independent Gaussian samples, we still must have

$$E(\bar{Y}_n^a) = E(Y),$$

so the antithetic estimator is unbiased. Also, as $G(Z)$ and $G(-Z)$ have identical variance,

$$\text{Var}(\bar{Y}_n^a) = n^{-1} \left[\frac{1}{2} \text{Var}(Y) + \frac{1}{2} \text{Cov}(G(Z), G(-Z)) \right] = \frac{\text{Var}(Y)}{n} \frac{(1 + \rho)}{2},$$

where ρ is the correlation between $G(Z)$ and $G(-Z)$. Recalling that the regular sample average has variance

$$\text{Var}(\bar{Y}_n) = \frac{\text{Var}(Y)}{n},$$

we conclude that $\text{Var}(\bar{Y}_n^a) < \text{Var}(\bar{Y}_n)$ as long as $\rho < 1$ (which is obviously likely).

While use of antithetic variates can always be expected to lower the standard error, it is not necessarily more efficient than regular Monte Carlo, in the sense defined in Section 3.4.1. For instance, if generation of the Z_i 's is of negligible cost relative to the evaluation of $G(Z)$, computation of \bar{Y}_n^a will take about twice as long as the regular sample average \bar{Y}_n . For this case, the results in Section 3.4.1 show that for antithetic variates to constitute an improvement in computational efficiency, we must require that

$$\text{Var}(\bar{Y}_n^a) \leq \frac{1}{2}\text{Var}(\bar{Y}_n),$$

or

$$\rho \leq 0.$$

A sufficient condition for $\rho < 0$ is that G be monotone in all q elements of Z . Given this, we would expect antithetic variates to be most suitable for option payouts that depend monotonically on prices.

3.4.2.2 General Case

While the method of antithetic variates is primarily associated with the idea of changing signs on Gaussian variables, the method can, in fact, be extended to other distributions. At the most basic level, most simulation trials involve a series of uniform draws that are translated to other random variables, using techniques described in Section 3.1.1. In this case, we can focus our attention on estimating the mean of a random variable $Y = H(U)$, where H is a function and U is a vector of independent uniformly distributed random variables. We notice that if $U = (U_1, \dots, U_q)^\top$ is a vector of independent uniform random variables on $[0, 1]$, then so is $\tilde{U} = (1 - U_1, \dots, 1 - U_q)^\top$. The pair $\{U, \tilde{U}\}$ is thereby antithetic (negatively dependent) in the same way as the Gaussian pair $\{Z, -Z\}$ above, and we can estimate the mean of Y as the average of independent samples of the form

$$\frac{H(U) + H(\tilde{U})}{2}.$$

From the discussion above, it follows that if H is monotonic in U , the resulting scheme will exhibit better computational efficiency than regular Monte Carlo.

As an aside, we note that the simple “reflection” of a vector of uniforms advocated above is, as should be obvious, not the only possible way of generating an antithetic sample — for instance, we could have chosen to reflect only select dimensions of the U -vector. A similar observation holds for the case of vector-valued Gaussian variables. The general idea of applying deterministic transformations to a vector-valued sample of random numbers as a way to reduce variance is sometimes known as *systematic sampling*.

3.4.3 Control Variates

3.4.3.1 Basic Idea

While we may need to use Monte Carlo simulation to estimate the unknown mean of a random variable Y , there may be random variables “close” to Y with means that can be computed analytically. It seems reasonable that the additional information about Y revealed by these random variables could be useful in improving our estimate of $E(Y)$. While a number of strategies are possible²², we shall here focus on the so-called *control variate* method.

Formally, let

$$Y^c = (Y_1^c, \dots, Y_q^c)^\top$$

be a vector of *control variates* (or just *controls*), ideally with strong negative or positive correlation to a variable Y . The mean of Y^c is known to be

$$E(Y^c) = \mu^c = (\mu_1^c, \dots, \mu_q^c)^\top.$$

Now, introduce an arbitrary constant vector

$$\beta = (\beta_1, \dots, \beta_q)^\top$$

and consider forming the linear combination

$$X = Y - \beta^\top (Y^c - \mu^c). \quad (3.83)$$

Clearly

$$E(X) = E(Y) - \beta^\top (E(Y^c) - \mu^c) = E(Y),$$

so using Monte Carlo sampling to estimate the mean of X will provide an unbiased estimate of $E(Y)$, regardless of the choice of β .

To analyze the variance of the new variable X , let Σ_{Y^c} be the $q \times q$ covariance matrix of the vector Y^c , and let Σ_{Y,Y^c} be the q -dimensional vector of covariances between Y and the components of Y^c . The variance of X can then be shown to be

$$\text{Var}(X) = \text{Var}(Y) - 2\beta^\top \Sigma_{Y,Y^c} + \beta^\top \Sigma_{Y^c} \beta. \quad (3.84)$$

Whether or not this constitutes an improvement (in the sense that $\text{Var}(X) < \text{Var}(Y)$) is largely a matter of what β is chosen to be. We have the following easily proven lemma.

Lemma 3.4.1. *The function $\text{Var}(X) = \text{Var}(Y) - 2\beta^\top \Sigma_{Y,Y^c} + \beta^\top \Sigma_{Y^c} \beta$ is minimized at*

$$\beta^* = \Sigma_{Y^c}^{-1} \Sigma_{Y,Y^c}$$

²²Other methods include *moment matching* and *importance sampling*. We shall cover the latter strategy shortly; the former is discussed in Boyle et al. [1997], where it is concluded that control variates are superior, at least asymptotically.

with minimum value

$$\min_{\beta} \text{Var}(X) = (1 - R^2)\text{Var}(Y), \quad R^2 \triangleq \frac{\Sigma_{Y,Y^c}^\top \Sigma_{Y^c}^{-1} \Sigma_{Y,Y^c}}{\text{Var}(Y)} \geq 0. \quad (3.85)$$

In the lemma, we recognize the scalar R^2 as the R -squared of a multi-dimensional regression of Y against Y^c . Similarly, the components of the optimal vector β^* are the regression coefficients (the slopes) on the vector Y^c . In practice, we may not know Σ_{Y^c} and Σ_{Y,Y^c} explicitly, in which case we simply replace these with empirical estimates, as obtained by an n -sample Monte Carlo trial. We note that if the random samples used to estimate β^* are the same as those used to estimate $E(X)$, a small bias is typically introduced. This can be circumvented by using separate random numbers for the estimates of β^* and $E(X)$, but in practice this is rarely worth the effort. Nelson [1990], among others, analyzes this issue in more detail.

While the usage of control variates will always lower variance (unless Y and Y^c are perfectly uncorrelated), an improvement of computational efficiency over standard Monte Carlo is, of course, not guaranteed. Consider, for instance, the case where the computational effort involved in generating a single sample of X is $q + 1$ times that of generating Y itself. This will be the case, if i) the effort of drawing random numbers is small relative to computing Y itself; and ii) each of the components of Y^q take about the same time to compute as Y . According to the result in Section 3.4.1, for this special case the control variate method will only entail an increase in efficiency, if

$$(1 - R^2)\text{Var}(Y)(q + 1) < \text{Var}(Y)$$

or

$$1 - R^2 < \frac{1}{q + 1}.$$

As q grows large, this requirement obviously becomes increasingly difficult to satisfy. Rather than indiscriminately adding multiple controls, it is therefore normally best to properly analyze a given problem and use only a few well-chosen variables with strong (negative or positive) correlation to the variable in question.

3.4.3.2 Non-Linear Controls

Our discussion of the control variate method has so far only considered linear controls (3.83), where the modified estimator involves a linear combination of control variates. The resulting estimate of $E(Y)$ are n -point sample averages of the type

$$\bar{Y}_n - \beta^\top (\bar{Y}_n^c - \mu^c).$$

A more general formulation than (3.83) approximates $E(Y)$ with

$$f(\bar{Y}_n^c, \bar{Y}_n) \quad (3.86)$$

for some function f satisfying

$$f(\mu^c, y) = y. \quad (3.87)$$

The requirement (3.87) ensures that $f(\bar{Y}_n^c, \bar{Y}_n)$ approaches $E(Y)$ in the large-sample limit; unlike the regular control variate formulation, however, (3.86) may involve a bias for finite sample sizes.

If f is smooth, a result by Glynn and Whitt [1989] demonstrates that for sufficiently large samples, any non-linear control variate estimator of the type (3.86) is equivalent to an ordinary linear control variate estimator. Still, there may be situations where a non-linear control variate estimator is appropriate, either because i) the sample size is not large enough to justify the result in Glynn and Whitt [1989]; or ii) because the “effective” β weighting of \bar{Y}_n^c implied by f is close to optimal, allowing us to skip the estimation of β^* .

To give an example of non-linear control variates, let us consider the “delta” method of Clewlow and Carverhill [1994]. To state the basic idea, consider the estimate of

$$V(0) = E(g(X(T))),$$

where $X(t)$ is a p -dimensional vector process and $g : \mathbb{R}^p \rightarrow \mathbb{R}$ is a smooth function. Assume that all components of X are martingales, as is the case when X represents assets deflated by a numeraire. We recall from Section 1.7 that, under certain regularity conditions, we have

$$V(T) = V(0) + \int_0^T \sum_{i=1}^p V_{x_i}(t) dX_i(t),$$

where we use the notation $V_{x_i}(t)$ from Section 1.7 to denote, informally, $V_{x_i}(t) = \partial V(t)/\partial X_i(t)$. On a simulation time line $\{t_j\}_{j=1}^m$, we can write, in the style of an Euler scheme,

$$V(T) \approx V(0) + \sum_{j=1}^m \sum_{i=1}^p V_{x_i}(t_{j-1}) (X_i(t_j) - X_i(t_{j-1})).$$

As the zero-mean quantity

$$\sum_{j=1}^m \sum_{i=1}^p V_{x_i}(t_{j-1}) (X_i(t_j) - X_i(t_{j-1}))$$

is likely to have high correlation to $V(T)$, we can consider using it as a control variate. One obstacle is the fact that the derivatives $V_{x_i}(t)$ are likely

to be unknown (as the function $V(t)$ is unknown). Often, however, we can provide an inspired guess for these derivatives, based on perhaps a simpler model or on regression information. The former idea is outlined in Clewlow and Carverhill [1994], and the latter shall be discussed further in Chapter 25. In any case, the resulting scheme ends up effectively using the increments $X_i(t_j) - X_i(t_{j-1})$ as controls, with non-constant weights $V_{x_i}(t_{j-1})$ being functions of the X_i themselves.

3.4.4 Importance Sampling

3.4.4.1 Basic Idea

The basic idea of the *importance sampling method* is to use a measure shift to reduce variance. For a given measure P , consider estimating

$$\mu = E^P(Y), \quad (3.88)$$

where Y is a scalar random variable. Let \widehat{P} be a measure equivalent to P . From the Radon-Nikodym theorem in Chapter 1, we have

$$\mu = E^{\widehat{P}}(Y/R), \quad (3.89)$$

where R is the Radon-Nikodym derivative

$$R = d\widehat{P}/dP, \quad E^P(R) = 1.$$

While (3.88) and (3.89) are both valid expressions for μ , it is possible that the variance of Y/R under measure \widehat{P} is lower than the variance of Y under P , making (3.89) potentially more efficient for Monte Carlo purposes. As an extreme case, consider setting (assuming $Y > 0$ a.s.)

$$R = \frac{Y}{E^P(Y)} = \frac{Y}{\mu}. \quad (3.90)$$

In this case

$$Y/R = \mu$$

and non-random, implying that the measure shift from P to \widehat{P} has removed *all* variance. The problem with the “perfect” choice (3.90) is obviously that we do not know μ — if we did, there would be no need to estimate it by Monte Carlo methods. Nevertheless, we may be able to provide a good guess for μ , allowing us to use (3.90) in an approximate sense.

3.4.4.2 Density Formulation

Importance sampling methods are often most conveniently (and most intuitively) treated in terms of probability densities, so let us cast the description

of Section 3.4.4.1 in such terms. Specifically, let us assume that Y can be represented as $g(X)$, where $g : \mathbb{R}^p \rightarrow \mathbb{R}$ is a well-behaved function and X is p -dimensional with probability density $f : \mathbb{R}^p \rightarrow \mathbb{R}$. We then write

$$\mu = \mathbb{E}^P(g(X)) = \int_{\mathbb{R}^p} g(x) f(x) dx,$$

to which corresponds a regular Monte Carlo estimator

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n g(X_i),$$

where the X_i are independent samples of X , drawn from the density f . Let $h : \mathbb{R}^p \rightarrow \mathbb{R}$ be another density, satisfying the continuity requirement that $h(x) > 0$ whenever $f(x) > 0$. We can then also represent μ as

$$\mu = \int_{\mathbb{R}^p} g(x) \frac{f(x)}{h(x)} h(x) dx,$$

which we can interpret as

$$\mu = \mathbb{E}^{\widehat{P}} \left(g(X) \frac{f(X)}{h(X)} \right),$$

where \widehat{P} is a measure under which X has density $h(x)$. Comparison to the results above identifies the so-called *likelihood ratio* $l(x) = f(x)/h(x)$ as the Radon-Nikodym derivative $dP/d\widehat{P}$ (or $1/R$) governing the shift from P to \widehat{P} . If now X_1, \dots, X_n are independent draws from h (and *not* f), the importance sampling Monte Carlo estimator for μ takes the form

$$\bar{\mu}_n^h = \frac{1}{n} \sum_{i=1}^n g(X_i) \frac{f(X_i)}{h(X_i)}.$$

Let us investigate under which circumstances importance sampling will lead to an improvement in variance. We have

$$\begin{aligned} \text{Var}(\bar{\mu}_n^h) &= \frac{1}{n} \left[\mathbb{E}^{\widehat{P}} \left(g(X)^2 \frac{f(X)^2}{h(X)^2} \right) - \mu^2 \right] \\ &= \frac{1}{n} \left[\mathbb{E}^P \left(g(X)^2 \frac{f(X)}{h(X)} \right) - \mu^2 \right], \end{aligned}$$

and

$$\text{Var}(\bar{\mu}_n) = \frac{1}{n} \left[\mathbb{E}^P \left(g(X)^2 \right) - \mu^2 \right].$$

Hence, importance sampling will lower variance, provided that

$$\mathbb{E}^P \left(g(X)^2 \frac{f(X)}{h(X)} \right) < \mathbb{E}^P \left(g(X)^2 \right).$$

Choosing the importance sampling density $h(x)$ wisely is key to the efficiency of the importance sampling. As an extreme, suppose we could set

$$h(x) = Cf(x)g(x), \quad (3.91)$$

where the constant C is dictated by the need for $h(x)$ to integrate to 1:

$$C^{-1} = \int_{\mathbb{R}^p} g(x)f(x) dx = \mu.$$

In this case,

$$\mathbb{E}^P \left(g(X)^2 \frac{f(X)}{h(X)} \right) = C^{-1} \mathbb{E}^P (g(X)) = \mu^2$$

and

$$\text{Var} (\bar{\mu}_n^h) = 0.$$

This replicates a similar argument in Section 3.4.4.1 (see equation (3.90)), and is equally useless in practice: to compute (3.91) we need to normalize by the constant $1/\mu$, where μ is the quantity that we are trying to estimate in the first place. Nevertheless, (3.91) provides some useful practical guidance: a good choice of likelihood density will sample in proportion to f and g . That is, values of X where both the density $f(X)$ and the payout $g(X)$ are high should be assigned a high value of $h(X)$ (high “importance”), and values of X where either $f(X)$ or $g(X)$ (or both) are low should be assigned a low value of $h(X)$ (low “importance”). This rule is often particularly easy and efficient to apply to situations where $g(X)$ is significant only for a set $X \in \mathcal{A}$, where $P(X \in \mathcal{A})$ is small. Such rare-event problems are a classical application of importance sampling; we give a simple example in Section 3.4.4.5. Related applications to barrier options can be found later, in Chapter 25, with more such examples in Boyle et al. [1997].

3.4.4.3 Importance Sampling and SDEs

Consider now a dynamic setting where we are given a P-measure SDE

$$dX(t) = \mu(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad (3.92)$$

where X is p -dimensional and W is d -dimensional. We wish to evaluate

$$\mathbb{E}^P (g(X(T)))$$

for a real-valued function g . To shift measure, we introduce the density process

$$d\varsigma(t) = -\varsigma(t)\theta(t)^\top dW(t), \quad \varsigma(0) = 1, \quad (3.93)$$

for some adapted d -dimensional process $\theta(t)$, sufficiently regular to make $\varsigma(\cdot)$ a martingale (see Chapter 1). Let

$$\varsigma(t) = E_t^P \left(\frac{d\hat{P}}{dP} \right),$$

for a new measure \hat{P} . By the Girsanov theorem, under \hat{P} ,

$$dX(t) = [\mu(t, X(t)) - \sigma(t, X(t))\theta(t)] dt + \sigma(t, X(t)) d\hat{W}(t), \quad (3.94)$$

where \hat{W} is a Brownian motion in \hat{P} . Also, by the Radon-Nikodym theorem,

$$E^P(g(X(T))) = E^{\hat{P}} \left(\frac{g(X(T))}{\varsigma(T)} \right). \quad (3.95)$$

In a Monte Carlo setting, rather than simulating (3.92) (using methods from Section 3.2) and computing the sample mean of $g(X(T))$, we can instead jointly simulate (3.93) and (3.94) and compute the sample mean of $g(X(T))/\varsigma(T)$. The validity of this approach is independent of the choice of θ in (3.93), and we can use θ as a parameter to minimize the variance of $g(X(T))/\varsigma(T)$ under \hat{P} .

To find the optimal choice for θ , define

$$u(t, X(t)) = E_t^P(g(X(T))), \quad t \leq T,$$

and consider setting

$$\varsigma(t)u(0, X(0)) = u(t, X(t)).$$

By Ito's lemma,

$$d\varsigma(t) = -\varsigma(t)\theta(t)^\top dW(t), \quad \varsigma(0) = 1,$$

where

$$\theta(t) = -u(t, X(t))^{-1} \sigma(t, X(t))^\top \frac{\partial u(t, X(t))}{\partial x}, \quad (3.96)$$

with $\partial u(t, X(t))/\partial x$ being a p -dimensional vector of partial derivatives $\{\partial u(t, X(t))/\partial x_i\}$. The choice for θ in (3.96) is optimal as we have

$$g(X(T))/\varsigma(T) = u(0, X(0)) = E^P(g(X(T))),$$

which is non-random with zero variance. As in earlier examples, the optimal choice for $\theta(t)$ cannot be applied directly as it requires knowledge of $E_t^P(g(X(T)))$ for all t , knowledge which we never possess in practice. In many applications, however, we can often make an educated guess for u , based perhaps on either a simpler SDE than (3.92) or on a simpler payout function than g . We shall see an example of this in Chapter 25; another application can be found in Schoenmakers and Heemink [1997].

3.4.4.4 More on SDE Path Simulation

Let us consider an alternative point of view about SDE simulations, where we assume that the SDE (3.92) is simulated by an m -dimensional Euler scheme (or similar), such that we can write (see also Section 3.4.2.1) for some function $G : \mathbb{R}^{p \times m} \rightarrow \mathbb{R}$,

$$g(X(T)) = G(Z_1, \dots, Z_m),$$

where the Z_i are independent p -dimensional Gaussian vectors. With the Gaussian density of Z_i being denoted $\phi(z)$, $z \in \mathbb{R}^p$, the independence of the Z_i 's allows us to write

$$\mathbb{E}^P(g(X(T))) = \int_{\mathbb{R}^{p \times m}} G(z_1, \dots, z_m) \prod_{i=1}^m \phi(z_i) dz, \quad z \triangleq (z_1, \dots, z_m).$$

If we apply a change of measure that preserves independence of Z_i but alters the common marginal density from $\phi(z)$ to $h(z)$, the likelihood ratio is easily seen to be

$$l(z) = \prod_{i=1}^m \frac{\phi(z_i)}{h(z_i)},$$

such that

$$\mathbb{E}^P(g(X(T))) = \mathbb{E}^{\widehat{P}} \left(G(Z_1, \dots, Z_m) \prod_{i=1}^m \frac{\phi(Z_i)}{h(Z_i)} \right).$$

It is understood that the Z_i used to advance the SDE simulation under \widehat{P} are drawn from the density h , rather than ϕ .

To give a concrete example of a measure shift, assume for simplicity that $p = 1$ and consider shifting the means of the Z_i from zero to some scalar²³ μ , but retaining unit variance. For this, we must set

$$h(z_i) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}(z_i - \mu)^2 \right),$$

whereby

$$l(z) = l(z; \mu) = \exp \left(-\mu \sum_{i=1}^m z_i + \frac{m}{2} \mu^2 \right). \quad (3.97)$$

Here μ is a free variable, which can be set to minimize the variance of the term

$$G(Z)l(Z; \mu), \quad Z \triangleq (Z_1, \dots, Z_m)$$

²³It is also straightforward to introduce a measure shift that moves the means of the Z_i to *different* means μ_i , $i = 1, \dots, m$.

under \widehat{P} . Sometimes this minimization problem can be handled analytically (see Section 3.4.4.5), but most often numerical methods are required. Examples of how to perform this minimization by Monte Carlo simulation can be found in, for example, Su and Fu [2002] and Capriotti [2007]. The approach in Capriotti [2007] (called *least-squares importance sampling*) is particularly straightforward, as the optimization problem is here cast as a least-squares regression problem for which well-known numerical schemes exist such as, e.g., the Levenberg-Marquardt routine in Press et al. [1992]. Both Su and Fu [2002] and Capriotti [2007] point out that, when computing variance, it is advantageous to cast the problem back into the original probability measure P by using

$$E^{\widehat{P}}(G(Z)^2 l(Z; \mu)^2) = E^P(G(Z)^2 l(Z; \mu)).$$

Let us finally note that the measure transformation employed above is a special case of so-called *exponential twisting* (also known as *Esscher transform*), under which a density $f(x)$, $x \in \mathbb{R}$, is transformed into

$$f_\theta(x) = e^{\theta x - \gamma(\theta)} f(x),$$

where θ is a twisting parameter and γ is the *cumulant-generating function*

$$\gamma(\theta) = \ln \left(\int_{\mathbb{R}} e^{\theta x} dx \right).$$

For a standard Gaussian variable, $\gamma(\theta) = \theta^2/2$, demonstrating that the shift of mean employed above is indeed a special case of exponential twisting. We notice that exponential twisting is often a very convenient starting point when working with parametric families of Radon-Nikodym derivatives.

3.4.4.5 Rare Event Simulation and Linearization

For illustrative purposes, consider finally the problem of estimating by Monte Carlo

$$P(Z > c), \quad (3.98)$$

where $Z \sim \mathcal{N}(0, 1)$ is standard Gaussian under the measure P , and c is a big number. In ordinary Monte Carlo, we write

$$P(Z > c) = E^P(1_{\{Z>c\}})$$

and use the sample mean estimator

$$P(Z > c) \approx \frac{1}{n} \sum_{i=1}^n 1_{\{Z_i > c\}},$$

where Z_1, \dots, Z_n are independent standard Gaussian samples. We notice that

$$\begin{aligned}\text{Var}^P(1_{\{Z>c\}}) &= E^P((1_{\{Z>c\}})^2) - E^P(1_{\{Z>c\}})^2 \\ &= E^P(1_{\{Z>c\}}) - E^P(1_{\{Z>c\}})^2 \\ &= P(Z > c)(1 - P(Z > c)),\end{aligned}$$

with sample mean estimator variance being n times smaller. Consider now introducing a probability measure that shifts the mean of Z from 0 to μ . The likelihood ratio is seen from (3.97) to be

$$l(z) = e^{-\mu z + \mu^2/2},$$

such that

$$P(Z > c) = E^{\widehat{P}}(e^{-\mu Z + \mu^2/2} 1_{\{Z>c\}}),$$

where $Z \sim \mathcal{N}(\mu, 1)$ in the measure \widehat{P} . A Monte Carlo estimator for this is then

$$\frac{1}{n} \sum_{i=1}^n e^{-\mu(Z_i + \mu) + \mu^2/2} 1_{\{Z_i + \mu > c\}},$$

where Z_1, \dots, Z_n are again independent standard Gaussian samples. Notice that we have added to the Z_i the mean μ to reflect the shift of measure from P to \widehat{P} . As for variance, we have

$$\begin{aligned}\text{Var}^{\widehat{P}}(e^{-\mu Z + \mu^2/2} 1_{\{Z>c\}}) &= E^{\widehat{P}}(e^{-2\mu Z + \mu^2} (1_{\{Z>c\}})^2) - P(Z > c)^2 \\ &= E^{\widehat{P}}(e^{-2\mu Z + \mu^2} 1_{\{Z>c\}}) - P(Z > c)^2 \\ &= E^P(e^{-\mu Z + \mu^2/2} 1_{\{Z>c\}}) - P(Z > c)^2 \\ &= e^{\mu^2} P(Z > c + \mu) - P(Z > c)^2,\end{aligned}\tag{3.99}$$

where the last equation follows from the properties of the standard Gaussian density. The choice of μ that minimizes the variance under \widehat{P} is the solution to

$$\min_{\mu} e^{\mu^2} P(Z > c + \mu).$$

Differentiating with respect to μ and setting the resulting expression to zero shows that the variance is minimized at μ^* , where

$$2\mu^* [1 - \Phi(c + \mu^*)] - \phi(c + \mu^*) = 0,\tag{3.100}$$

with Φ and ϕ being the standard Gaussian distribution function and density, respectively. This expression can be solved for μ^* with the aid of a numerical root solver. Alternatively, we can use the fact that c is large to rely on the asymptotic approximation

$$1 - \Phi(c + \mu^*) \approx \frac{\phi(c + \mu^*)}{c + \mu^*},$$

which leads to

$$2\mu^* \frac{\phi(c + \mu^*)}{c + \mu^*} \approx \phi(c + \mu^*) \Rightarrow \mu^* \approx c. \quad (3.101)$$

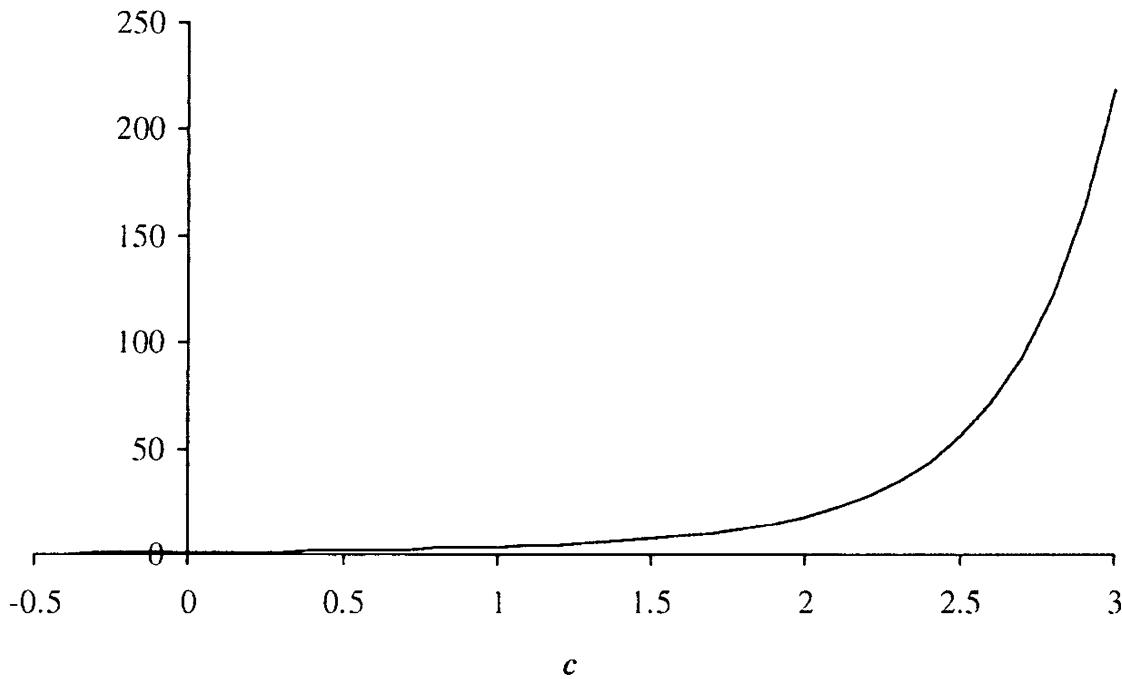
Note that this implies that the probability of Z exceeding c in measure \widehat{P} is approximately $\frac{1}{2}$. This is an intuitive result²⁴, consistent with the discussion at the end of Section 3.4.4.2.

To measure the efficacy of importance sampling, we can use (3.99) to define a variance efficiency ratio as

$$\frac{P(Z > c)(1 - P(Z > c))}{e^{\mu^2} P(Z > c + \mu) - P(Z > c)^2}. \quad (3.102)$$

Figure 3.1 graphs this ratio when μ is set to c , as prescribed in (3.101). For large c , the improvements to variance associated with using importance sampling can be seen to be extremely significant.

Fig. 3.1. Variance Ratio



Notes: The figure graphs the ratio (3.102), with μ set according to (3.101).

It is also illustrative to consider the multi-variate extension to the problem above. Here, we are interested in estimating

$$E^P(1_{\{X>c\}}),$$

²⁴For a somewhat more accurate approximation to μ^* , see Jäckel [2004].

where c is a p -dimensional constant and X is a p -dimensional vector of Gaussian random variables with mean 0 and covariance matrix Σ . Let C be the Cholesky decomposition of Σ , such that

$$\mathbb{E}^P(1_{\{X>c\}}) = \mathbb{E}^P(1_{\{CZ>c\}}) = \mathbb{E}^P(1_{\{Z>c'\}}), \quad c' \triangleq C^{-1}c,$$

where Z is a p -dimensional vector of independent standard Gaussian variables. Let us introduce a measure \widehat{P} where the mean of Z has been shifted to μ , a p -dimensional vector. Following the same steps as for the univariate case, we have

$$\mathbb{E}^P(1_{\{Z>c'\}}) = \mathbb{E}^{\widehat{P}}\left(\exp\left(-\mu^\top Z + \frac{1}{2}\mu^\top\mu\right)1_{\{Z>c'\}}\right),$$

with variance

$$\begin{aligned} \text{Var}^{\widehat{P}}\left(\exp\left(-\mu^\top Z + \frac{1}{2}\mu^\top\mu\right)1_{\{Z>c'\}}\right) \\ = e^{\mu^\top\mu}\mathbb{E}^P(1_{\{Z>c'+\mu\}}) - P(X > c)^2. \end{aligned}$$

A direct optimization of this expression in μ involves multi-dimensional Gaussian integrals, so we wish to resort to approximations. We can use the arguments of Section 3.4.4.2 to argue that the optimal importance sampling density should be proportional to

$$1_{\{z>c'\}}\exp\left(-\frac{1}{2}z^\top z\right), \quad (3.103)$$

since $\exp(-z^\top z/2)$ is proportional to the normal density. Following the idea in Glasserman et al. [1999], we can choose μ such that the location of the peak of an $\mathcal{N}(\mu, I)$ distribution coincides with the peak of (3.103). In other words, we approximate the optimal μ as the value μ^* of z that solves

$$\max_z \left\{ 1_{\{z>c'\}} \exp\left(-\frac{1}{2}z^\top z\right) \right\} = \min_{z>c'} \{z^\top z\}. \quad (3.104)$$

If we assume, say, that all components of c' are larger than 0, then obviously

$$\mu^* = c' = C^{-1}c,$$

consistent with the approximative univariate result (3.101).

We note that the idea behind (3.104) is not limited to situations where we evaluate expectations of an indicator function. For instance, suppose we, as in Section 3.4.4.4, wish to estimate

$$\mathbb{E}^P(G(Z)),$$

for a smooth function $G : \mathbb{R}^p \rightarrow \mathbb{R}$. Restricting ourselves again to the class of measure shifts that only move the mean of Z , the approximately optimal mean shift μ solves

$$\max_z \left\{ G(z) \exp \left(-\frac{1}{2} z^\top z \right) \right\}$$

or, if $G(Z)$ is strictly positive,

$$\max_z \left\{ w(z) - \frac{1}{2} z^\top z \right\}, \quad w(z) \triangleq \ln(G(z)).$$

The first-order condition for the optimum is

$$\nabla w(\mu^*) = (\mu^*)^\top, \quad (3.105)$$

where ∇ is the gradient operator, $\nabla = (\partial/\partial z_1, \dots, \partial/\partial z_p)$ (row vector). This is a fixed-point condition that can be solved by numerical methods. The result (3.105) is exact if w is linear in its argument; the method above can thus be seen as a linearization through a first-order Taylor approximation. Glasserman et al. [1999] demonstrate that, under some conditions, (3.105) satisfies a certain asymptotic optimality property.

3.5 Some Notes on Bermudan Security Pricing

As alluded to in the beginning of this chapter, one drawback of Monte Carlo methods is the difficulty associated with the pricing of securities with early exercise rights. We demonstrated earlier in the chapter that Monte Carlo path generation runs *forward* in time, making direct application of dynamic programming and backward induction (see Chapters 1 and 2) impossible. Indeed, until the early 1990's, it was generally believed that Monte Carlo techniques were inherently incompatible with the pricing of early exercise rights. In the last decade, however, this belief has been overturned, with the advent of several different techniques for Monte Carlo pricing of options with early exercise rights. Most of these techniques are rather advanced and a detailed description will be postponed until later in this book, when the interest rate modeling foundation has been properly laid and the details of callable interest rate securities have been covered. For now, we only provide a brief discussion of certain generic principles, with additional details to be filled in later, in Chapters 18 and 19, among others. We start by establishing some notation and reminding the reader of some basic results from Chapter 1.

3.5.1 Basic Idea

For the remainder of this section, we consider the pricing of a Bermudan security C , with a payout function²⁵ $U(t) = U(t, x(t))$, where $x(t)$ is a

²⁵For many exotic interest rate options, the function $U(t, x(t))$ may actually not be known in closed form. We deal with this complication in Chapter 18.

p -dimensional vector of Markovian state variables. The allowed discrete set of exercise dates is denoted $\mathcal{D} = \{T_1, T_2, \dots, T_B\}$, with $T_B = T$ being the terminal maturity of C . We fix a numeraire N , assumed to be a function of $x(t)$, $N(t) = N(t, x(t))$. From Section 1.10, we recall that

$$C(0) = N(0) \sup_{\tau \in \mathcal{T}} \mathbb{E}^N \left(\frac{U(\tau)}{N(\tau)} \right),$$

where \mathbb{E}^N denotes expectation in the measure Q^N induced by the numeraire N , and \mathcal{T} is the set of stopping time strategies taking values in \mathcal{D} . More generally, we write

$$C(t) = N(t) \sup_{\tau \in \mathcal{T}(t)} \mathbb{E}_t^N \left(\frac{U(\tau)}{N(\tau)} \right), \quad (3.106)$$

where $\mathcal{T}(t)$ is the set of stopping time strategies in \mathcal{D} for which $\tau \geq t$. We also recall that when $t \in (T_{i-1}, T_i]$, we have

$$C(t) = N(t) \mathbb{E}_t^N \left(N(T_i)^{-1} \max(H_i(T_i), U(T_i)) \right), \quad (3.107)$$

where the hold value (see Section 1.10) $H_i(T_i)$ is defined as

$$H_i(T_i) = N(T_i) \mathbb{E}_{T_i}^N \left(N(T_{i+1})^{-1} C(T_{i+1}) \right).$$

Notice that (3.107) establishes that the optimal exercise strategy, as seen from time t , is

$$\tau^* = \inf \{T_i \geq t : U(T_i) \geq H_i(T_i)\}. \quad (3.108)$$

3.5.2 Parametric Lower Bound Methods

Assuming that we are able to simulate the p -dimensional vector $x(t)$ through time, it follows from (3.106) that a lower bound for $C(0)$ can be computed by Monte Carlo method through any exogenous guess for the optimal exercise strategy τ^* . One fairly intuitive approach involves a user-supplied specification of a parametric stopping rule, $\tau(\alpha) \in \mathcal{T}$, where $\alpha \in A \subset \mathbb{R}^m$ is an m -dimensional parameter vector. Defining $x = (x(T_1), \dots, x(T_B))$, for a given value of α , we have the following algorithm.

1. Generate n independent paths $x^{(k)}$, $k = 1, \dots, n$. For path k , let $\tau^{(k)}(\alpha)$ be the exercise date suggested by the parametric stopping rule.
2. For each path k , set $U^{(k)} = U(\tau^{(k)}(\alpha), x(\tau^{(k)}(\alpha)))$ and $C_\alpha^{(k)} = N(\tau^{(k)}(\alpha), x(\tau^{(k)}(\alpha)))^{-1} U^{(k)}$.
3. Return $\bar{C}_\alpha(0) = N(0)n^{-1} \sum_{k=1}^n C_\alpha^{(k)}$ as our estimate for the Bermudan option value $C(0)$.

Let $C_\alpha(0) \triangleq \mathbb{E}^N(\bar{C}_\alpha(0))$. As $\tau(\alpha)$ in general will be sub-optimal, it is clear that

$$C_\alpha(0) \leq C(0). \quad (3.109)$$

To get as close as possible to $C(0)$, it is, of course, preferable to use the value $\alpha^* \in A$ for which $C_\alpha(0)$ is optimized, i.e.

$$\alpha^* = \operatorname{argsup}_{\alpha \in A} C_\alpha(0).$$

A tempting way of estimating $C_{\alpha^*}(0)$ would be to modify step 3 in the algorithm above to

3a. Return $\bar{C}_{\alpha^*}(0) = \sup_{\alpha \in A} \bar{C}_\alpha(0)$.

Leaving aside the question of how one might execute the optimization in Step 3a, we notice that the estimator $\bar{C}_{\alpha^*}(0)$ is biased high relative to $C_{\alpha^*}(0)$:

$$\mathbb{E}^N(\bar{C}_{\alpha^*}(0)) \geq \sup_{\alpha \in A} C_\alpha(0). \quad (3.110)$$

This inequality states that the expected value of the maximum over α must be at least as large as the maximum over α of expected values, a consequence of Jensen's inequality. We may interpret the bias of $\mathbb{E}^N(\bar{C}_{\alpha^*}(0))$ as a *perfect foresight bias*: by using in-sample information to estimate α^* , we effectively "cheat" by making the optimum specific to the same n samples that are also used to determine the option value.

The combination of inequalities (3.109) and (3.110) shows that the quantity $\bar{C}_{\alpha^*}(0)$ from Step 3a has an *indeterminate* bias relative to the true option price. As a bias is generally inevitable when using parametric exercise strategies, in practice it is preferable to at least know its sign. To accomplish this, we can retain the estimated value of α^* found as described above, but draw a *separate* set of Monte Carlo paths when pricing the option. That is, we replace Step 3a with the following two steps:

3b. Set $\hat{\alpha}^* = \operatorname{argsup}_{\alpha \in A} \bar{C}_\alpha(0)$.

4b. Draw a fresh set of n_2 independent paths for x and N , with α locked at the value $\hat{\alpha}^*$. Return $\bar{C}_{\hat{\alpha}^*}(0) = N(0)/n_2 \sum_{k=1}^{n_2} C_{\hat{\alpha}^*}^{(k)}$, where the $C_{\hat{\alpha}^*}^{(k)}$ are computed on the new set of paths.

As the parameter $\hat{\alpha}^*$ will a.s. never equal α^* , it follows that

$$\mathbb{E}^N(\bar{C}_{\hat{\alpha}^*}(0)) \leq C_{\alpha^*}(0) \leq C(0),$$

i.e. we are now assured that $\bar{C}_{\hat{\alpha}^*}(0)$ is low-bound estimator.

3.5.3 Parametric Lower Bound: An Example

What constitutes a good parametric exercise rule is strongly instrument-specific and typically requires case-by-case analysis. Even for simple Markov models and standard option payouts, the topology of exercise and continuation regions can be highly complicated (see e.g. Broadie and Detemple [1997]), so this exercise is by no means straightforward. As a first approximation, however, one can always attempt to use a simple rule based on outright “moneyness” of the underlying option payout, as in Andersen [2000a]. According to this rule, one sets $\alpha = (h_1, h_2, \dots, h_B)$ (i.e. $m = B$) and writes

$$\tau(\alpha) = \inf \{T_i : U(T_i, x(T_i)) > h_i\}. \quad (3.111)$$

That is, exercise of the option takes place when it is sufficiently deep in the money, with the term “sufficiently deep” quantified through unknown trigger thresholds $h_i \geq 0$, $i = 1, \dots, B$.

While α is B -dimensional, finding its optimal value is not truly a B -dimensional optimization problem. Rather, due to the Markov assumption on $x(t)$, we may decompose it into a series of $B - 1$ *one-dimensional* optimization problems. Specifically, working backwards in time, suppose that the optimal values of $h_{j+1}, h_{j+2}, \dots, h_B$ are known. We then find the optimal value of h_j , by optimizing on $\bar{C}_\alpha(0)$, but subject to the constraint that exercise is *not* allowed to take place before time j . As $h_{j+1}, h_{j+2}, \dots, h_B$ are assumed known — and h_1, \dots, h_{j-1} do not come into play — the only variable involved in this optimization is h_j . The algorithm starts with $j = B - 1$ and the known²⁶ boundary condition $h_B = 0$.

A couple of comments on the algorithm above are in order. First, we notice that $U(T_i, x(T_i)) > h_i$ can be replaced with any one-parameter boolean function $g(T_i, x(T_i); h_i)$ without affecting the basic algorithm — see Andersen [2000a] for some examples. Second, if in such a boolean function $g(T_i, x(T_i); h_i)$ each h_i is allowed to be q -dimensional, the optimization problem reduces to $B - 1$ q -dimensional optimization problems. And third, for a finite-path simulation, the objective functions in each of the $B - 1$ optimization problems will not be smooth; consequently, the optimization is best performed by an iterative search rather than a derivative-based method. Andersen [2000a] uses the golden section search (see Press et al. [1992]), but simpler strategies based on, say, outright sorting are also possible.

3.5.4 Regression-Based Lower Bound

According to (3.108), an approximation for the optimal exercise strategy can always be constructed through an estimate for hold values $H_i(T_i)$ at all $i = 1, 2, \dots, B - 1$. In our Markov setting, we know that

²⁶At the last possible exercise date, we would, of course, always exercise the option if it is in-the-money.

$$H_i(T_i) = q_i(x(T_i)) = q_i(x_1(T_i), \dots, x_p(T_i))$$

for a set of $B - 1$ functions $q_i : \mathbb{R}^p \rightarrow \mathbb{R}$, $i = 1, 2, \dots, B - 1$; the problem of estimating hold values is equivalent to the problem of estimating the functions q_i .

From Section 3.5.1, we know that

$$q_i(x) = N(T_i, x) E^N(C(T_{i+1})/N(T_{i+1}, x(T_{i+1})) | x(T_i) = x), \quad (3.112)$$

which can be interpreted as the *regression* of $C(T_{i+1})/N(T_{i+1})$ on the Markov state variables $x(T_i)$. Several authors — including Carriére [1996], Longstaff and Schwartz [2001], and Tsitsiklis and Roy [2001] — have used this observation to suggest that $q_i(x)$ be estimated by a linear combination of exogenously specified (basis) functions of $x(T_i)$, with least-squares regression on Monte Carlo paths used to determine the best weights for these functions. That is, we fundamentally assume that

$$q_i(x) = \sum_{j=1}^d \beta_{i,j} \psi_j(x), \quad (3.113)$$

for a set of d basis-functions $\psi_j : \mathbb{R}^p \rightarrow \mathbb{R}$, $j = 1, 2, \dots, d$. Setting $\beta_i = (\beta_{i,1}, \dots, \beta_{i,d})^\top$ and $\psi(x) = (\psi_1(x), \dots, \psi_d(x))^\top$, we can rewrite (3.113) as $q_i(x) = \psi(x)^\top \beta_i$ or, from (3.112),

$$\begin{aligned} E^N(N(T_i, x(T_i)) C(T_{i+1})/N(T_{i+1}, x(T_{i+1})) | x(T_i) = x) \\ = E^N(\psi(x(T_i))^\top | x(T_i) = x) \beta_i. \end{aligned}$$

This, in turn, implies that

$$\Omega_i = \Psi_i \beta_i \Rightarrow \beta_i = \Psi_i^{-1} \Omega_i, \quad (3.114)$$

where Ω_i is the d -dimensional vector

$$\Omega_i = E^N\left(\psi(x(T_i)) \frac{N(T_i, x(T_i))}{N(T_{i+1}, x(T_{i+1}))} C(T_{i+1})\right),$$

and Ψ_i is the $d \times d$ matrix

$$\Psi_i = E^N\left(\psi(x(T_i)) \psi(x(T_i))^\top\right).$$

The rationale for rewriting (3.113) into the seemingly more convoluted representation (3.114) is that the latter leads naturally to the algorithm for a least-squares estimation of β_i : one simply replaces the expectations in Ψ_i and Ω_i with sample averages $\bar{\Psi}_i$ and $\bar{\Omega}_i$ computed on a set of Monte Carlo paths. That is, one uses²⁷

²⁷In practice, a direct solution of linear equations in this fashion can be suboptimal if the matrix $\bar{\Psi}_i$ is ill-conditioned. Instead, one would use either truncated singular value decomposition (TSVD) or Tikhonov regularization to find $\hat{\beta}_i$. We return to this issue in Chapter 18.

$$\widehat{\beta}_i = \overline{\Psi}_i^{-1} \overline{\Omega}_i \quad (3.115)$$

as the sample estimate.

We shall discuss the details of (and many variations on) the regression approach later, in Chapter 18. For now, let us just notice that computation of $\overline{\Omega}_i$ requires estimation of $C(T_{i+1})$, which naturally encourages running the estimation of the $\widehat{\beta}_i$ backwards in i , starting from $i = B - 1$. We also notice that the success of the regression approach depends critically on the choice and number of basis functions ψ_j . We give specific advice on this topic in Chapters 18 and 19.

3.5.5 Upper Bound Methods

Given a martingale M in measure Q^N , we recall from Section 1.10.2 that an M -specific upper bound $C_M(0)$ for a Bermudan option can always be constructed as

$$C_M(0) = N(0) \left\{ M(0) + E^N \left(\max_{t \in \mathcal{D}} \left(\frac{U(t, x(t))}{N(t, x(t))} - M(t) \right) \right) \right\} \geq C(0). \quad (3.116)$$

Let $\mathbf{M} = (M(T_1), \dots, M(T_B))^T$. As long as $M(\cdot)$ can be simulated along with the vector $x(\cdot)$, (3.116) suggests the following Monte Carlo algorithm:

1. Generate n independent paths $x^{(k)}$, $\mathbf{M}^{(k)}$, $k = 1, \dots, n$. For path k , let $\gamma^{(k)}$ be the maximum value of $U(T_i, x(T_i))/N(T_i, x(T_i)) - M(T_i)$ over $i = 1, 2, \dots, B$.
2. Return $\overline{C}_M(0) = N(0)\{M(0) + n^{-1} \sum_{k=1}^n \gamma^{(k)}\}$.

For the upper bound method to be practically useful, we would want the gap $E^N(\overline{C}_M(0)) - C(0)$ to be small. This, in turn, requires that we specify the martingale $M(t)$ to be “reasonable”. More specifically, from results in Section 1.10.2, we would like $M(t)$ to represent the martingale component of a good approximation to the supermartingale $C(t)$. For instance, if we happen to think that $C(t)$ is well-approximated by some known function v of time and x ,

$$C(t) = v(t, x(t)), \quad (3.117)$$

we would set

$$M(t) = \sum_{j=1}^p \frac{\partial v(t, x(t))}{\partial x_j} (dx_j(t) - E_t^N(dx_j(t))). \quad (3.118)$$

As an example, suppose that W^N is a (possibly vector-valued) Brownian motion in Q^N and $dx_j(t) = \mu_j(t, x(t)) dt + \sigma_j(t, x(t)) dW^N(t)$, in which case we could use

$$M(t) = \sum_{j=1}^p \frac{\partial v(t, x(t))}{\partial x_j} \sigma_j(t, x(t)) dW^N(t). \quad (3.119)$$

Occasionally, a natural analytical guess for the function v may exist, but most often the only estimate for v is given only implicitly, through a low-bound estimator based on an approximation of the optimal exercise strategy. A completely generic algorithm to turn a guess for the optimal exercise strategy into a proxy for $M(t)$ is developed in Andersen and Broadie [2004]; we shall discuss this algorithm in detail in Chapter 18. If the evolution of $x(t)$ is described by an SDE, a regression approach to estimation of the terms multiplying $dW^N(t)$ in (3.119) can be found in Belomestny et al. [2007].

3.5.6 Confidence Intervals

Suppose that we simultaneously apply a lower bound and an upper bound method to provide two sample estimates \bar{C}_{lo} and \bar{C}_{up} , with

$$\mathbb{E}^N(\bar{C}_{\text{lo}}) \leq C(0) \leq \mathbb{E}^N(\bar{C}_{\text{up}}). \quad (3.120)$$

Let us assume that the sample standard errors on \bar{C}_{lo} and \bar{C}_{up} have been computed as s_{lo} and s_{up} , respectively. For a sufficiently large number of Monte Carlo trials, we can then use the central limit theorem from Section 3.1 to set up a confidence interval

$$[\bar{C}_{\text{lo}} - u_{\gamma/2} \cdot s_{\text{lo}}, \bar{C}_{\text{up}} + u_{\gamma/2} \cdot s_{\text{up}}],$$

where $\Phi(u_{\gamma/2}) = 1 - \gamma/2$. It is clear from (3.120) that the likelihood of this interval bracketing the true price $C(0)$ is *at least* $1 - \gamma$. It is also clear that this confidence interval will not shrink to zero — even in the limit of an infinite number of samples where $s_{\text{lo}} \rightarrow 0$ and $s_{\text{up}} \rightarrow 0$ — unless \bar{C}_{lo} and \bar{C}_{up} simultaneously achieve the unlikely feat of being perfectly unbiased estimators for $C(0)$.

Finally, let us note that any number inside the interval $[\bar{C}_{\text{lo}}, \bar{C}_{\text{up}}]$ can reasonably be used as an estimator for $C(0)$. To the extent that we have reason to believe²⁸ that \bar{C}_{lo} and \bar{C}_{up} have roughly opposite biases, a natural estimator is $(\bar{C}_{\text{lo}} + \bar{C}_{\text{up}})/2$.

3.5.7 Other Methods

The methods described so far are those that, in our opinion, are most useful for practical Monte Carlo pricing of interest rate options with early exercise rights. Several other methods, however, have been proposed in the literature, some of which have interesting theoretical properties. We highlight the *random tree* method (Broadie and Glasserman [1997]) which builds a random

²⁸There is some evidence that the upper-bound method in Andersen and Broadie [2004] produces a bias that is often roughly opposite of that of the low-bound method from which the martingale M in (3.116) is extracted.

non-recombining lattice by Monte Carlo methods; backward induction arguments are then used to construct high- and low-biased estimators, both of which are convergent to the true price as the number of Monte Carlo paths are increased. The drawback of the method is its computational complexity which increases exponentially in the number of exercise dates (B), ruling out its practical usage for many realistic applications. Broadie and Glasserman [2004] suggest a recombining *stochastic mesh* method that grows only linearly in the number of exercise weights; in its basic form, this method requires explicit knowledge of transition densities as it relies on likelihood ratios to set weights on nodes in the mesh (see Section 3.3.3). As discussed in Glasserman [2004], the concept of stochastic meshes can, however, be broadened to include several other methods, include the regression approach in Section 3.5.4.

3.A Appendix: Constants for Φ^{-1} Algorithm

a_0	2.50662823884	c_0	0.3374754822726147
a_1	-18.61500062529	c_1	0.9761690190917186
a_2	41.39119773534	c_2	0.1607979714918209
a_3	-25.44106049637	c_3	0.0276438810333863
b_0	-8.47351093090	c_4	0.0038405729373609
b_1	23.08336743743	c_5	0.0003951896511919
b_2	-21.06224101826	c_6	0.0000321767881768
b_3	3.13082909833	c_7	0.0000002888167364
		c_8	0.0000003960315187

Fundamentals of Interest Rate Modeling

The purpose of this brief chapter is twofold. First, we introduce notations to characterize prices and yields of basic fixed income market securities. In addition to providing the foundation for a more expansive discussion of fixed income markets (which we shall undertake in Chapter 5), this part of the chapter serves to identify and characterize a number of probability measures that are of fundamental importance in models for the term structure of interest rates. A brief discussion of measures used in a two-currency setting is also provided.

In the second part of the chapter, we discuss general characteristics of models with dynamics driven by vector-valued Brownian motions. This analysis leads to the fundamental class of Heath-Jarrow-Morton (HJM) (see Heath et al. [1992]) models of continuously compounded forward rates. Among other special cases, we discuss in some detail tractable HJM models with Gaussian volatility structure, and provide some results for the case where such models are Markovian. These discussions continue in Chapters 10 through 12 where we consider one- and multi-factor short rate models, and in the Chapter 13 where we introduce the important class of quasi-Gaussian HJM models with local and stochastic volatility.

4.1 Fixed Income Notations

4.1.1 Bonds and Forward Rates

As in earlier chapters, let $P(t, T)$ denote the time t price of a zero-coupon bond (also known as a *discount bond*) delivering for certain \$1 at maturity $T \geq t$. Suppose we are interested in purchasing at some future time T a zero-coupon bond maturing at $T + \tau$, $\tau > 0$. At time $t < T$, the price of such a bond can be locked in by i) purchasing at time t one $(T + \tau)$ -maturity zero-coupon bond; and ii) selling short (“shorting”) $P(t, T + \tau)/P(t, T)$

T -maturity zero-coupon bonds. The time t cost of executing this strategy is zero,

$$-1 \cdot P(t, T + \tau) + P(t, T + \tau)/P(t, T) \cdot P(t, T) = 0,$$

but a flow of

$$-P(t, T + \tau)/P(t, T)$$

will take place at time T as the T -maturity short position matures. This is compensated by an inflow of \$1 at time $T + \tau$. In other words, our trading strategy effectively fixes the time T purchase price of the $(T + \tau)$ -maturity bond at

$$P(t, T, T + \tau) \triangleq P(t, T + \tau)/P(t, T), \quad \tau > 0,$$

a quantity known as the time t *forward price* for the zero-coupon bond spanning $[T, T + \tau]$.

It is often convenient to characterize a forward bond price by a discount rate. One such rate is the *continuously compounded forward yield* $y(t, T, T + \tau)$, defined by

$$e^{-y(t, T, T + \tau)\tau} = P(t, T, T + \tau). \quad (4.1)$$

The time between the maturity of the forward bond and the expiry of the forward contract, i.e. τ , is often called the *tenor* of the forward bond or the forward yield. In the definition of the continuously compounded yield lies an implicit, and idealized, assumption of continuous reinvestment of investment proceeds. Most actual market quotes, however, are based on discrete-time compounding of proceeds. Accordingly, we define a *simple forward rate* $L(t, T, T + \tau)$ as

$$1 + \tau L(t, T, T + \tau) = 1/P(t, T, T + \tau). \quad (4.2)$$

Again, τ is the *tenor* of the forward rate. For an arbitrary set of dates $T = T_0 < T_1 < T_2 < \dots < T_n$, notice that forward bond prices can be recovered from forward rates by simple compounding,

$$P(t, T_n)/P(t, T) = \prod_{i=1}^n \frac{1}{1 + (T_i - T_{i-1}) L(t, T_{i-1}, T_i)}.$$

Unless we state otherwise, throughout this book we shall typically make the assumption that spot rates $L(T, T, T + \tau)$ are the *Libor (London Interbank Offered Rate) rates* quoted in the interbank market. Libor rates are quoted on values of τ ranging from one week ($\tau = 1/52$)¹ to 12 months ($\tau = 1$), and form the basis for a number of floating-rate derivative contracts, such as interest rate swaps and Eurodollar futures. We shall examine these securities in more detail in Chapter 5.

¹Note that in reality the calculation of year fractions τ are governed by fairly complicated market conventions. A brief discussion of this topic can be found in Appendix 5.A.

In the limit $\tau \downarrow 0$,

$$L(t, T, T + \tau) \rightarrow f(t, T),$$

where the quantity $f(t, T)$ is the time t *instantaneous forward rate* to time T . We think of $f(t, T)$ as the forward rate spanning $[T, T + dT]$, observed at time t . The relation between instantaneous forward rates and bond prices is given by the continuous compounding formula

$$P(t, T, T + \tau) = \exp \left(- \int_T^{T+\tau} f(t, u) du \right), \quad (4.3)$$

such that

$$f(t, T) = - \frac{\partial \ln P(t, T)}{\partial T}, \quad (4.4)$$

and, from (4.1),

$$y(t, T, T + \tau) = \tau^{-1} \int_T^{T+\tau} f(t, u) du, \quad f(t, T) = \lim_{\tau \downarrow 0} y(t, T, T + \tau).$$

We also notice the relationship

$$f(t, T) = \frac{\partial (y(t, t, T)(T - t))}{\partial T} = y(t, t, T) + (T - t) \frac{\partial y(t, t, T)}{\partial T}.$$

The quantity

$$r(t) \triangleq f(t, t) \quad (4.5)$$

is an \mathcal{F}_t -measurable random variable known as the *short rate* or sometimes the *spot rate*. Loosely speaking, we can think of $r(t)$ as the overnight rate in effect at time t .

Finally, let us note that interest rates of various flavors and related quantities are typically quoted in percentage points or sometimes in *basis points*, where 1 basis point = 1/100 of one percent.

4.1.2 Futures Rates

Through the market for Eurodollar futures (see Chapter 5), investors can enter into securities that will pay at time T an amount of

$$1 - L(T, T, T + \tau). \quad (4.6)$$

At time 0, a Eurodollar futures contract can be entered into at no upfront cost, but with an implicit obligation of the holder to pay at time T per unit of notional

$$1 - F(0, T, T + \tau)$$

in return for the payout (4.6). Here, $F(t, T, T + \tau)$ is the time t *simple futures rate* for the period $[T, T + \tau]$. Importantly, the futures rate is *marked to market* (or *resettled*) each day, with the day's change in the futures rate immediately credited to or debited from the contract holder's account with the futures exchange. Specifically, after holding the contract for a period of $\Delta = 1$ day, the futures contract holder would thus experience a cash flow of

$$(1 - F(\Delta, T, T + \tau)) - (1 - F(0, T, T + \tau)) \\ = - (F(\Delta, T, T + \tau) - F(0, T, T + \tau)).$$

Continuing the mark-to-market process to maturity shows that the total amount of cash flow received by the holder on $[0, T]$ is

$$- (F(T, T, T + \tau) - F(0, T, T + \tau)) = - (L(T, T, T + \tau) - F(0, T, T + \tau)) \quad (4.7)$$

where we have used the fact that $F(T, T, T + \tau)$ must equal $L(T, T, T + \tau)$ to avoid a *delivery arbitrage*.

The fact that the net cash flow payment (4.7) on a Eurodollar futures contract has been made incrementally on a daily basis has important valuation consequences, and causes the futures rate to differ from the forward rate defined earlier. For instance, under a scenario of rising interest rates, the holder of a Eurodollar futures contract must make payments to the futures exchange. As rates are rising, the contract holder will be faced with a high-rate — and thus unfavorable — borrowing environment for funding these payments. Conversely, when interest rates fall, the reinvestment of received funds will take place at increasingly low rates. Due to the adverse behavior of funding costs and reinvestment gains, we would expect the purchaser of a Eurodollar futures contract to pay less for these instruments than for a comparable instrument without daily mark-to-market. Consequently, we would expect the futures rate to be *above* the corresponding forward rate. We shall quantify this effect in Section 4.5.1 and, with more advanced models, in Section 16.8.

We notice that we can define *instantaneous futures rates* $q(t, T)$ in the same fashion as we defined instantaneous forward rates:

$$q(t, T) = \lim_{\tau \downarrow 0} F(t, T, T + \tau).$$

4.1.3 Annuity Factors and Par Rates

Most fixed income securities involve multiple cash flows taking place on a pre-set schedule of dates, often referred to as a *tenor structure*,

$$0 \leq T_0 < T_1 < \dots < T_N.$$

Given a tenor structure, for any two integers k, m satisfying $0 \leq k < N$, $m > 0$, and $k + m \leq N$, we can define an *annuity factor* $A_{k,m}$ by

$$A_{k,m}(t) = \sum_{n=k}^{k+m-1} P(t, T_{n+1}) \tau_n, \quad \tau_n = T_{n+1} - T_n. \quad (4.8)$$

Annuity factors provide for compact notation when pricing coupon-bearing securities. For instance, a security making m coupon payments of $c\tau_n$ at all $T_{n+1}, n = k, \dots, k + m - 1$, is easily seen to have time t value of

$$cA_{k,m}(t), \quad t \leq T_k.$$

If the security also involves a back-end return of notional at time T_{k+m} (as is the case for a regular coupon-bearing bond), the pricing expression is

$$cA_{k,m}(t) + P(t, T_{k+m}), \quad (4.9)$$

where we assume that the bond has been normalized to have a unit notional. The time t forward price to T_k of the security (4.9) is

$$cA_{k,m}(t)/P(t, T_k) + P(t, T_{k+m})/P(t, T_k);$$

the value of the coupon c for which this expression equals 1 is known as the *forward par rate* or, when used in the context of swap pricing, as the *forward swap rate*. With $S_{k,m}(t)$ denoting the time t swap rate, we apparently have

$$S_{k,m}(t) = \frac{P(t, T_k) - P(t, T_{k+m})}{A_{k,m}(t)}, \quad t \leq T_k. \quad (4.10)$$

From the definition of $L(t, T_n, T_{n+1})$ in (4.2), a little thought shows that the numerator of the expression for $S_{k,m}(t)$ can be expanded into a weighted sum of forward rates, leading to the alternative expression

$$S_{k,m}(t) = \frac{\sum_{n=k}^{k+m-1} \tau_n P(t, T_{n+1}) L_n(t)}{A_{k,m}(t)}, \quad t \leq T_k, \quad (4.11)$$

where we have introduced the useful shorthand

$$L_n(t) \triangleq L(t, T_n, T_{n+1}).$$

It follows that the forward swap rate can be loosely interpreted as a weighted average of simple forward rates on the specified tenor structure. We note for the future that the time T_k is sometimes referred to as the *fixing date*, or *expiry*, of the swap rate $S_{k,m}$, while the length of the corresponding swap, $T_{k+m} - T_k$, is sometimes called the *tenor* of the swap rate.

4.2 Fixed Income Probability Measures

As discussed in Chapter 1, selecting an equivalent martingale measure is largely a matter of choosing a *numeraire*, an asset price process used to

re-normalize the prices of other traded securities. For later reference, this section lists and names a number of important numeraires and measures used in fixed income pricing. Throughout the section, we assume that the market is complete, and we use $V(t)$ to denote the time t price of a derivative security making an \mathcal{F}_T -measurable payment of $V(T)$.

4.2.1 Risk Neutral Measure

The numeraire defining the risk-neutral measure Q is the continuously compounded money market account $\beta(t)$, satisfying the locally deterministic SDE

$$d\beta(t) = r(t)\beta(t) dt, \quad \beta(0) = 1, \quad (4.12)$$

where $r(t)$ is the short rate, $r(t) = f(t, t)$. Solving this equation yields

$$\beta(t) = e^{\int_0^t r(u) du}.$$

From the results of Chapter 1, in the absence of arbitrage the numeraire-deflated process $V(t)/\beta(t)$ must be a martingale, implying the derivative security valuation formula

$$V(t)/\beta(t) = E_t^Q (V(T)/\beta(T)), \quad t \leq T, \quad (4.13)$$

or equivalently

$$V(t) = E_t^Q \left(e^{-\int_t^T r(u) du} V(T) \right). \quad (4.14)$$

If we apply (4.14) to the special case of $V(T) = 1$, we obtain a fundamental bond pricing formula. We highlight the importance of this result by listing it in a lemma.

Lemma 4.2.1. *In the absence of arbitrage, the time t price $P(t, T)$ of a T -maturity zero-coupon bond is*

$$P(t, T) = E_t^Q \left(e^{-\int_t^T r(u) du} \right). \quad (4.15)$$

It follows from Lemma 4.2.1 that specification of the dynamics of $r(t)$ under Q suffices to determine the prices of discount bonds at all times and maturities. Models that are based on such a direct specification of $r(t)$ dynamics are known as *short rate models* and are the subject of Chapters 10 through 12. Notice the resemblance between expressions (4.3) and (4.15); if $r(t)$ is deterministic, the two expressions will agree as $r(u) = f(t, u)$, $u \geq t$. If $r(t)$ is random, one may wonder whether this result will hold in expectation. The answer to this is negative, i.e.

$$f(t, u) \neq E_t^Q (r(u)), \quad (4.16)$$

provided r is random. We prove this in the section below. Under certain idealized conditions, however, equality holds in (4.16) provided $f(t, u)$ is replaced by the *futures* rate $q(t, u)$. The exact result is as follows.

Lemma 4.2.2. Assume that mark-to-market takes place continuously. Under regularity conditions on the short rate $r(\cdot)$ — it suffices that $r(\cdot)$ is positive and bounded — the futures rate $F(\cdot, T, T + \tau)$ is a Q-martingale, and

$$F(t, T, T + \tau) = E_t^Q(L(T, T, T + \tau)). \quad (4.17)$$

Proof. Over a small interval $[t, t + dt]$, we have earlier shown that the cash proceeds from a futures contract are proportional to

$$dF(t, T, T + \tau) = F(t + dt, T, T + \tau) - F(t, T, T + \tau).$$

Suppose that we hold the futures contract up to some arbitrary horizon $t < \mathcal{T} \leq T$ at which point we exit (e.g., by selling the futures contract). Deflating all cash proceeds from this strategy with the numeraire $\beta(t)$ and integrating provides us with the time t value of the futures contract as

$$V_{\text{fut}}(t) = \beta(t) E_t^Q \left(\int_t^{\mathcal{T}} \beta(s)^{-1} dF(s, T, T + \tau) + \beta(\mathcal{T})^{-1} V_{\text{fut}}(\mathcal{T}) \right).$$

As it is always costless to enter into a futures contract, $V_{\text{fut}}(t) = V_{\text{fut}}(\mathcal{T}) = 0$ by definition, so for arbitrary $t < \mathcal{T} \leq T$ we must have (since $\beta(t)$ is positive)

$$E_t^Q \left(\int_t^{\mathcal{T}} \beta(s)^{-1} dF(s, T, T + \tau) \right) = 0. \quad (4.18)$$

Provided that $\beta(s)^{-1}$ is almost surely positive (which is the case if r is bounded), the fact that (4.18) holds for arbitrary horizons $t < \mathcal{T} \leq T$ shows that

$$E_t^Q(dF(s, T, T + \tau)) = 0, \quad t \leq s \leq T,$$

which demonstrates that F is a Q-martingale. The result (4.17) then immediately follows. \square

Equation (4.17) states that the futures rate is the Q-expectation of the time T spot rate $L(T, T, T + \tau)$. A similar relation must then hold for the instantaneous futures rate, i.e.

$$q(t, u) = E_t^Q(r(u)) \quad (4.19)$$

as stated earlier².

Lemma 4.2.2 was first proven by non-probabilistic methods in Cox et al. [1981], who employed a direct, and quite instructive, hedging argument to

²This result should not be confused with the classical *expectations hypothesis* which states that futures (or sometimes forward) rates are unbiased estimators of future spot rates, in the *real-life* probability measure P: $E_t^P(r(T)) = q(t, T)$. The expectations hypothesis amounts to a strong assumption about the market price of risk (see Chapter 1), whereas equation (4.19) is a preference-free arbitrage relationship.

show the result. The assumption of continuous resettlement in the lemma may appear idealized, but the difference between daily and continuous settlement is quite small, as shall be demonstrated in Chapter 16. Explicit modeling of discrete resettlement is nevertheless quite straightforward, and basically involves shifting measure, from the risk-neutral measure to the so-called spot measure, defined below. We return to this issue in Chapter 16.

4.2.2 T -Forward Measure

The T -forward measure Q^T was introduced in Jamshidian [1991b] (see also Geman et al. [1995]), and uses a T -maturity zero-coupon bond as the numeraire asset. As is customary, we let $E^T(\cdot)$ denote expectations in measure Q^T , such that

$$V(t)/P(t, T) = E_t^T(V(T)/P(T, T)), \quad t \leq T.$$

As obviously $P(T, T) = 1$, this expression simplifies to the convenient form

$$V(t) = P(t, T)E_t^T(V(T)). \quad (4.20)$$

Comparison of (4.20) and (4.14) shows that shifting to the T -forward measure in a sense decouples the expectation of the terminal payout $V(T)$ from that of the numeraire. As we shall see, this is often very convenient when we attempt to construct analytical formulas for prices of certain simple interest rate derivatives. From the results of Section 1.3, we note that the explicit connection between the risk-neutral and T -forward measures is given by the density

$$E_t^Q \left(\frac{dQ^T}{dQ} \right) = \frac{P(t, T)/P(0, T)}{\beta(t)}. \quad (4.21)$$

As $P(t, T + \tau)$ is the price of a traded asset, from the definition of the T -forward measure it follows that forward bond prices

$$P(t, T, T + \tau) = P(t, T + \tau)/P(t, T)$$

are martingales in the T -forward measure. We highlight a related result for forward rates below.

Lemma 4.2.3. *In the absence of arbitrage the forward Libor rate $L(t, T, T + \tau)$ is a martingale under $Q^{T+\tau}$, such that*

$$L(t, T, T + \tau) = E_t^{T+\tau}(L(T, T, T + \tau)), \quad t \leq T. \quad (4.22)$$

Proof. By definition (see (4.2))

$$L(t, T, T + \tau) = \tau^{-1} (P(t, T)/P(t, T + \tau) - 1).$$

As $P(t, T)/P(t, T + \tau)$ is a martingale under $Q^{T+\tau}$, so is $L(t, T, T + \tau)$. The result follows. \square

Taking the limit $\tau \downarrow 0$ and setting $T = u$ yields

$$f(t, u) = E_t^u(f(u, u)) = E_t^u(r(u)), \quad (4.23)$$

which should be compared to the result (4.19).

4.2.3 Spot Measure

When working with a multitude of forward rates on a tenor structure $0 = T_0 < T_1 < \dots < T_N$, it is often convenient to introduce a numeraire that can be extended to arbitrary horizons by compounding. While the continuously compounded money market account β would accomplish this, working with a continuously compounded numeraire is inherently awkward in a setting with a discrete tenor structure. As an alternative, we can introduce a discrete-time equivalent of the continuously compounded money market account to be the value of the following trading strategy. At time 0, \$1 is invested in $1/P(0, T_1)$ T_1 -maturity discount bonds, returning the amount

$$1/P(0, T_1) = 1 + \tau_0 L(0, 0, \tau_0)$$

at time T_1 . This amount is then reinvested (“rolled”) at time T_1 in T_2 -maturity bonds, returning

$$1/P(0, T_1) \cdot 1/P(T_1, T_2) = (1 + \tau_0 L(0, 0, T_1))(1 + \tau_1 L(T_1, T_1, T_2))$$

at time T_2 . Repeating this re-investment strategy at each date in the tenor structure gives rise to an asset price process $B(t)$, where $B(0) = 1$ and

$$B(t) = \prod_{n=0}^i (1 + \tau_n L_n(T_n)) P(t, T_{i+1}), \quad T_i < t \leq T_{i+1}, \quad (4.24)$$

where we used the already introduced short-hand notation $L_n(t) = L(t, T_n, T_{n+1})$. The process $B(t)$ is effectively a rolling certificate of deposit, and can be interpreted as a discrete-time equivalent of $\beta(t)$. $B(t)$ will approach $\beta(t)$ as the time spacing of the tenor structure is made increasingly fine.

The measure induced by $B(t)$ is known as the *spot measure* (or sometimes *spot Libor measure*), denoted $E^B(\cdot)$. With $E^B(\cdot)$ denoting expectations in this measure we have

$$V(t) = E_t^B \left(V(T) \frac{B(t)}{B(T)} \right),$$

where

$$\begin{aligned} \frac{B(t)}{B(T)} &= \prod_{n=i+1}^j (1 + \tau_n L_n(T_n))^{-1} \frac{P(t, T_{i+1})}{P(T, T_{j+1})}, \\ T_i < t \leq T_{i+1}, \quad T_j < T \leq T_{j+1}. \end{aligned}$$

The similarity between the discrete and continuous money market accounts makes the spot Libor measure resemble the risk-neutral measure in many ways. For example, as we recall from Lemma 4.2.2, the risk-neutral measure is characterized by the fact that a continuously resettled futures rate is a martingale. In close parallel, the futures rate that is marked to market (resettled) discretely on dates T_0, \dots, T_N turns out to be a martingale in the spot Libor measure. We show this in Section 16.8.

4.2.4 Terminal and Hybrid Measures

One advantage of the spot measure over an arbitrary T -forward measure is the fact that the numeraire asset $B(t)$ will remain alive throughout the span of the tenor structure $\{T_n\}_{n=0}^N$. This property of $B(t)$ is necessary for the valuation of securities which may mature (randomly) at any date in the tenor structure. Securities of this type include for instance barrier options (such as range accruals) and options with early exercise rights (such as Bermudan swaptions). On the other hand, if we pick the T -forward measure corresponding to the *last* maturity in the tenor structure, $T = T_N$, this also yields a numeraire asset — the T_N -maturity zero-coupon bond — that is certain to remain alive at all dates in the tenor structure. The measure induced by $P(t, T_N)$ (Q^{T_N}) is often referred to as the *terminal measure*. For a security V maturing at a date $T \leq T_N$ we get, from the usual martingale pricing formula,

$$V(t) = P(t, T_N) E_t^{T_N} (V(T)/P(T, T_N)), \quad t \leq T \leq T_N. \quad (4.25)$$

In (4.25) it is useful to notice that $V(T)/P(T, T_N)$ is the time T_N proceeds of rolling at time T the security payout $V(T)$ into a zero-coupon bond maturing at time T_N , effectively aligning the maturity of the numeraire and the cash flow date of the underlying asset. As an alternative, $V(T)$ could be rolled into the spot numeraire asset $B(T)$, leading to a T_N payout of $V(T)/B(T) \cdot B(T_N)$. This gives rise to the equivalent formula

$$V(t) = P(t, T_N) E_t^{T_N} (V(T)B(T_N)/B(T)), \quad t \leq T \leq T_N. \quad (4.26)$$

We note that this formula can also be derived from the basic relationship between the measures Q^B and Q^{T_N} by simply noting that

$$P(T, T_N) E_T^{T_N} (B(T_N)/B(T)) = B(T) E_T^B (B(T_N)/B(T)/B(T_N)) = 1,$$

such that, by iterated conditional expectations³,

³The *law of iterated conditional expectations*, sometimes known as the *tower rule*, states that for an \mathcal{F}_T -measurable random variable X , $E(E(X|\mathcal{F}_s)|\mathcal{F}_t) = E(X|\mathcal{F}_t)$, where $t \leq s \leq T$.

$$\begin{aligned}
V(t) &= P(t, T_N) \mathbb{E}_t^{T_N} (V(T)/P(T, T_N)) \\
&= P(t, T_N) \mathbb{E}_t^{T_N} \left(V(T) \mathbb{E}_T^{T_N} (B(T_N)/B(T)) \right) \\
&= P(t, T_N) \mathbb{E}_t^{T_N} (V(T) B(T_N)/B(T)),
\end{aligned}$$

as before.

As mentioned, equations (4.25) and (4.26) effectively involve reinvestment of the proceeds $V(T)$ to align cash payment with the numeraire $P(t, T_N)$. If the numeraire expires before the derivative security, we can apply the same reinvestment idea, but this time to the numeraire asset. Consider for instance a derivative security maturing at time T_N (paying $V(T_N)$), and suppose we wish to extend the T -forward measure to price this option. For instance, we can define a numeraire asset as follows:

$$\tilde{P}(t, T) = \begin{cases} P(t, T), & t \leq T, \\ B(t)/B(T), & t > T. \end{cases}$$

This asset corresponds to an investment strategy where we i) at time 0 purchase the T -maturity zero-coupon bond; and ii) at time T invest the proceeds from the zero-coupon bond (\$1) in the spot measure numeraire asset (4.24). Letting $\tilde{\mathbb{Q}}^T$ denote the measure induced by $\tilde{P}(t, T)$, we can write

$$\begin{aligned}
V(t) &= \tilde{P}(t, T) \tilde{\mathbb{E}}_t^T \left(V(T_N)/\tilde{P}(T_N, T) \right) \\
&= \tilde{P}(t, T) \tilde{\mathbb{E}}_t^T (V(T_N) B(T)/B(T_N)), \quad T < T_N, \quad t < T_N,
\end{aligned}$$

where $\tilde{\mathbb{E}}^T$ is the expectations operator for the measure $\tilde{\mathbb{Q}}^T$. If also $t \leq T$, this expression becomes (compare to (4.26))

$$V(t) = P(t, T) \tilde{\mathbb{E}}_t^T (V(T_N) B(T)/B(T_N)), \quad t \leq T < T_N,$$

which, in effect, uses $B(T)/B(T_N)$ to discount $V(T_N)$ back to time T . The equivalent result in the T -forward measure is

$$V(t) = P(t, T) \mathbb{E}_t^T (B(T) \mathbb{E}_T^B (V(T_N)/B(T_N))).$$

Notice that if V matures at time T , rather than at T_N , we simply have

$$V(t) = P(t, T) \tilde{\mathbb{E}}_t^T (V(T)) = P(t, T) \mathbb{E}_t^T (V(T)),$$

which is obvious from the definition of the numeraire $\tilde{P}(t, T)$.

The measure $\tilde{\mathbb{Q}}^T$ is by construction a hybrid between the spot measure and the T -forward measure. Obviously, many other such measures exist, corresponding to different reinvestment strategies of expiring numeraire assets.

4.2.5 Swap Measures

Being a linear combination of zero-coupon bonds (see (4.8)), an annuity factor $A_{k,m}(t)$ on a tenor structure qualifies as a numeraire asset. The measure $Q^{k,m}$ induced by this numeraire is known as a *swap measure* or an *annuity measure*. In the absence of arbitrage we have

$$V(t) = A_{k,m}(t) E_t^{k,m} (V(T)/A_{k,m}(T)),$$

where $E^{k,m}(\cdot)$ denotes expectation under $Q^{k,m}$.

Lemma 4.2.4. *In the absence of arbitrage, the forward swap rate $S_{k,m}(t)$ is a martingale in measure $Q^{k,m}$.*

Proof. By definition

$$S_{k,m}(t) = \frac{P(t, T_k) - P(t, T_{k+m})}{A_{k,m}(t)}.$$

As the numeraire deflated assets $P(t, T_k)/A_{k,m}(t)$ and $P(t, T_{k+m})/A_{k,m}(t)$ must both be martingales, so must be their difference. \square

As we shall see later, swap measures are very useful for analytical manipulations of price formulas for options on swaps.

4.3 Multi-Currency Markets

While this book is primarily dedicated to the study of single-currency interest rate derivatives, occasionally it will be necessary to consider certain effects associated with trading in a multi-currency economy. For instance, in Chapter 6 we touch upon issues of yield curve constructions in non-domestic currencies, and in Chapter 16 we discuss the important practical case where a derivative pays out in a foreign currency, but has a payout function that depends on one or more domestic interest rate variables. This brief section provides background material and notation required for these and other cross-currency applications.

4.3.1 Notations and FX Forwards

We consider two economies, a “domestic” economy and a “foreign” economy. Let $P_d(t, T)$ and $P_f(t, T)$ denote time t zero-coupon bond prices in the domestic and foreign economies, respectively. So, $P_f(t, T)$ (say) is the time t price, in foreign currency, of one unit of foreign currency delivered for certain at time T . Translation of values in foreign currency to domestic currency takes place at a foreign exchange (FX) rate of $X(t)$, measured in units of domestic currency per unit of foreign currency. In other words, the value \tilde{P}_d to a domestic investor of one foreign zero-coupon bond is

$$\tilde{P}_d(t, T) = X(t)P_f(t, T).$$

The quantity

$$X_T(t) = \frac{\tilde{P}_d(t, T)}{P_d(t, T)} = X(t) \frac{P_f(t, T)}{P_d(t, T)}$$

is known as the *forward FX rate* to time T . The name is motivated by the following arbitrage strategy:

- Buy one foreign zero-coupon bond, at a cost of $\tilde{P}_d(t, T)$ in domestic currency.
- Finance the purchase by selling short domestic zero-coupon bonds on a notional of $\tilde{P}_d(t, T)/P_d(t, T)$.

With no outlay at time t , the strategy will generate a net cash flow at time T of one unit of foreign currency and $-X_T(t)$ units of domestic currency, such that the trading strategy in effect has locked in a time t a future time T exchange rate of $X_T(t)$.

4.3.2 Risk Neutral Measures

Let $\beta_d(t)$ and $\beta_f(t)$ be the continuously compounded money market accounts in the domestic and foreign economies, respectively. $\beta_d(t)$ and $\beta_f(t)$ induce two separate risk-neutral measures, denoted Q^d and Q^f ; let us investigate how these measures are related. If $g(T)$ is a random payout at time T made in foreign currency, in a complete market the value (in units of foreign currency) of this payout to a foreign investor is, from standard principles,

$$V_f(t) = \beta_f(t)E_t^f(g(T)\beta_f(T)^{-1}), \quad (4.27)$$

where E_t^f denotes expectations in the foreign risk-neutral measure Q^f . For a domestic investor, the payout of $g(T)$ must be translated to domestic currency units at a rate of $X(T)$, making the effective domestic payout function $g(T)X(T)$. Thereby,

$$V_d(t) = \beta_d(t)E_t^d(g(T)X(T)\beta_d(T)^{-1}), \quad (4.28)$$

where E_t^d denotes expectations in measure Q^d . Importantly, the expressions in (4.27) and (4.28) are linked by the spot exchange rate, as the absence of a cross-currency arbitrage dictates that

$$V_d(t) = X(t)V_f(t),$$

or

$$\beta_d(t)E_t^d(g(T)X(T)\beta_d(T)^{-1}) = X(t)\beta_f(t)E_t^f(g(T)\beta_f(T)^{-1}). \quad (4.29)$$

We use this result to establish the following lemma:

Lemma 4.3.1. *The domestic and foreign risk-neutral probability measures Q^d and Q^f are related by the density process*

$$E_t^d \left(\frac{dQ^f}{dQ^d} \right) = \frac{\beta_f(t)X(t)}{\beta_d(t)X(0)}, \quad t \geq 0.$$

Proof. For an \mathcal{F}_T -measurable variable $Y(T) = g(T)X(T)\beta_d(T)^{-1}$ satisfying regularity conditions, a rearrangement of the basic relation (4.29) yields

$$E_t^d(Y(T)) = X(t) \frac{\beta_f(t)}{\beta_d(t)} E_t^f \left(\frac{Y(T)}{X(T)} \frac{\beta_d(T)}{\beta_f(T)} \right),$$

From the results of Section 1.3, the density relating measures Q^d and Q^f is then as given in the lemma. \square

With $\beta_f(t)X(t)/\beta_d(t)$ being a martingale in the domestic risk-neutral measure, we note that if $X(t)$ is an Ito process, it must take the form

$$dX(t) = X(t)(r_d(t) - r_f(t)) dt + \sigma_X(t)^\top dW(t),$$

where $r_d(t) - r_f(t)$ is the spread between domestic and foreign short rates, $W(t)$ is a (vector-valued) Q^d -Brownian motion, and $\sigma_X(t)$ is some adapted stochastic process satisfying regularity conditions.

4.3.3 Other Measures

Having established the Radon-Nikodym derivative relating the domestic and foreign risk-neutral measures, relations between various other domestic and foreign probability measures are easy to establish. For instance, the following result is easily proven the same way as Lemma 4.3.1.

Lemma 4.3.2. *Let $E_t^{T,d}$ denote expectations in the domestic T -forward probability measure. The domestic and foreign T -forward probability measures $Q^{T,d}$ and $Q^{T,f}$ are related by the density process*

$$E_t^{T,d} \left(\frac{dQ^{T,f}}{dQ^{T,d}} \right) = \frac{P_f(t, T)P_d(0, T)X(t)}{P_d(t, T)P_f(0, T)X(0)} = \frac{X_T(t)}{X_T(0)}, \quad t \geq 0.$$

We highlight the fact that the forward FX rate $X_T(t)$ is a $Q^{T,d}$ -martingale satisfying $X_T(T) = X(T)$. For an \mathcal{F}_T -measurable variable $Y(T)$, we thereby have the convenient expression

$$E_t^{T,d}(Y(T)) = X_T(t) E_t^{T,f} \left(\frac{Y(T)}{X(T)} \right),$$

where $E_t^{T,f}$ denotes expectations in the foreign T -forward probability measure.

4.4 The HJM Analysis

Having defined notations and established basic arbitrage relationships, let us turn to assigning dynamics to the many quantities we have introduced so far. We shall here follow the Heath et al. [1992] (Heath-Jarrow-Morton, or HJM) approach, where all information in the economy is assumed to originate with a finite number of Brownian motions. The resulting class of models is quite broad, and in much of the rest of this book we shall deal with ways to reduce the general HJM model to specific, and tractable, special cases. For now, however, we concentrate on a general analysis, although we keep our treatment fairly informal.

4.4.1 Bond Price Dynamics

In the HJM framework, we concern ourselves with the modeling of how an entire continuum of T -indexed bond prices $P(\cdot, T)$ jointly evolves over time, starting from a known condition $P(0, T)$. We consider models of finite horizon, i.e. with $T \in [0, \mathcal{T}]$, $\mathcal{T} < \infty$, and specialize to a filtration generated by a d -dimensional Brownian motion. We assume that a risk-neutral measure Q exists and is unique. Let $W(t)$ be an adapted d -dimensional Q -Brownian motion, and define deflated bond values as $P_\beta(t, T) = P(t, T)/\beta(t)$, where $\beta(t)$ as always is the continuously rolled money market account. In the absence of arbitrage, $P_\beta(t, T)$ is a martingale in the risk-neutral measure, and the martingale representation theorem then implies that

$$dP_\beta(t, T) = -P_\beta(t, T)\sigma_P(t, T)^\top dW(t), \quad t \leq T, \quad (4.30)$$

where $\sigma_P(t, T) = \sigma_P(t, T, \omega)$ is a d -dimensional stochastic process adapted to the filtration generated by W . We assume that $\sigma_P(t, T)$ is regular enough for $P_\beta(t, T)$ to be a square-integrable martingale. Also, as the bond $P(t, T)$ must equal \$1 at $t = T$ (“pull to par”), we impose the consistency condition

$$\sigma_P(T, T) = 0.$$

Using (4.12) and Ito’s lemma, it follows from (4.30) that

$$dP(t, T)/P(t, T) = r(t) dt - \sigma_P(t, T)^\top dW(t), \quad (4.31)$$

where $r(t)$ is the short rate process. Equation (4.31) defines the class of d -dimensional HJM models.

Another application of Ito’s lemma shows that forward bond prices $P(t, T, T + \tau) = P(t, T + \tau)/P(t, T)$ must satisfy

$$\begin{aligned} dP(t, T, T + \tau)/P(t, T, T + \tau) &= -[\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top \sigma_P(t, T) dt \\ &\quad - [\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top dW(t). \end{aligned} \quad (4.32)$$

In the T -forward measure \mathbb{Q}^T , $P(t, T, T + \tau)$ is a martingale (see Section 4.2.2), and

$$dP(t, T, T + \tau)/P(t, T, T + \tau) = -[\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top dW^T(t), \quad (4.33)$$

where $W^T(t)$ is a \mathbb{Q}^T -Brownian motion. Comparison of (4.32) and (4.33) shows that

$$dW^T(t) = dW(t) + \sigma_P(t, T) dt \quad (4.34)$$

which by Girsanov's theorem identifies the density process for the measure shift between \mathbb{Q}^T and \mathbb{Q} in the HJM setting:

$$\varsigma(t) = \mathbb{E}_t^{\mathbb{Q}} \left(\frac{d\mathbb{Q}^T}{d\mathbb{Q}} \right) = \mathcal{E} \left(- \int_0^t \sigma_P(u, T)^\top dW(u) \right), \quad (4.35)$$

or

$$d\varsigma(t)/\varsigma(t) = -\sigma_P(t, T)^\top dW(t).$$

This result could, of course, have been established from the first principles as well — see equation (4.21).

4.4.2 Forward Rate Dynamics

Traditionally, HJM models are stated in terms of instantaneous forward rates, rather than bond prices. Besides eliminating the need to consider the short rate r , this also reveals a number of fundamental properties of the class of HJM models. By Ito's lemma, in measure \mathbb{Q} ,

$$d \ln P(t, T) = O(dt) - \sigma_P(t, T)^\top dW(t),$$

where for convenience we have omitted writing out the drift term. Differentiating the right- and left-hand sides of this equation with respect to T , we get from equation (4.4),

$$df(t, T) = \mu_f(t, T) dt + \sigma_f(t, T)^\top dW(t),$$

where

$$\sigma_f(t, T) = \frac{\partial}{\partial T} \sigma_P(t, T), \quad (4.36)$$

and $\mu_f(t, T)$ is listed below.

Lemma 4.4.1. *The process for $f(t, T)$ in the T -forward measure is*

$$df(t, T) = \sigma_f(t, T)^\top dW^T(t). \quad (4.37)$$

In the risk-neutral measure, the process is

$$\begin{aligned} df(t, T) &= \sigma_f(t, T)^\top \sigma_P(t, T) dt + \sigma_f(t, T)^\top dW(t) \\ &= \sigma_f(t, T)^\top \int_t^T \sigma_f(t, u) du dt + \sigma_f(t, T)^\top dW(t). \end{aligned} \quad (4.38)$$

Proof. The SDE (4.37) follows directly from the martingale relation (4.23). The risk-neutral process (4.38) then can be derived from the relations (4.34) and (4.35), with the second equality following from (4.36). \square

The equation (4.38) is often considered to be the main result of Heath et al. [1992]. It demonstrates that an HJM model is fully specified once the forward rate diffusion coefficients $\sigma_f(t, T)$ have been specified for all t and T . Note that HJM models take initial forward rates $f(0, T)$ as exogenous inputs, ensuring that these models are automatically consistent with discount bond prices at time 0. This is true irrespective of the choice of $\sigma_f(t, T)$, which can be set freely (subject to regularity conditions) from either empirical analysis, or from a calibration to market prices of fixed income derivatives.

While it is convenient that HJM models are automatically calibrated to initial bond prices, a number of other features of the general HJM model are less attractive. Particularly problematic is the sheer dimensionality of the model: to describe the time t state of a discount bond curve spanning $[t, T]$, we need to keep track of a continuum of forward rates $\{f(t, u), t \leq u \leq T\}$. By Lemma 4.4.1 the forward rate curve follows an infinite-dimensional diffusion process, leaving us with an infinite number of state variables to diffuse. In practice, the implementation of an HJM model will require either making special assumptions about the σ_f process that permit a finite-dimensional Markovian representation of the forward rate curve; or moving from infinitesimal forward rates to continuously compounded forward rates that span time-buckets of finite length. Chapters 10 through 13 and Section 4.5.2 below give examples of the former idea, and Andersen [1995] discusses the latter approach in a Monte Carlo setting. An idea closely related to the discussion in Andersen [1995] is to build a model around a finite set of simple (Libor) forward rates on a fixed tenor structure. This approach has a number of computational and theoretical advantages, and is the subject of Chapter 14. For now, we note that *any* arbitrage-free interest rate model set in a filtration generated exclusively by Brownian motions must be a special case of an HJM model. In particular, any such model must correspond to a particular choice of $\sigma_f(t, T)$.

4.4.3 Short Rate Process

As discussed earlier, specification of a short rate process is, in principle, sufficient to completely specify a full yield curve model. In the HJM framework, it follows from (4.38) that the short rate $r(t)$ in measure Q is

$$r(t) = f(t, t) = f(0, t) + \int_0^t \sigma_f(u, t)^\top \int_u^t \sigma_f(u, s) ds du + \int_0^t \sigma_f(u, t)^\top dW(u).$$

The process for $r(t)$ is generally not Markovian, as can be seen by focusing on the path-dependent term

$$D(t) = \int_0^t \sigma_f(u, t)^\top dW(u)$$

for which we must have

$$\begin{aligned} D(T) &= D(t) + \int_t^T \sigma_f(u, T)^\top dW(u) \\ &\quad + \left\{ \int_0^t \sigma_f(u, T)^\top dW(u) - \int_0^t \sigma_f(u, t)^\top dW(u) \right\}. \end{aligned} \quad (4.39)$$

Thereby

$$E^Q(D(T)|D(t)) \neq E_t^Q(D(T))$$

unless the bracketed term in (4.39) is either non-random, or a deterministic function of $D(t)$ (which is generally not the case).

An interesting area of investigation concerns the conditions under which either $r(t)$ is outright Markov⁴ or, less restrictively, can be written as

$$r(t) = h(t, x(t)),$$

for a deterministic function h and a finite-dimensional Markovian vector of state variables $x(t)$. Definitive results are given in Björk [2001], building on earlier (and considerably less abstract) work by Jamshidian [1991b], Cheyette [1991], and Ritchken and Sankarasubramanian [1995]. Section 4.5.2 and Chapter 13 list some of the results of these papers.

4.5 Examples of HJM Models

4.5.1 The Gaussian Model

In the HJM bond price dynamics (4.31), we now assume that $\sigma_P(t, T)$ is a bounded (d -dimensional) deterministic function of t and T . It follows from (4.32) and (4.33) that forward bond prices are then log-normally distributed in both Q and Q^T . The forward rate process in Q is

$$df(t, T) = \sigma_f(t, T)^\top \sigma_P(t, T) dt + \sigma_f(t, T)^\top dW(t), \quad \sigma_f(t, T) = \frac{\partial}{\partial T} \sigma_P(t, T), \quad (4.40)$$

which implies that $r(T) = f(T, T)$ is Gaussian with Q -moments

$$\begin{aligned} E_t^Q(f(T, T)) &= \int_t^T \sigma_f(u, T)^\top \sigma_P(u, T) du, \\ \text{Var}_t^Q(f(T, T)) &= \int_t^T \sigma_f(u, T)^\top \sigma_f(u, T) du. \end{aligned}$$

⁴In the sense that the time t expectations of functionals of $r(T)$ only require knowledge of $r(t)$ itself. In this case the process for $r(t)$ must be a diffusion characterized by an SDE $dr(t) = \mu_r(t, r(t))dt + \sigma_r(t, r(t))^\top dW(t)$.

The simple form of the Gaussian HJM model makes it quite tractable, permitting analytical price formulas for a number of European options and futures contracts⁵. While the Gaussian HJM model suffers from the drawback of allowing negative forward and spot rates, analytical results derived in the model are often very useful in gaining a deeper understanding of a given contract, even if ultimately a more realistic model will be required for serious pricing purposes. Indeed, results derived for the Gaussian HJM model can often be used as a starting point for development of closed-form approximations in other models; we shall see many examples of this later in the book.

For illustration, we list a few select analytical results below. More formulas can be found in numerous sources, including Chapter II in Andersen [1996], Jamshidian [1991b], and Jamshidian [1993], to name a few.

Proposition 4.5.1 (Option on Zero-Coupon Bond). *Consider a European call option paying at maturity T the amount*

$$V(T) = (P(T, T^*) - K)^+, \quad T^* > T.$$

In the Gaussian HJM model (4.40), we have

$$V(t) = P(t, T^*)\Phi(d_+) - P(t, T)K\Phi(d_-), \quad (4.41)$$

where

$$d_{\pm} = \frac{\ln(P(t, T^*) / (KP(t, T))) \pm v/2}{\sqrt{v}},$$

$$v = \int_t^T |\sigma_P(u, T^*) - \sigma_P(u, T)|^2 du.$$

Proof. In the T -forward measure \mathbb{Q}^T we have, from (4.20),

$$\begin{aligned} V(t) &= P(t, T)\mathbb{E}_t^T \left((P(T, T^*) - K)^+ \right) \\ &= P(t, T)\mathbb{E}_t^T \left((P(T, T, T^*) - K)^+ \right). \end{aligned}$$

From the discussion in Section 4.4.1 we know that $P(t, T, T^*)$ is a \mathbb{Q}^T -martingale characterized by the SDE

$$dP(t, T, T^*)/P(t, T, T^*) = -[\sigma_P(t, T^*) - \sigma_P(t, T)]^\top dW^T(t),$$

where W^T is a d -dimensional \mathbb{Q}^T -Brownian motion. As this is just a GBM process with time-dependent coefficients, the Black-Scholes-Merton results in Section 1.9 apply and lead to (4.41). \square

⁵Indeed, we have already used this model in an equity context — see Section 1.9.3.2.

Proposition 4.5.2 (Caplet). Consider a European call option paying at $T + \tau$ the amount (a caplet)

$$V(T + \tau) = \tau (L(T, T, T + \tau) - K)^+, \quad \tau > 0.$$

In the Gaussian HJM model (4.40), we have

$$V(t) = \tau P(t, T + \tau) (L(t, T, T + \tau) \Phi(d_+) - K \Phi(d_-)),$$

$$d_{\pm} = \frac{\ln(L(t, T, T + \tau)/K) \pm v/2}{\sqrt{v}}, \quad v = \int_t^T |\sigma_P(u, T + \tau) - \sigma_P(u, T)|^2 du.$$

Proof. From Lemma 4.2.3 we know that $L(t, T, T + \tau)$ is a martingale in the $(T + \tau)$ -forward measure. An application of Ito's lemma to the definition (4.2) reveals that

$$dL(t, T, T + \tau)/L(t, T, T + \tau) = [\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top dW^T(t),$$

and the result follows immediately along the same lines as in the proof of Proposition 4.5.1. \square

Proposition 4.5.3 (Futures Rate). In the Gaussian HJM model (4.40), futures rates are given by

$$F(t, T, T + \tau) = \tau^{-1} \left((1/P(t, T, T + \tau)) e^{\Omega(t, T)} - 1 \right), \quad (4.42)$$

where

$$\Omega(t, T) = \int_t^T [\sigma_P(u, T + \tau) - \sigma_P(u, T)]^\top \sigma_P(u, T + \tau) du.$$

Proof. From Lemma 4.2.2,

$$\begin{aligned} F(t, T, T + \tau) &= E_t^Q (L(T, T, T + \tau)) \\ &= \tau^{-1} E_t^Q (1/P(T, T + \tau) - 1) \\ &= \tau^{-1} E_t^Q (G(T) - 1), \end{aligned} \quad (4.43)$$

where we have introduced an auxiliary variable

$$G(t) \triangleq P(t, T)/P(t, T + \tau) = 1/P(t, T, T + \tau).$$

Ito's lemma shows that (see also (4.32)) in measure Q

$$\begin{aligned} dG(t)/G(t) &= [\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top \sigma_P(t, T + \tau) dt \\ &\quad + [\sigma_P(t, T + \tau) - \sigma_P(t, T)]^\top dW(t), \end{aligned}$$

such that

$$\mathbb{E}_t^Q(G(T)) = G(t)e^{\Omega(t,T)} = (1/P(t, T, T + \tau)) e^{\Omega(t,T)},$$

where Ω is as given above. The result of Proposition 4.5.3 then follows directly from (4.43). \square

In any rational model $\Omega(t, T) \geq 0$, such that $F(t, T, T + \tau) \geq L(t, T, T + \tau)$, consistent with the qualitative discussion in Section 4.1.2. As shown in Chapter II of Andersen [1996], the spread (also known as futures *convexity*) between futures and forward rates can be decomposed into two components: i) a term originating from the mark-to-market mechanism of a futures contract; and ii) a term originating from the fact that a futures contract — unlike a regular forward rate agreement — pays out the rate at the date it settles (at time T) rather than one period ahead (at time $T + \tau$). Andersen [1996], Chapter II, additionally contains a number of numerical examples examining typical futures-forward spreads, and also investigates the pricing of options on futures rates.

Section 16.8 looks in detail into pricing interest rate futures under more advanced models.

4.5.2 Gaussian HJM Models with Markovian Short Rate

Although quite tractable, the Gaussian HJM model generally does not allow for a finite-dimensional Markovian representation, and typically does not imply Markov-diffusive behavior of the short rate. As shown in Carverhill [1994], the short rate can be made Markovian, however, by imposing certain conditions on the deterministic forward rate volatility function $\sigma_f(t, T)$. To explore this, first recall from Section 4.4.3 the relation

$$r(t) = f(0, t) + \int_0^t \sigma_f(u, t)^\top \int_u^t \sigma_f(u, s) ds du + \int_0^t \sigma_f(u, t)^\top dW(u),$$

where now σ_f is deterministic. Consider imposing the special choice

$$\sigma_f(t, T) = g(t)h(T), \quad (4.44)$$

where h is a positive real function and $g : \mathbb{R} \rightarrow \mathbb{R}^{d \times 1}$ can take any sign. For this case we have

$$\sigma_P(t, T) = \int_t^T \sigma_f(t, u) du = g(t) \int_t^T h(u) du,$$

and

$$\begin{aligned} r(t) &= f(0, t) + h(t) \int_0^t g(u)^\top g(u) \left(\int_u^t h(s) ds \right) du + h(t) \int_0^t g(u)^\top dW(u) \\ &\triangleq f(0, t) + h(t) \int_0^t m_f(t, u) du + h(t) \int_0^t g(u)^\top dW(u). \end{aligned} \quad (4.45)$$

Importantly, the term

$$D(t) = \int_0^t \sigma_f(u, t)^\top dW(u) = h(t) \int_0^t g(u)^\top dW(u)$$

is now Markov, since

$$D(T) = h(T) \int_0^T g(u)^\top dW(u) = \frac{h(T)}{h(t)} D(t) + h(T) \int_t^T g(u)^\top dW(u),$$

which should be compared to the general (non-Markov) expression (4.39).

To show that the short rate is Markovian, we differentiate (4.45) with respect to t , yielding

$$\begin{aligned} dr(t) &= \frac{\partial f(0, t)}{\partial t} dt + h'(t) \left(\int_0^t m_f(t, u) du + \frac{D(t)}{h(t)} \right) dt \\ &\quad + h(t) \frac{\partial}{\partial t} \int_0^t m_f(t, u) du dt + h(t) g(t)^\top dW(t) \\ &= \frac{\partial f(0, t)}{\partial t} dt + h'(t) \left(\frac{r(t) - f(0, t)}{h(t)} \right) dt \\ &\quad + h(t)^2 \int_0^t g(u)^\top g(u) du dt + h(t) g(t)^\top dW(t) \\ &= \left(\frac{\partial f(0, t)}{\partial t} - \frac{h'(t)}{h(t)} f(0, t) + h(t)^2 \int_0^t g(u)^\top g(u) du + \frac{h'(t)}{h(t)} r(t) \right) dt \\ &\quad + h(t) g(t)^\top dW(t), \end{aligned} \tag{4.46}$$

where the second equality follows from rearrangement of (4.45). This leads to the following result.

Proposition 4.5.4. *In the d -dimensional Gaussian HJM model, when (4.44) holds the short rate satisfies an SDE of the type*

$$dr(t) = (a(t) - \varkappa(t)r(t)) dt + \sigma_r(t)^\top dW(t),$$

where $\varkappa : \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma_r : \mathbb{R} \rightarrow \mathbb{R}^{d \times 1}$ are deterministic functions of time, and

$$\begin{aligned} a(t) &= \frac{\partial f(0, t)}{\partial t} + \varkappa(t)f(0, t) + \int_0^t e^{-2 \int_u^t \varkappa(s) ds} \sigma_r(u)^\top \sigma_r(u) du \\ &= \frac{\partial f(0, t)}{\partial t} + \varkappa(t)f(0, t) + \int_0^t \sigma_f(u, t)^\top \sigma_f(u, t) du. \end{aligned}$$

Proof. First, by way of defining \varkappa and σ_r , we set

$$h(T) = e^{-\int_0^T \varkappa(s) ds}; \quad g(t) = e^{\int_0^t \varkappa(s) ds} \sigma_r(t),$$

such that $h'(t)/h(t) = -\varkappa(t)$ and

$$\sigma_f(t, T) = e^{-\int_t^T \varkappa(s) ds} \sigma_r(t).$$

The result of Proposition 4.5.4 then follows directly by insertion into (4.46).

□

4.5.3 Log-Normal HJM Models

To avoid the negative forward rates inherent in Gaussian HJM models, it is tempting to consider forward rate specifications of the type

$$\sigma_f(t, T) = f(t, T)\sigma(t, T), \quad (4.47)$$

where $\sigma(t, T)$ is deterministic and bounded. In the T -forward measure

$$df(t, T) = f(t, T)\sigma(t, T)^\top dW^T(t)$$

such that $f(t, T)$ is log-normally distributed. While avoiding negative rates, the specification (4.47) has severe technical problems: in Q , forward rates will explode to infinity with non-zero probability. Attempts to apply the valuation formula (4.15) will thereby result in all zero-coupon bond prices being zero, implying obvious arbitrage opportunities. To suggest a rationale for the exploding rates, consider the Q -dynamics

$$df(t, T) = \left(f(t, T)\sigma(t, T)^\top \int_t^T f(t, u)\sigma(t, u)du \right) dt + f(t, T)\sigma(t, T)^\top dW(t).$$

Loosely speaking, the drift-term is proportional to forward rates *squared*, which, in the light of the linear growth condition in Theorem 1.6.1, may cause us to suspect problems with the existence of a non-exploding solution. Morton [1988] confirms this rigorously.

One solution to the explosion problem involves enforcing a strict upper bound on $\sigma_f(t, T)$, as in

$$\sigma_f(t, T) = \min(f(t, T), M)\sigma(t, T),$$

where M is a large positive constant. For the one-factor case ($d = 1$) Heath et al. [1992] demonstrate that this specification will ensure non-negative forward rates⁶ and will prevent rate explosions. Nevertheless, the model is clearly awkward in its dependence on the arbitrary constant M . A more satisfying solution is discussed in Chapter 14, where we show that the explosion problem can be circumvented by working with simply — rather than continuously — compounded forward rates. A related issue in short rate models is also discussed in Chapter 11.

⁶To see this, notice that $M > 0$ guarantees that $df(t, T) = 0$ if $f(t, T)$ should ever reach 0.

Fixed Income Instruments

At this point, we have established the mathematical and numerical prerequisites needed for the remaining part of the book, much of which is devoted to the development of models for fixed income derivatives. Before delving into the modeling exercise, this final foundational chapter provides a tour of actual fixed income markets as well as an overview of the types of products traded. The simpler (and more liquid) of these products will typically serve as calibration targets to parameterize the models we develop; others (the more complicated and illiquid ones) will constitute the contracts that our model are ultimately meant to price and hedge. Throughout the chapter — and, indeed, this book — our focus is on the securities tied to the so-called *Libor rate*; this will include essentially all high-end exotic securities as well as more basic instruments such as FRAs, caps, and swaptions. Our priorities dictate that we leave out government, corporate, and mortgage bonds, as well as the derivatives associated with these types of securities. A discussion of these classes of securities, along with many more details on the organization and workings of fixed income markets, can be found in specialist literature, such as Fabozzi and Modigliani [1996], Fabozzi [1985], Fabozzi [2001], and Fabozzi and Fabozzi [1989].

5.1 Fixed Income Markets and Participants

At the most fundamental level, interest rates determine the economic cost of borrowing and lending, and as such define present values of future cash flows. In general, cash flows occurring at different times are discounted at different rates, reflecting market fluctuations in demand for money and risk preferences of market participants. The dependence of interest rates on time is described by the so-called *term structure of interest rates*, easily visualized as a curve that assigns a particular interest rate (or, equivalently, a discount factor) to each future date.

For a given entity, the cost of borrowing money will depend on its credit quality. Governments of developed countries, perceived to have virtually no possibility of default, issue bonds at comparatively low interest rates that reflect this perception. While the market in government debt is vast, corporations typically find it more convenient to use and originate fixed income instruments linked to rates that are more reflective of their own financing costs (i.e., credit quality). By far the most common of such reference rates is the London Interbank Offered rate, commonly known as the *Libor rate*. The Libor rate is a filtered average of bank estimates of rates at which they can borrow for a given term in the *interbank money market*, i.e. the wholesale market in which banks provide unsecured short-term credit to each other. Libor rates are quoted for multiple deposit maturities ranging from one day to one year, and are set every business day by averaging polling results from a number of large banks. Libor rates are available for deposits in different currencies, so that there is a USD-Libor rate, a EUR-Libor rate, and so on.

While Libor rates are probably the most used reference rates for interest rate contracts, there are other important rates to be aware of. For example, in the United States, banks are required to hold certain balances (“Federal funds”) with the Federal Reserve, the central bank of the US. If a bank does not have sufficient balances, it can borrow them from another bank that has an excess on its account. The overnight interest rate charged in this case is called the (*effective*) *Federal funds rate*¹, or sometimes simply the Fed funds rate. This rate is often considered the best available proxy for a risk-free USD rate, in part because the Fed funds rate is normally the contractual rate used to accrue interest on posted collateral², as explained in Piterbarg [2010]. It is worth noting that the Fed funds rate used to be closely linked to the overnight Libor rate, with the spread between the two in the single basis points. However, in the subprime crisis of 2007–2009 the two have diverged significantly; the implications of this for interest rate curve construction are discussed in Section 6.5.3. Instruments linked to averages of the (*effective*) Fed fund rate over different terms are actively traded, giving rise to a term structure of Fed funds linked rates.

A special feature of the US public debt markets gives rise to another set of rates. In particular, interest on bonds issued by states and other local governments of the US is often free of the federal tax. The Bond Market Association, a trade association of the bond industry, publishes the *BMA rate* (or *BMA index*) which is the estimate of borrowing by such municipalities.

¹The target rate, set by the Federal reserve, is aptly called the *target* Fed funds rate.

²To mitigate credit risk, many derivatives transactions require posting of collateral (normally cash or Treasury bonds) in the amount of the current mark-to-market. ISDA [2005] contains a detailed description of collateral agreements; according to ISDA [2009], in 2009 about 65% of all OTC derivatives transactions involved such agreements.

There is a well-developed market in interest rate derivatives that are linked to the BMA rate.

The Euro and GBP markets do not have the same mechanism as the US does for Federal funds, but overnight rates that are proxies for risk-free borrowing in those currencies do exist. They are called *Eonia* (Euro OverNight Index Average) in the Eurozone and *Sonia* (Sterling OverNight Index Average) in Great Britain, and are computed as averages of all *actual* overnight lending/borrowing transactions by qualifying banks weighted by the size of the transactions. We emphasize that these rates reflect the actual transactions that have happened, in contrast to Libor which reflect banks' estimates of rates at which borrowing (for a given term) might take place. In the crisis of 2007–2009 there have been serious concerns about the integrity of the Libor rate and whether it really reflected the actual cost of funding for banks, and even some calls to scrap the Libor rate altogether. While the Libor rate has survived the crisis, the importance of overnight rates has increased dramatically, with the market in FedFunds/Eonia/Sonia linked derivatives, most importantly in *overnight index swaps*, or OIS, of various maturities growing dramatically. As with the Fed funds rate, Eonia and Sonia have diverged significantly from the corresponding Libor rates during market turbulence, and the decoupling continues to persist. As with the Fed funds rate, the implications of these developments on interest rate curve construction are discussed in Section 6.5.3.

Interest rates change day-to-day in response to changing macroeconomic and market conditions. With the cost of borrowing and lending money affecting all aspects of the economy, it is no surprise that a vast market in derivatives on interest rates has developed. Motivations of participants are diverse, ranging from locking in the cost of financing to pure speculation.

The fixed income market can be broadly split into two (overlapping) segments: the *exchange* market and the *over-the-counter*, or *OTC*, market. Contracts linked to the level of interest rates are traded on many securities exchanges. The exchanges attract all types of investors, including market makers, hedgers and speculators; see Hull [2006] for details on all. As of March 2008, notional amounts outstanding were \$26 trillion in exchange traded interest rate futures, and \$45 trillion in exchange traded interest rate options. While these are impressive numbers, far more fixed income derivatives trade in OTC markets than in exchange markets: as of December 2007, the notional amounts outstanding of OTC interest rate derivatives amounted to \$393 trillion³. The OTC market can loosely be visualized as a network of banks that trade with each other under terms governed by agreements spelled out by the trade organization International Swaps and Derivatives Association (ISDA). Central to OTC markets are the *interest*

³All figures from the report “Semiannual OTC derivatives statistics at end-December 2007” by Bank for International Settlements, available from www.bis.org.

rate dealers, banks with trading desks specializing in fixed income trading. The dealers provide liquidity in various types of securities, and are typically the most sophisticated players in the market. The dealers trade either on their own account or on behalf of customers such as *financial institutions* and *corporates*.

Financial institutions include mortgage companies (organizations that originate, package or service residential and commercial mortgage loans), pension funds, mutual funds, insurance companies, hedge funds, and other entities whose primary activities are related to financial markets. Financial institutions seek to either make money directly by engaging in trading activities (hedge funds), or to hedge their exposures (mortgage originators or servicers), or to achieve superior returns on their investments (pension funds, insurance companies). Among financial institutions, an important role is played by *issuers*, companies that issue structured notes for private and public placement. Structured notes deliver appealing return profiles to investors, returns that are essentially financed by selling options back to issuers. Issuance of increasingly complicated structured notes drives the exotic end of the fixed income markets.

Corporates are companies with primary activities not directly linked to fixed income markets, but whose operational results may be affected by the interest rate environment. For instance, many companies raise funds by borrowing from banks or by selling bonds, and are therefore affected by the prevailing levels of interest rates. Corporates often seek to lock in favorable interest rates for borrowing money, to hedge their interest rate exposures, to transform their liabilities from one type (e.g., a fixed rate liability) to another (a floating rate liability), or to design custom borrowing schemes around their expected future borrowing needs.

5.2 Certificates of Deposit and Libor Rates

Having identified the main types of market participants, we now proceed to define the universe of securities that this book will cover. For technical precision, we shall occasionally need to refer to the risk-neutral measure Q , as well as its associated expectation operator $E = E^Q$ and its numeraire $\beta(t)$.

We start with the *certificate of deposit* (or *CD*), a deposit of money for a pre-specified term at a pre-specified interest rate. Terms may range from one week to one year or more, with the most popular being a 3 month or a 6 month term, depending on the currency of the deposit. If 1 (dollar) is deposited at time T for a period of τ years, then the amount of capital to be returned at time $T + \tau$ is given by⁴

⁴As was mentioned earlier, the computation of τ from given start- and end-dates will involve certain formal day counting rules, see Appendix 5.A.

$$1 + \tau L,$$

where L is, by definition, the interest rate for the CD. The rate is quoted as a simple rate, i.e. a rate with the compounding frequency equal to the term of the deposit. Notice that the average value of L for CDs quoted in the interbank market will, by definition, be equal to the (spot) Libor rate for tenor τ . Spot Libor rates for various tenors are calculated daily and are published by major news services such as Bloomberg or Reuters. As mentioned above, Libor serves as the primary reference rate in fixed income markets.

If $P(T, T + \tau)$ is the (Libor-based) discount factor to date $T + \tau$ as observed at T , then the discounted value of receiving $1 + \tau L$ at time $T + \tau$ should be equal to 1 at time T , i.e.

$$1 = P(T, T + \tau)(1 + \tau L).$$

In particular, recalling the definition (4.2) of $L(t, T, T + \tau)$, the rate L paid on the CD is a simple spot rate

$$L = L(T, T, T + \tau) = \frac{1}{\tau} \left(\frac{1}{P(T, T + \tau)} - 1 \right). \quad (5.1)$$

5.3 Forward Rate Agreements (FRA)

A certificate of deposit allows a market participant to lock in an interest rate for a given period of time, effective immediately. Many market participants, however, find it convenient to lock in interest rates for a given period of time that starts in the future. Contracts that provide such a rate guarantee are known as *forward contracts* or, in a fixed income context, *forward rate agreements* (FRAs). An FRA for the period $[T, T + \tau]$ is a contract to exchange fixed rate payment (agreed at the initiation of the contract) against a payment based on the time T spot Libor rate of tenor τ . While all payments on an FRA are exchanged at, or near⁵, time T , the contract is structured so that the payments are made in $T + \tau$ dollars.

Formally, consider the origination at time t , $t \leq T$, of a unit notional FRA contract with a rate of k . Ignoring payment delays, from the perspective of the fixed rate payer the net payment at time T will be

$$V_{\text{FRA}}(T) = \tau(L(T, T, T + \tau) - k) / (1 + \tau L(T, T, T + \tau)),$$

with the (contractually specified) factor $1/(1 + \tau L(T, T, T + \tau))$ applied to roll the payment to the future date $T + \tau$. We note that

⁵Typical market conventions call for a two business day payment delay, see Appendix 5.A for more details.

$$1 / (1 + \tau L(T, T, T + \tau)) = P(T, T + \tau)$$

so, by the fundamental pricing result (4.13), the value of this contract at time t is equal to

$$V_{\text{FRA}}(t) = \beta(t) E_t (\beta(T)^{-1} \tau (L(T, T, T + \tau) - k) P(T, T + \tau))$$

(recall that $\beta(\cdot)$ is the money market account). Substituting (5.1) we obtain

$$V_{\text{FRA}}(t) = \beta(t) E_t (\beta(T)^{-1} (1 - P(T, T + \tau) - \tau k P(T, T + \tau))).$$

Since $P(\cdot, T + \tau)$ is a traded asset, its price deflated by the numeraire $\beta(\cdot)$ is a martingale. Thus

$$\begin{aligned} V_{\text{FRA}}(t) &= P(t, T) - P(t, T + \tau) - \tau k P(t, T + \tau) \\ &= \tau P(t, T + \tau) \left(\frac{P(t, T) - P(t, T + \tau)}{\tau P(t, T + \tau)} - k \right). \end{aligned} \quad (5.2)$$

Most often, FRAs are issued at no cost to either party at the time of origination. The value of k that makes the FRA contract have value 0 at the contract initiation time t is given by the *forward Libor rate* (see (4.2)),

$$k = L(t, T, T + \tau) = \frac{P(t, T) - P(t, T + \tau)}{\tau P(t, T + \tau)}.$$

Thus, a forward Libor rate has the financial interpretation of being a break-even rate on an FRA contract in interbank markets.

5.4 Eurodollar Futures

FRAs, being forward contracts on Libor rates, allow market participants to either lock in favorable rates for future periods, or to speculate on the future direction of rates. FRAs trade in the OTC market, and are open only to institutions that participate in this market. Alternatively, *futures contracts* on Libor rates are available on a number of international exchanges, including the Chicago Mercantile Exchange (CME), London International Financial Futures and Options Exchange (LIFFE), and Marché à Terme International de France (MATIF). The CME interest rate futures contract on a three-month spot Libor rate on US dollar denominated deposits is called the *Eurodollar futures* or, simply, *ED futures* contract.

At maturity T , an ED futures contract is settled at

$$100 \times (1 - L(T, T, T + \tau)).$$

The *futures rate* $F(t, T, T + \tau)$ at time t (see (4.1.2)) is defined to be the rate such that the *quoted futures price* at time t is equal to⁶

⁶So, if the futures rate is 5%, the quoted futures price is 95.

$$100 \times (1 - F(t, T, T + \tau)).$$

As is the case for all futures contracts, ED futures are settled (marked to market) daily. Confusing matters somewhat, the actual amount of money that is settled between holders of the long and the short positions in an ED future is determined by the daily change in the *actual futures price* defined by

$$N_{ED} \times \left[1 - \frac{1}{4} F(t, T, T + \tau) \right],$$

where N_{ED} is the notional principal of the contract (\$1,000,000 for the CME's ED futures). In particular, for 1 basis point (0.01%) increase in the rate $F(t, T, T + \tau)$, the CME contract buyer pays $1,000,000 \times 0.25 \times 0.0001 = 25$ dollars to the seller.

As explained in Chapter 4, futures rates $F(t, T, T + \tau)$ are generally different from forward Libor rates $L(t, T, T + \tau)$. The problem of computing the difference, the *ED convexity adjustment*, is considered in Section 16.8.

Unlike FRAs, for which the deposit period is negotiated between two parties, ED futures are standardized. Available contracts expire on four specific dates, one each in March, June, September and December, over the next ten years. Such standardization increases liquidity in each particular contract.

5.5 Fixed-for-Floating Swaps

A *swap* is a generic term for an OTC derivative in which two counterparties agree to exchange one stream of cash flows against another stream. These streams are called the *legs* of the swap. A *plain vanilla fixed-for-floating interest rate swap* (a *plain vanilla swap*, or just a *swap* if there is no confusion) is a swap in which one leg is a stream of fixed rate payments and the other a stream of payments based on a floating rate, most often Libor. The legs are denominated in the same currency, have the same notional, and expire on the same date. Payment streams are made on a pre-defined schedule of contiguous time intervals, known as *periods*. Typically, the floating rate is observed (or *fixed*) at the beginning of each period, with both fixed and floating rate coupons being paid out at the end of the period. A plain-vanilla swap is economically equivalent⁷ to a multi-period FRA, and serves the same purpose in the market as regular FRAs. Between interest rate dealers

⁷This is true up to subtle but potentially important discounting issues. As we have pointed out in Section 5.3, the net payment of an FRA is *contractually* discounted using Libor rate from $T + \tau$ to T , whereas in a swap, the net payment for a given period is discounted at the money market account rate from the end to the beginning of the accrual period. The two types of discounting can in fact be different in the presence of discounting-index *basis*, see Sections 6.5.2 and 6.5.3.

and financial institutions, swaps of different maturities are often traded to adjust interest risk positions of the parties involved, or to simply make bets on future direction of interest rates. Swaps are also used by corporates, often in conjunction with bond or note issuance, to transform fixed rate obligations into floating ones, or vice versa.

To formally define a fixed-floating swap, one specifies a tenor structure, i.e. an increasing sequence of maturity times, normally spaced roughly equidistantly (see Section 4.1.3)

$$0 \leq T_0 < T_1 < T_2 < \dots < T_N, \quad \tau_n = T_{n+1} - T_n. \quad (5.3)$$

In a fixed-floating swap with fixed rate k , one party (the fixed rate payer) pays simple interest based on the rate k in return for simple interest payments computed from the Libor rate fixing on date T_n , for each period $[T_n, T_{n+1}]$, $n = 0, \dots, N - 1$. The payments are exchanged at the end of each period, i.e. at time T_{n+1} . In practice, the payments are netted, and only their difference changes hands. From the perspective of the fixed rate payer, the net cash flow of the swap at time T_{n+1} is therefore given by (on a unit notional)

$$\tau_n (L_n(T_n) - k), \quad L_n(t) = L(t, T_n, T_{n+1}),$$

for $n = 0, \dots, N - 1$. Dates when the Libor rates are observed are typically called *fixing dates*; dates when payments occur are called *payment dates*.

By the fundamental valuation result (4.13), the value of a swap is equal to the expected discounted value of its (netted) payments. Specifically, the value to the fixed rate payer of a unit notional fixed-floating swap at time t , $0 \leq t \leq T_0$, is given by⁸

$$\begin{aligned} V_{\text{swap}}(t) &= \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (L_n(T_n) - k) \right) \\ &= \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_n)^{-1} (L_n(T_n) - k) P(T_n, T_{n+1}) \right). \end{aligned}$$

Using the definition of Libor rates $L_n(T_n)$,

$$V_{\text{swap}}(t) = \beta(t) \sum_{n=0}^{N-1} E_t \left(\beta(T_n)^{-1} (1 - P(T_n, T_{n+1}) - \tau_n k P(T_n, T_{n+1})) \right).$$

For each n , $P(\cdot, T_n)$ is a traded asset, so its price deflated by the numeraire $\beta(\cdot)$ is a martingale. Hence

$$V_{\text{swap}}(t) = \sum_{n=0}^{N-1} (P(t, T_n) - P(t, T_{n+1}) - \tau_n k P(t, T_{n+1})).$$

⁸This is a somewhat idealized expression. See Appendix 5.A for more details on market day counting conventions and related topics.

Recalling the definition of $L_n(t)$, this can be rewritten as

$$V_{\text{swap}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) (L_n(t) - k).$$

An important observation is that a vanilla fixed-floating swap can be valued on date t using only the term structure of interest rates observed on that date. In particular, swap values are not affected by the *dynamics* of interest rates, only their current levels.

The swap valuation formula above can be rewritten as follows,

$$V_{\text{swap}}(t) = \left(\sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) \right) \left(\frac{\sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) L_n(t)}{\sum_{n=0}^{N-1} \tau_n P(t, T_{n+1})} - k \right).$$

Using the definitions (4.8), (4.10) and (4.11) from Chapter 4:

$$A(t) \triangleq A_{0,N}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}), \quad (5.4)$$

$$S(t) \triangleq S_{0,N}(t) = \frac{\sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) L_n(t)}{\sum_{n=0}^{N-1} \tau_n P(t, T_{n+1})}, \quad (5.5)$$

we obtain the convenient formula

$$V_{\text{swap}}(t) = A(t) (S(t) - k). \quad (5.6)$$

The quantity $A(\cdot)$ is the *annuity* of the swap (or its *PVBP*, for Present Value of a Basis Point), and the quantity $S(t)$ is the *forward swap rate*. Clearly, $S(t)$ is the value of the fixed rate that makes the swap have value 0 to both parties at time t ; S is consequently often referred to as a *par* or *break-even* rate.

For plain-vanilla swaps, the fixed rate and the swap notional are constant through time. More general swaps are, however, not bound by such restrictions and both the fixed rate and the notional may vary from period to period. A non-standard swap with a notional schedule $\{q_n\}_{n=0}^{N-1}$ (non-constant but deterministic) and a fixed rate schedule $\{k_n\}_{n=0}^{N-1}$ has the value

$$\begin{aligned} V_{\text{genswap}}(t) &= \beta(t) \sum_{n=0}^{N-1} \tau_n q_n E_t \left(\beta(T_{n+1})^{-1} (L_n(T_n) - k_n) \right) \\ &= \sum_{n=0}^{N-1} \tau_n q_n P(t, T_{n+1}) (L_n(t) - k_n). \end{aligned}$$

Certain general swaps have dedicated names, such as *amortizing swaps* (notional decreases with time) and *accreting swaps* (notional increases with time).

As we mentioned in Section 5.1, swaps linked to overnight rates (Fed-Funds/Eonia/Sonia) have recently become more popular. Among them the overnight index swap (OIS) is probably the most liquid, and is defined as a swap that pays a *compounded* overnight rate against fixed rate payments. To write down its definition, let us assume that a tenor structure (5.3) is given, and denote by $\{t_{n,i}\}_{i=1}^{K_n}$ the collection of all business days in the period $[T_n, T_{n+1})$, so that $T_n = t_{n,1} < \dots < t_{K_n} < T_{n+1}$. Then the net payment of the OIS with fixed rate k at time T_{n+1} is given by

$$\tau_n (\bar{L}_n - k),$$

where the floating rate \bar{L}_n for the n -th period of OIS is given by

$$\bar{L}_n = \frac{1}{\tau_n} \left(\prod_{i=1}^{K_n-1} (1 + (t_{i+1} - t_i)L(t_i, t_i, t_{i+1})) - 1 \right). \quad (5.7)$$

Here we used the notation $L(t_i, t_i, t_{i+1})$ to denote the overnight rate. Equating the overnight rate with the short rate, we can use a more mathematically convenient (although not exactly correct) expression

$$\bar{L}_n = \frac{1}{\tau_n} \left(e^{\int_{T_n}^{T_{n+1}} r(t) dt} - 1 \right). \quad (5.8)$$

5.6 Libor-in-Arrears Swaps

Allowing the fixed rate and the notional to vary through time is not the only way to generalize a swap. For a *Libor-in-arrears swap*, Libor rates are observed (fixed) at the end of each period rather than at the beginning. Thus, a value of a Libor-in-arrears payer swap is equal to

$$V_{\text{LIA}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (L_{n+1}(T_{n+1}) - k) \right).$$

Interestingly, this seemingly innocuous modification makes the value of a swap model-dependent, in contrast to the standard fixed-floating swap. We will discuss pricing of in-arrears swaps in Chapter 16.

Libor-in-arrears swaps are popular in upward-sloping interest rate curve environments, i.e. when long-tenor rates are higher than shorter-tenor ones. In such a scenario, the break-even fixed rate on the Libor-in-arrears swap tends to look more “attractive” than that of a standard fixed-floating swap, thus increasing the desirability of the swap to those seeking to receive fixed rate payments.

5.7 Averaging Swaps

Libor rates are not restricted to being observed on either the start date or the end of the pay period. A popular example is the *averaging swap*, i.e. a swap where the floating rate is determined as an average of Libor rate observations taken at regular intervals over each coupon period. For example, let $\{(t_{n,i}^f, t_{n,i}^s, t_{n,i}^e)\}_{i=1}^{K_n}$ be a collection of date triplets (fixing, start and end date) that define the rates to be used in calculating the payment in period n . Defining a set of weights $w_{n,i}$, $i = 1, \dots, K_n$, the floating rate \bar{L}_n for the period $[T_n, T_{n+1}]$ may be defined as

$$\bar{L}_n = \sum_{i=1}^{K_n} w_{n,i} L(t_{n,i}^f, t_{n,i}^s, t_{n,i}^e).$$

For the fixed rate swap payer, the averaging swap value is therefore

$$V_{\text{average}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (\bar{L}_n - k) \right). \quad (5.9)$$

As a rule, the weights $w_{n,i}$ sum up to 1, $\sum_{i=1}^{K_n} w_{n,i} = 1$; the weights usually reflect the number of days (using the appropriate day counting conventions) that a given rate $L(t_{n,i}^f, t_{n,i}^s, t_{n,i}^e)$ is supposed to be in effect. Computation of the valuation expression (5.9) can be done using techniques similar to those required for in-arrears swaps; see Chapter 16 for details.

Swaps linked to the average of the Federal funds rate are common examples of an averaging swap. Particularly noteworthy is the *Fed funds/Libor basis swap* which pays the average of the Fed funds rate (over a given period) against a payment based on a Libor rate for that period. This instrument is an example of a *floating-floating single-currency basis swap*, i.e., a swap that exchanges payments based on two different floating rates in the same currency. Closely related to Fed funds basis swaps are the *Fed funds futures* contracts traded on the Chicago Board of Trade (CBOT) exchange. These contract uses the 30 day running average of the Federal funds rate for settlement.

Remark 5.7.1. Going forward, in our product descriptions we shall normally assume that all cash flows pay at the end of the periods in which they fix. While this is common practice, as we have just seen the “pay-in-arrears” rule can be broken at will depending on the client’s needs — the only (self-evident) restriction is that payments should be fixed by the time they are made.

5.8 Caps and Floors

A firm with liabilities funded at a floating (i.e., Libor) rate is naturally concerned with the possibility that interest rates, and thus its interest rate

payments, may increase in the future. One way to immunize against this risk is to pay fixed on a fixed-floating interest rate swap, in effect turning floating rate payments into fixed ones. While this will guarantee a fixed rate for funding payments for the duration of the swap, it will also mean forgoing the possibility of benefiting from a potential future drop in rates. An *interest rate cap* is a security that allows one to benefit from low floating rates yet be protected from high rates. Similarly, for an investor with assets earning a floating rate, a low-rate scenario is unfavorable. An *interest rate floor* is an instrument designed to protect against low interest rates yet allow the holder to benefit from high rates.

Formally, a cap is a strip of *caplets*, call options on successive Libor rates, and a floor is a strip of *floorlets*, put options on successive Libor rates. We encountered caplets already in Section 4.5.1 and recall that this instrument pays

$$\tau_n (L_n(T_n) - k)^+$$

per unit notional at time T_{n+1} . Similarly, a floorlet pays

$$\tau_n (k - L_n(T_n))^+$$

per unit notional at time T_{n+1} . Then, N -period caps and floors have values at time t of

$$V_{\text{cap}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (L_n(T_n) - k)^+ \right),$$

$$V_{\text{floor}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (k - L_n(T_n))^+ \right).$$

By switching to the T_{n+1} -forward measure (see Section 4.2.2) for the n -th caplet/floorlet, the valuation formulas can be written in a more convenient form

$$V_{\text{cap}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}} \left((L_n(T_n) - k)^+ \right),$$

$$V_{\text{floor}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}} \left((k - L_n(T_n))^+ \right).$$

By Lemma 4.2.3, the Libor rate $L_n(\cdot)$ is a martingale under the T_{n+1} -forward measure. Hence, caplets/floorlets can be priced using “vanilla” models⁹, such as the log-normal Black model (see Remark 1.9.4).

⁹By a *vanilla* model we mean a model that specifies the dynamics (or just the terminal distribution) of only a single rate, or at most a few rates, in contrast to term structure models that specify consistent dynamics for the entire term structure of interest rates. Often vanilla models are borrowed from equity or FX modeling; having the underlying rate a martingale makes such borrowing painless. We discuss vanilla models in Chapters 7, 8, 9 and 17.

The OTC market in caps/floors is very liquid. While individual caplets/floorlets are not traded, caps/floors are available in a number of maturities. This allows the volatility information for individual forward Libor rates to be extracted from market quotes for caps/floors of different maturities, at least in principle¹⁰. Once extracted, these volatilities may be combined with the volatilities observed from European swaption quotes (see below), to form a set of market inputs to which interest rate models for exotics are calibrated.

5.9 Digital Caps and Floors

Digital caps and floors work like regular caps and floors, except that the n -th digital caplet pays

$$\tau_n \times 1_{\{L_n(T_n) > k\}}.$$

Similarly, the n -th digital floorlet pays

$$\tau_n \times 1_{\{L_n(T_n) < k\}}.$$

Digital caps and floors provide a leveraged way to bet on the future direction of interest rates, more so than through standard caps and floors.

5.10 European Swaptions

Caps and floors have an asymmetric exposure to interest rates, a characteristic used by both hedgers and speculators. A similar exposure profile is provided by options on swaps, the so-called *European swaptions*. A European swaption gives the holder a right, but not an obligation, to enter a swap at a future date at a given fixed rate. A *payer* swaption is an option to pay the fixed leg on a fixed-floating swap; a *receiver* swaption is an option to receive the fixed leg.

Assuming the underlying swap starts on the expiry date T_0 of the option (a typical situation), the payoff for a payer swaption at time T_0 then equals

$$V_{\text{swaption}}(T_0) = (V_{\text{swap}}(T_0))^+ = \left(\sum_{n=0}^{N-1} \tau_n P(T_0, T_{n+1}) (L_n(T_0) - k) \right)^+. \quad (5.10)$$

The value at an intermediate time t , $t < T_0$, must then equal

$$\begin{aligned} V_{\text{swaption}}(t) &= \beta(t) E_t (\beta(T_0)^{-1} V_{\text{swaption}}(T_0)) \\ &= \beta(t) E_t \left(\beta(T_0)^{-1} \sum_{n=0}^{N-1} \tau_n P(T_0, T_{n+1}) (L_n(T_0) - k) \right)^+, \end{aligned}$$

¹⁰This “volatility bootstrap” is by no means trivial; we discuss it in Section 16.2.

which, using (5.6), can be rewritten in the more compact form

$$V_{\text{swaption}}(t) = \beta(t) E_t \left(\beta(T_0)^{-1} A(T_0) (S(T_0) - k)^+ \right). \quad (5.11)$$

Moreover, switching to the annuity measure, also known as the *swap measure*, Q^A from Section 4.2.5, the swaption value can be expressed as

$$V_{\text{swaption}}(t) = A(t) E_t^A (S(T_0) - k)^+, \quad (5.12)$$

with the forward swap rate $S(\cdot)$ being a martingale in the swap measure Q^A ; see Lemma 4.2.4.

It is evident from (5.12) that a payer European swaption is a call option — and a receiver European swaption is a put option — on the forward swap rate, struck at the fixed rate of the swap. Hence, swaptions could be priced using a vanilla model (see footnote 9), such as the Black model or similar. Conversely, values of European swaptions can be translated into market-implied distributional characteristics of forward swap rates, a topic discussed at length in Section 7.1.2. In particular, it is universal practice to quote swaption prices in terms of *implied* Black volatilities, i.e. volatilities that recover market price when used in the Black formula. In some markets (e.g., the US), it is also common to quote implied *Gaussian* volatilities, defined in the same way with regard to a Gaussian (rather than log-normal) model for the distribution of interest rates, see (7.16).

The market in swaptions is very liquid, with many different option maturities and swap underlyings actively traded. To characterize the full universe of traded instruments, given a tenor structure (5.3) we consider swaptions of different expiries $\{T_n\}_{n=0}^{N-1}$ that can be exercised into swaps that start at T_n and cover m periods¹¹, i.e. their last payment date is T_{n+m} . For a convenient way to denote the various swaptions, recall definitions (4.8), (4.10) and (4.11) and introduce

$$A_{n,m}(t) = \sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}), \quad (5.13)$$

$$S_{n,m}(t) = \frac{\sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}) L_i(t)}{\sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1})}, \quad (5.14)$$

for $n = 0, \dots, N-1$, $m = 1, \dots, N-n$. Then the value of the (n, m) -swaption (a short-hand for an “ m -period swaption with expiry T_n ”) is equal to

$$A_{n,m}(t) E_t^{n,m} \left((S_{n,m}(T_n) - k)^+ \right),$$

where $E_t^{n,m}$ denotes time t expectation in the appropriate swap measure, $Q^{n,m}$. Note that in trader parlance, a (vanilla) T_n -maturity European swaption on a swap that runs from T_n to T_m is said to be a “ T_n into $T_{m+n} - T_n$ ”

¹¹A bit confusingly, such a swaption is often said to have *tenor* $T_{n+m} - T_n$, a characterization it inherits from the underlying swap rate.

swaption. For instance, a 5 year option on a 10 year swap would be a “5-into-10” (or “5y-into-10y”, or simply “5y10y”) swaption.

Clearly, when $m = 1$, the (n, m) -swaption reduces to a caplet (or floorlet) on the Libor rate $L_n(\cdot)$, so caplets and floorlets can be thought of as one-period swaptions. Whenever in this book swaptions are discussed or used, caplets and floorlets are thereby implicitly included. Collectively, all (n, m) -swaptions constitute *the swaption grid*.

Market quotes on swaptions, typically in terms of implied volatilities, in the swaption grid provide the most readily-available information on the volatility structure of interest rates. As swaptions in the grid cover overlapping sections of the term structure of interest rates, extracting clean volatility information from market quotes is a non-trivial exercise that forms the foundation for calibration of models used for exotic interest rate derivative pricing. We will have much to say about such *volatility calibration* later on.

While options on plain-vanilla swaps comprise the bulk of the liquid (“vanilla”) interest rate market, options on general swaps (i.e. on swaps with non-constant notional and fixed rates) also trade and are properly treated as exotic derivatives. Often, general swaps can be decomposed into baskets of standard swaps, in which case options on general swaps become *basket options*. Valuation of basket options requires information on the co-dependence structure of securities in the basket, information that is not readily available from the vanilla options markets. We demonstrate how to handle this complication in Section 19.4.

5.10.1 Cash-Settled Swaptions

The swaption contract discussed in the previous section involves *physical settlement*, in the sense that an actual interest rate swap is entered into, should the option be exercised at its expiry. Physically-settled swaptions are also known as *swap-settled swaptions*. An economically equivalent swaption contract is one that instead settles into a cash payment equal to the PV (present value) of the swap as observed at time T_0 . Indeed, for both types, the swaption payoff (for a payer) is given by

$$A(T_0)(S(T_0) - k)^+, \quad (5.15)$$

see (5.10) and (5.11). In the European markets, a third variety of swaptions is common, the so-called *cash-settled* swaptions. For this type of option, rather than entering into a swap, the option holder will receive a cash payout upon exercise. The settlement amount is calculated by a formula similar to (5.15), except the annuity $A(\cdot)$ is not calculated by (5.4), but instead by discounting fixed rate payments at the swap rate $S(T_0)$. Specifically,

$$V_{\text{css}}(T_0) = a(S(T_0))(S(T_0) - k)^+,$$

where

$$a(x) = \sum_{n=0}^{N-1} \frac{\tau_n}{\prod_{i=0}^n (1 + \tau_i x)}.$$

Notice that the cash settlement mechanism ensures a well-defined present value of the option payout, as long as the swap rate $S(T_0)$ is observable. In contrast, the value of exercise of a physically settled swaption — the computation of which requires knowledge of a strip of discount factors — may be estimated differently by different dealers, due to bid-ask spread effects and differences in curve building technology (see Chapter 6). Technically, however, the cash settlement mechanism induces certain valuation complications, and cash-settled swaptions cannot, strictly speaking, be considered vanilla options that can be priced using, e.g., a Black-type formula¹². This follows from the fact that in the measure associated with the deflator $X(t) = a(S(t))$, the swap rate $S(\cdot)$ is *not* a martingale, and certain drift adjustments are required. We discuss valuation of cash-settled swaptions in Section 16.6.12. As they are the most liquidly-traded OTC interest rate options in the European market, cash-settled swaptions still could (and should) be used to extract information on the volatility structure of interest rates; the procedure, however, is necessarily more involved.

5.11 CMS Swaps, Caps and Floors

As the market in plain vanilla swaps is both deep and very active, market quotes of corresponding swap rates can be used as “indexes”, i.e. market variables that can themselves be used in defining payoffs of other securities. The demand for such products is often driven by particular segments of fixed income markets. For example, mortgage lenders are primarily concerned with hedging interest rate risk arising from holding residential loans, some of which may have maturities as long as thirty years. Because of potential prepayments, the interest rate risk of a pool of such mortgages is often assumed to be closely connected to movements in the 10 year swap rate; hence, mortgage lenders are natural consumers of interest rate securities linked to the 10 year swap rate.

A constant-maturity swap (CMS) rate is defined as a break-even swap rate (see (5.5)) on a standard swap of a fixed maturity, e.g. 10 years or 30 years. A *CMS swap* works just like a standard fixed-floating (Libor) swap, except for the fact that floating leg payments are based on CMS, rather than Libor, rates. Formally, let $S_{n,m}(\cdot)$ be the m -period swap rate with the first fixing date T_n , as defined by (5.14). Then an m -period (payer) CMS swap’s value is given by

¹²Nevertheless, this practice has been widespread until recently, and may still be in use in some institutions.

$$V_{\text{cmsswap}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta (T_{n+1})^{-1} (S_{n,m}(T_n) - k) \right)$$

or, using the T_{n+1} -forward measure for each period,

$$V_{\text{cmsswap}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}} ((S_{n,m}(T_n) - k)).$$

While standard swaps can be valued solely from knowledge of the term structure of interest rates, CMS swaps require an interest rate model for valuation; we return to a complete discussion in Chapter 16.

CMS caps and floors are defined as strips of European options on CMS rates, just like regular caps and floors are strips of European options on Libor rates:

$$V_{\text{cmscap}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}} ((S_{n,m}(T_n) - k)^+),$$

$$V_{\text{cmsfloor}}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}} ((k - S_{n,m}(T_n))^+).$$

CMS caplets are related to European swaptions, as both are European-style options on swap rates. The connection between the two types of securities is, however, subtle, as we shall discuss later in this book.

5.12 Bermudan Swaptions

A Bermudan swaption is an option to enter into a fixed-floating swap on any (or any from a given subset) of its fixing dates. For a given tenor structure (5.3), the holder of a standard Bermudan swaption has the right to exercise it on any of the dates $\{T_n\}_{n=0}^{N-1}$. Once exercised on date T_n , say, the option goes away, and the holder enters the swap with the first fixing date T_n and the final payment date T_N . The period up to $T_0 > 0$ is known as the *lockout* or *no-call* period. In common jargon, a Bermudan swaption on, say, a 10 year swap with a 2 year lockout period (at inception) is known as a “10 no-call 2”, or “10nc2”, Bermudan swaption.

Formally, at time T_n , the value of a payer¹³, if exercised, is therefore

$$U_n(T_n) = \beta(t) \sum_{i=n}^{N-1} \tau_i E_{T_n} \left(\beta (T_{i+1})^{-1} (L_i(T_i) - k) \right)$$

$$= \sum_{i=n}^{N-1} \tau_i P(T_n, T_{i+1}) (L_i(T_n) - k).$$

¹³Upon exercise, the holder of a payer (receiver) Bermudan swaption will pay the fixed (floating) leg of the swap.

Here, $U_n(T_n)$ here denotes the *exercise* value of the Bermudan swaption; loosely speaking, a Bermudan swaption contract is an option to chose between $U_n(T_n)$ for different $n = 0, \dots, N-1$. More succinctly, we recall from Section 1.10 that the Bermudan option value at time T_n will be the maximum of $U_n(T_n)$ and the *hold value* $H_n(T_n)$, the latter defined as the value of a Bermudan swaption with the exercise dates $\{T_i\}_{i=n+1}^{N-1}$ only (compare to Sections 1.10 and 3.5).

Demand for Bermudan swaptions comes from different segments of fixed income markets. Mortgage companies use them to hedge pools of mortgages, with the flexibility of Bermudan exercise convenient in matching the uncertain timing of prepayments in mortgage pools. Investors seeking higher current income sell Bermudan-style options on swaps to increase the coupons they receive, as explained later in the context of callable Libor exotics. Bermudan swaptions are also used as hedges for callable coupon bonds.

While it may be tempting to think of Bermudan swaptions as straightforward generalizations of European swaptions, they are substantially more difficult to model and price. Indeed, it is fair to say that many valuation methods and techniques covered in this book were developed in response to the need to value and risk manage Bermudan swaptions. Bermudan swaptions are, by far, the most liquid exotic fixed income securities, with all interest rate dealers holding large inventories.

5.13 Exotic Swaps and Structured Notes

With market sophistication ever on the rise, clients demand increasingly complicated payouts, often in a familiar swap or bond format (although the appetite has waned somewhat post-crisis). In an *exotic swap*, a regular floating Libor leg is swapped against structured coupons that are allowed to be arbitrary functions of observed interest rates (such as Libor or CMS rates). A standard fixed-floating vanilla swap is an obvious and trivial example where the structured coupon simply is a fixed rate. A cap (or a floor) can be seen as another, less trivial, example. In particular, note that

$$\begin{aligned}(k - L_n(T_n))^+ &= \left((k - L_n(T_n))^+ + L_n(T_n) \right) - L_n(T_n) \\ &= \max(k, L_n(T_n)) - L_n(T_n),\end{aligned}$$

which demonstrates that a floor can be represented as an exotic swap in which a Libor rate is exchanged for a *floored payoff* $\max(k, L_n(T_n))$.

Exotic swaps often start their life as bonds, or notes, sold by banks to investors. In a structured note, the investor pays an up-front *principal amount* (e.g., \$10,000,000) to the issuer of the note, who in turn pays the investor a structured coupon, and repays the principal at the maturity of the note. The principal amount is invested by the issuer (or the trading

desk to which the issuer passes the note for risk management), and pays the Libor rate plus or minus a spread. From the perspective of the issuer (or the trading desk), the net cash flows of the note are those of an exotic swap.

In terms of valuation, if C_n is the structured coupon for the n -th period, the value of the exotic swap is equal to (from the perspective of structured leg buyer)

$$V_{\text{exotic}}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left(\beta(T_{n+1})^{-1} (C_n - L_n(T_n)) \right),$$

where we for brevity have assumed that both legs of the swap pay at the end date of each coupon period (see Remark 5.7.1). As discussed earlier, in this valuation equation, the coupon C_n can be a complicated function of interest rates, structured to reflect investors' views on the market, or to take advantage of current interest rate market conditions. For example, a floored payoff can be offered to an investor who believes that interest rates are poised for a fall in the future.

There is no universally agreed "taxonomy" for exotic swaps, but for our purposes we can distinguish between exotic swaps that are i) Libor-based, ii) CMS-based, iii) multi-rate, iv) range accruals, and v) generally path dependent. We proceed to described each type of swap in more details.

5.13.1 Libor-Based Exotic Swaps

In a Libor-based exotic swap, the structured coupon is a function of a Libor rate:

$$C_n = C_n(L_n(T_n)).$$

A large variety of structured coupons $C_n(\cdot)$ can be used. For example:

- A standard swap,

$$C_n(x) = k.$$

- Capped and floored floaters. For strike s , gearing g , cap c and floor f ,

$$C_n(x) = \max(\min(g \times x - s, c), f). \quad (5.16)$$

- Capped and floored inverse floaters. For spread s , gearing g , cap c and floor f ,

$$C_n(x) = \max(\min(s - g \times x, c), f). \quad (5.17)$$

- Digitals. For strike s and coupon k ,

$$C_n(x) = k \times 1_{\{x > s\}}$$

or

$$C_n(x) = k \times 1_{\{x < s\}}.$$

- “Flip-flops” or “tip-tops”. For strike s and two coupons, k_1 and k_2 ,

$$C_n(x) = \begin{cases} k_1, & x \leq s, \\ k_2, & x > s. \end{cases}$$

Different coupon types can be combined together to create new types of structured coupons.

A Libor-based exotic swap can usually be decomposed¹⁴ into a sum of simpler instruments such as ordinary swap floating legs, fixed legs, caps and floors, and digital caps and floors. Therefore, if the prices of these simple contracts are available in the market (as is typically the case), Libor-based exotic swaps can be perfectly replicated by a one-time transaction in market-available instruments, a strategy referred to as *static replication*. Hence, by themselves, these instruments rarely present major valuation challenges. They do, however, serve as building blocks for more complicated securities.

5.13.2 CMS-Based Exotic Swaps

The payoffs from the previous section can be applied to CMS, rather than Libor, rates. Structured coupons are then deterministic functions of CMS rates. If an m -period rate is used, then a structured coupon for period n can be defined by

$$C_n = C_n(S_{n,m}(T_n)),$$

with $C_n(x)$ as defined in the previous section.

CMS-based exotic swaps can be decomposed into linear combinations of CMS swaps and CMS caps/floors and rarely present any extra modeling difficulties beyond those already present in CMS swaps and caps.

5.13.3 Multi-Rate Exotic Swaps

Multi-rate exotic swaps differ from the structures in Sections 5.13.1 and 5.13.2 by referencing multiple market rates (Libor or CMS) for the calculation of structured coupons. The most common example is a *CMS spread coupon*. To describe this contract, let $S_{n,a}(\cdot)$ and $S_{n,b}(\cdot)$ be two collections of CMS rates, fixing on T_n , $n = 0, \dots, N - 1$, and covering a and b periods, respectively. A CMS spread coupon with gearing g , spread s , cap c and floor f is then defined by

$$C_n = \max(\min(g \times (S_{n,a}(T_n) - S_{n,b}(T_n)) + s, c), f).$$

A typical example would be a 10 year/2 year (often abbreviated as 10y2y) CMS call spread option where a is 40 (40 quarterly periods to cover 10 years) and b is 8, with the quarterly coupon given by

¹⁴Indeed, this is the case for all the payouts listed above, a fact that we invite the reader to verify.

$$C_n = \max(S_{10y}(T_n) - S_{2y}(T_n), 0)$$

(using somewhat loose notation). A relatively liquid broker market exists for spread options on Euro and US dollar CMS rates.

A more general example is obtained by using one of the payoff functions $C_n(x)$ defined in Section 5.13.1, applied to the spread $x = S_{n,a}(T_n) - S_{n,b}(T_n)$. In particular, digital and flip-flop CMS spread swaps are quite popular.

Multi-rate exotic swaps typically cannot be decomposed into “standard” instruments (such as vanilla swaps, caps, etc.). Therefore, they, as a rule, cannot be valued by replication arguments, and a valuation model is required. Such a model, however, does not always need to be a full-blown term structure model: we shall show later that some types of spread-linked payoffs can be efficiently valued and risk managed by vanilla models (see footnote 9).

It should be noted that more than two rates can be used in the definition of a coupon. For example, in the so-called *curve cap* one takes a standard capped and floored payoff on a Libor or CMS (or CMS spread!) rate — see (5.16) — and makes the cap c and the floor f functions of, potentially different, CMS spreads:

$$\begin{aligned} C_n(x) &= \max(\min(g \times x - s, c), f), \\ c &= \max(\min(g_1 \times (S_{n,a_1}(T_n) - S_{n,b_1}(T_n)) + s_1, c_1), f_1), \\ f &= \max(\min(g_2 \times (S_{n,a_2}(T_n) - S_{n,b_2}(T_n)) + s_2, c_2), f_2). \end{aligned} \quad (5.18)$$

5.13.4 Range Accruals

A *range accrual* structured coupon is defined as a given rate — fixed in the simplest case, but potentially a Libor, CMS or a CMS spread rate — that only “accrues” when a different *reference* rate is inside (or, sometimes, outside) a given range. So, let $R_n(t)$ be the payment rate and $X_n(t)$ be the reference rate, and let l be the low bound and u be the upper bound. A range accrual coupon then pays

$$C_n = R_n(T_n) \times \frac{\#\{t \in [T_n, T_{n+1}] : X_n(t) \in [l, u]\}}{\#\{t \in [T_n, T_{n+1}]\}}, \quad (5.19)$$

where $\#\{\cdot\}$ is used to denote the number of days that a given criteria is satisfied.

The most common choice of the payment rate $R_n(t)$ is either a constant or Libor, but a CMS rate or any other structured coupon rate are also occasionally used. The reference rate $X_n(t)$ can be any market-observable rate such as a Libor rate fixing at t , a CMS rate fixing at t , or even a CMS spread rate.

We note that a range accrual coupon can always be decomposed into simpler digital payoffs, because

$$\#\{t \in [T_n, T_{n+1}] : X_n(t) \in [l, u]\} = \sum_{t \in [T_n, T_{n+1}]} 1_{\{X_n(t) \in [l, u]\}}, \quad (5.20)$$

where the sum on the right-hand side is over all business days in the period. This decomposition is particularly useful for fixed rate ($R_n(T_n) \equiv k$) range accruals, as simple digital options can be priced directly from the market information on European options (see Section 7.1.2). For floating, or more complicated, range accruals the decomposition is useful but requires further work to turn it into valuation formulas — see Section 17.5 for further details.

The basic payout (5.19) can be extended to include more than one range condition. In a *dual range accrual*, the position of two different reference rates relative to the range are monitored, and (5.19) is generalized to

$$C_n = R_n(T_n) \frac{\#\{\{t \in I_n : X_{n,1}(t) \in [l_1, u_1]\} \diamond \{t \in I_n : X_{n,2}(t) \in [l_2, u_2]\}\}}{\#\{t \in I_n\}}$$

with $I_n = [T_n, T_{n+1}]$ and \diamond denoting either intersection \cap or union \cup . In the former case, one counts the number of days when *both* reference rates $X_{n,1}$ and $X_{n,2}$ are within their ranges; in the latter case, one counts the number of days when *either* of the two reference rates are within their ranges.

In a *curve cap range accrual*, the lower and upper bounds become functions of CMS spreads themselves, similar to (5.18).

A *product-of-ranges* range accrual multiplies up all range accrual factors to date to define the multiplier that is used for the current coupon, e.g.

$$C_n = R_n(T_n) \times Y_n,$$

where

$$Y_n = Y_{n-1} \times \frac{\#\{t \in [T_n, T_{n+1}] : X_n(t) \in [l, u]\}}{\#\{t \in [T_n, T_{n+1}]\}}, \quad (5.21)$$

$$Y_{-1} = 1.$$

5.13.5 Path-Dependent Swaps

The payoff (5.21) is an example of *path-dependence* in the payoff, where a coupon depends on rate observations from previous coupon periods. More commonly, path-dependence in exotic swaps is introduced by linking a structured coupon not only to interest rates observed during the coupon period, but to previous coupon(s) as well. This is often referred to as a *snowball* feature. The “original” snowball structure involved a coupon of an inverse floating type, with the n -th coupon C_n given by

$$C_n = (C_{n-1} + s_n - g_n \times L_n(T_n))^+. \quad (5.22)$$

Here $\{s_n\}$ and $\{g_n\}$ are contractually specified deterministic sequences of spreads and gearings. This type of a swap is sometimes also called a *ratchet* or a *ladder* swap.

The term snowball originates with the tendency of high initial coupons to spill into subsequent coupons, in a compounding or “snowballing” fashion”. Indeed, a little reflection reveals that a snowball has a highly leveraged exposure to its first few coupons, a feature that makes it more attractive to some investors, but also quite difficult to risk manage.

A large number of snowball-like payoffs have been created, often with “snow”-themed — and rather nonsensical — names, such as “snowrange”, “snowbear”, and “snowstorm”. For example, a “snowrange” combines a range-accrual feature and a snowball feature, in the following way

$$C_n = C_{n-1} \times Y_n + s_n + g_n \times X_{n,1}(T_n), \quad (5.23)$$

where $X_{n,1}(T_n)$ is some reference rate, and Y_n is a range-accrual factor depending on a second rate $X_{n,2}$,

$$Y_n = \frac{\#\{t \in [T_n, T_{n+1}] : X_{n,2}(t) \in [l, u]\}}{\#\{t \in [T_n, T_{n+1}]\}}.$$

The range accrual factor Y_n may be a product-of-ranges accrual factor, as in (5.21). Also, additional caps and/or floors are often added to the coupon (5.23).

Path-dependent swaps typically require a term structure model for valuation; for obvious reasons, Monte Carlo methods are often mandatory.

5.14 Callable Libor Exotics

5.14.1 Definitions

As described in Section 5.12, Bermudan swaptions are Bermudan-style options to enter a regular fixed-floating swap. If we alter the swap underlying the Bermudan swaption from a regular swap to an exotic swap (see previous section), then a so-called *callable Libor exotic* (CLE) is created. CLEs most often emerge as part of callable structured notes in which an issuer receives the principal from an investor and pays a structured coupon in return. In addition, the issuer has the right to cancel — or *call* — the note on a schedule of dates; typically, this call schedule will coincide with coupon fixing dates, after some initial lock-out (or *no-call*) period. Should a note be called by its issuer¹⁵, the principal is returned to the investor and no future coupons are paid.

A callable structured note is typically passed through by the issuer to an exotics trading desk (which could, but does not have to, be internal to the issuing bank) to deal with its risk management. Also, the principal is

¹⁵The call decision is most often made by the issuer’s swap counterparty who is actually managing the risk, see next paragraph.

invested and pays a Libor rate, plus or minus a spread depending on the cost of financing. From a trading desk perspective what is left is an exotic swap paying structured coupons and receiving Libor, plus a Bermudan-style right to cancel the swap. For clarity, Figures 5.1–5.3 list the cash flow diagrams of a callable structured note¹⁶.

Fig. 5.1. Callable Note: Flows at Inception

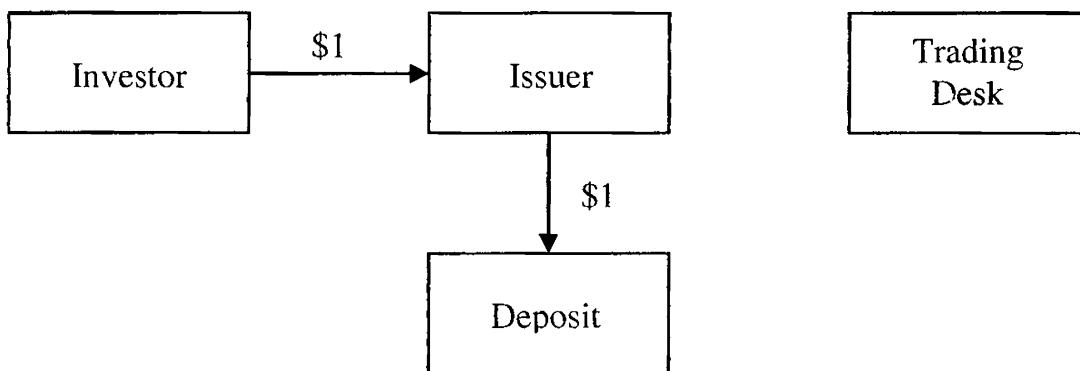
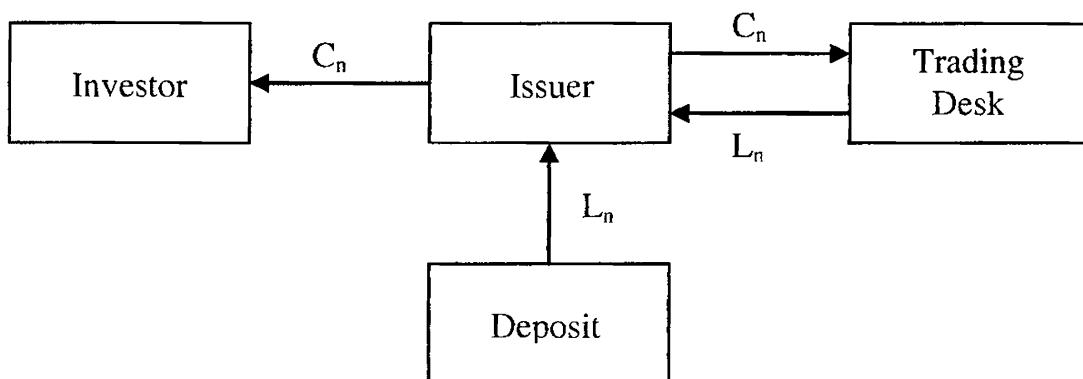


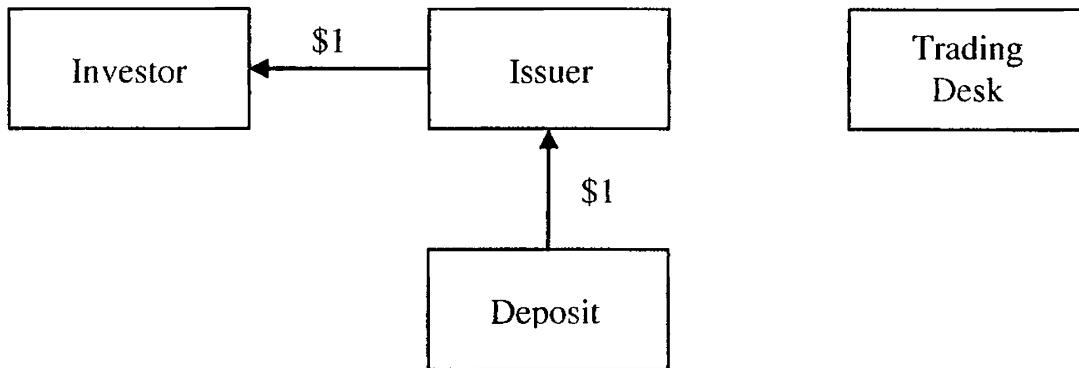
Fig. 5.2. Callable Note: Flows at Payment Times



Sometimes it is convenient to represent a cancelable exotic swap as a straight exotic swap, plus a Bermudan-style option to enter a reverse swap, i.e. a swap where legs are reversed relative to the original one. Beyond providing a break-down that is convenient for valuation purposes, this representation emphasizes the fact that the cancelability feature of a CLE benefits the party that owns it (typically a structured note issuer). Indeed, the feature is

¹⁶While, conceptually, the principal is deposited into a Libor-paying account, in practice it is used as part of cash management activities by the issuer. A structured note issuance program often provides cheaper funding to a bank than would be attainable by other means.

Fig. 5.3. Callable Note: Flows at Termination (Maturity or early Cancellation)



often added to a structured note as a way to offer a more attractive coupon to the investor, in return for the Bermudan-style option. Often, the coupons inside the non-call period are fixed rate coupons, and a typical way for the issuer to “pay” for the Bermudan option is to make these first coupons high, often much higher than the return available elsewhere. This “optical illusion” of high rate of return on investment is, at least in part, what drives investor interest in structured callable notes.

Consider a CLE on an exotic swap with structured coupons $\{C_n\}_{n=0}^{N-1}$. As for regular Bermudan swaptions, we denote the value of the exotic swap that one can exercise on date T_n by

$$U_n(T_n) = \beta(T_n) \sum_{i=n}^{N-1} \tau_i E_{T_n} \left(\beta(T_{i+1})^{-1} \times (C_i - L_i(T_i)) \right). \quad (5.24)$$

Here, we recall that from a trading desk prospective, the cancelability feature of CLE involves an option to enter a reverse swap, with receipt of structure coupons and payment of Libor. Hold values are also defined analogously to the Bermudan swaption case: the n -th hold value $H_n(T_n)$ is defined as the time T_n value of the CLE on the same exotic swap, but with exercise dates $\{T_i\}_{i=n+1}^{N-1}$ only. That is, $H_n(T_n)$ is the time T_n value of the CLE provided it has not been exercised on or before T_n .

5.14.2 Pricing Callable Libor Exotics

A significant part of this book is dedicated to efficient methods for pricing and risk-managing callable Libor exotics. To provide a brief preview of the difficulties involved, we notice that the call feature embedded in CLEs may suggest application of PDE methods, using backward induction arguments outlined in previous chapters. However, often a CLE has explicit path dependence that makes the application of PDE methods impractical. In other cases, as we shall describe in greater depth later in the book, models that admit an efficient PDE representation are often too inflexible for

application to anything but the simplest of CLEs. Hence, CLEs more often than not must be valued using Monte Carlo methods. We have seen a preview of how optimal exercise can be handled in Monte Carlo in Section 3.5; many more details are provided in Chapter 18.

5.14.3 Types of Callable Libor Exotics

Any exotic swap can be used as an underlying for a callable Libor exotic. For our purposes, the taxonomy of callable Libor exotics can follow closely that of exotic swaps, see Section 5.13. We can thereby distinguish various types of CLEs, e.g. Libor-based, CMS-based, multi-rate, callable range accruals, callable snowballs, and so on. Many variations on the basic CLE design exist, most of which are driven by a desire to increase the value of the option to cancel that the investor sells the trading desk, in order for a higher coupon to be paid. It is difficult to classify all the features that have been invented: we content ourselves with merely listing some of the more popular ones.

5.14.4 Callable Snowballs

A callable snowball is a CLE with a snowball (or snowrange, etc.) underlying. From a modeling prospective, they are notable for being one of the first widely popular instruments that combine both strong path-dependence and optimal exercise. It is possible to incorporate snowball-type path-dependence into a PDE framework by introducing auxiliary variables, following the principles of Section 2.7.5; Section 18.4.5 discusses details specific to snowballs. Alternatively — and often preferably — optimal exercise can be incorporated into the Monte Carlo method, as discussed in Section 3.5 and later on in Chapter 18.

5.14.5 CLEs Accreting at Coupon Rate

Typically the notional of the underlying swap of a CLE is fixed throughout the life of the deal, but it does not have to be. For instance, it is not uncommon for the notional to vary deterministically, e.g. increase or decrease by non-random additive or multiplicative amounts each coupon period. Such deterministic accretion/amortization rarely adds extra complications from a modeling prospective. Occasionally, however, a contract specifies that the notional of the swap accretes at the structured coupon rate, in which case the accretion rate will be random. For such CLEs, the exercise value in (5.24) must be amended. Specifically, if q_i is the notional in place for the period $[T_i, T_{i+1}]$, then q_i is obtained from the notional over the previous period q_{i-1} by multiplication with the structured coupon over the previous period. Formally,

$$U_n(t) = \beta(t) \sum_{i=n}^{N-1} \tau_i E_t \left(\beta(T_{i+1})^{-1} \times q_i \times (C_i - L_i(T_i)) \right),$$

$$q_i = q_{i-1} \times (1 + \tau_{i-1} C_{i-1}),$$

where the initial notional q_0 is contractually specified.

Interestingly, the random accretion feature above can be incorporated into a PDE-based scheme without any extra cost, see Section 18.4.5.

5.14.6 Multi-Tranches

The more optionality an investor can sell to the issuer, the better coupon she can receive. As described earlier, the option to call the note is already present in a callable structured note. Another option that is sometimes embedded is a right for the issuer to increase the size of the note, i.e. to put more of the same note to the investor, whether she wants it or not. The name of this feature, a “multi-tranche” callable structured note, originates with the fact that these possible notional increases are formalized as tranches¹⁷ of the same note that the issuer has the right to put to the investor. The times when the issuer has the right to increase the notional of the note typically come before the times when the note can be canceled altogether. Callability usually applies jointly to all tranches of the note.

By itself, the multi-tranche feature rarely presents modeling issues, although one must be mindful of it and plan for a pricing infrastructure that is flexible enough to handle it.

5.15 TARNs and Other Trade-Level Features

While sometimes the precise split is a little arbitrary, it is often helpful to think of a Libor exotic as being defined by

- A definition of its coupon, i.e. a formula that converts rates observed during a coupon period (and, sometimes, previous coupons) into the amount of money paid to the investor.
- A collection of trade-level features, i.e. features that cannot conveniently be expressed as coupon definitions, but instead “act” on the whole trade.

We have already seen examples of both features. For instance, a callable snowball CLE has a coupon definition given by the formula (5.22) on top of which callability has been added as a trade-level feature. In the next few sections we review some other trade-level features.

¹⁷“Slices” in French; here meaning “similar securities offered as part of the same transaction”.

5.15.1 Knock-out Swaps

A knock-out swap is just an exotic swap that disappears on the first fixing date on which some reference rate is above (or below) a given barrier. If the knock-out rate for the period n is denoted by $X_n(t)$, the coupon by C_n , the Libor rate by L_n , and the knock-out barrier by R , the value of a down-and-out¹⁸ knock-out swap is given by

$$V_{\text{KO}}(t) = \beta(t) E_t \left(\sum_{n=0}^{N-1} \tau_n \beta(T_{n+1})^{-1} \times (C_n - L_n(T_n)) \times \prod_{i=0}^n 1_{\{X_i(T_i) > R\}} \right).$$

5.15.2 TARNs

Callable structured notes have proved to be popular with investors, but suffer from the drawback that investors rarely know when the issuer will call the note — indeed, the decision to exercise a Bermudan-style option is driven by a model rarely accessible to the average investor. A relatively recent innovation, the *Targeted Redemption Note* (TARN), presents one possible solution to the problem. In a TARN (see Piterbarg [2004c]), the total investor return, defined as the sum of all structured coupons paid to date, is recorded over time. When the total return exceeds a pre-specified target level— hence the name of the structure — the note is terminated and the principal is returned to the investor.

As with callable notes, issuers do not keep TARN structures on their books, but swap them out with a trading desk. Since the principal payment from investors is invested at the Libor rate, to a trading desk a TARN looks like an exotic swap that knocks out on the total sum of structured coupons. Formally, let the structured coupon¹⁹ for the period $[T_n, T_{n+1}]$ be C_n . The coupon over the period $[T_n, T_{n+1}]$ is only paid if the sum of structured coupons up to (and not including) time T_n is below a total return R . Thus, the value of the TARN at time 0 from the investor's viewpoint is given by

$$V_{\text{tarn}}(t) = \beta(t) E_t \left(\sum_{n=1}^{N-1} \tau_n \beta(T_{n+1})^{-1} \times (C_n - L_n(T_n)) \times 1_{\{Q_n < R\}} \right), \quad (5.25)$$

$$Q_n = \sum_{i=1}^{n-1} \tau_i C_i, \quad Q_1 = 0.$$

¹⁸I.e. disappearing upon some variable breaching a barrier from above; compare to up-and-out options discussed in Chapter 2.

¹⁹In the original TARN product, an inverse floating coupon (5.17) was used, but any structured coupon can be employed.

We note that a TARN typically pays some fixed coupons to an investor before the knock-out feature starts, mirroring the non-call structure of CLEs, see Section 5.14. We omit these from the contract description as their present value can be computed statically off an interest rate curve separately, as the payments are known in advance.

Various features can be added to the description of the TARN we just described. For example, the last coupon, i.e. the coupon that pushes the total return over the target R , can be paid only partially to make the total return exactly R and not more. This feature is known as a *cap at trigger* or a *lifetime cap*. Also, if the total return over the life of the TARN never reaches the target R , a TARN equipped with a so-called *lifetime floor* will make a *make whole* payment at the TARN maturity to ensure that the total return exactly equals R , and not less. These features do not alter the general modeling framework for TARNs that we develop in later chapters, and we shall generally ignore them.

While it could be argued that a TARN is really just a swap with a different coupon definition (namely, $C_n \times 1_{\{Q_n < R\}}$)²⁰ we prefer to classify a TARN feature as trade-level, reflecting its historic importance and its relationship with the callability feature.

5.15.3 Global Cap

As discussed, a TARN can have a feature that restricts total return to an investor to be exactly the trigger level R . This feature can be decoupled from the TARN definition and used by itself, often called a *global cap*. Specifically, an exotic swap with a global cap R pays a structured coupon C_n to an investor until the sum of the coupons has reached the level R . Note that a swap typically does not terminate at this point, i.e. the trading desk will continue to receive Libor until the maturity of the trade or some other agreed termination event.

5.15.4 Global Floor

A *global floor* guarantees the investor a minimum cumulative return of the note. Specifically, if the sum of structured coupons paid to the investor does not reach a global floor value of F by the maturity of the deal, the issuer will pay a make whole amount to the client, equal to F minus the sum of actual coupons paid. The payment is made at the maturity of the swap or some other termination event (such as when the swap is canceled, in case of a callable global floor note).

²⁰The same could be said about knock-out swaps previously defined.

5.15.5 Pricing and Trade Representation Challenges

Trade-level features are often combined with each other, and with various coupon formulas. For example, one can be asked to price a callable, TARN’ed snowball with a global cap. As we shall argue in later chapters, the only modeling solution that is sufficiently scalable to accommodate such arbitrary combinations of various trade features involves a generic, flexible model that is calibrated to a broad collection of market volatility information. In such a framework, adding extra features to a trade is ultimately not much of a modeling problem, but could be a significant trade representation challenge. While outside the scope of this book, such a challenge should be addressed by a software framework that is flexible enough to represent any, current or future, trade-level features, and incorporate them into a pricing engine without significant extra effort. A successful implementation of such a trade representation framework requires careful planning and considerable investment. One fairly common route is to use a domain-specific programming language for trade representation, see for example Jones et al. [2000] for a commercially available version or Frankau et al. [2009] for an example of an in-house one.

5.16 Volatility Derivatives

In a nutshell, a *volatility derivative* is a contingent claim whose underlying is the volatility of a financial observable, rather than a financial observable itself. The simplest example of such a derivative is the *variance swap* (see Carr and Lee [2009b]), a structure that first emerged in equity and foreign exchange trading. In the last few years, similar ideas and structures have entered the fixed income derivative arena.

The demand for volatility derivatives in interest rates is driven by the same factors as in other asset classes; a common motivation is the desire of some market participants — hedge funds in particular — to have direct exposure to interest rate volatility, but not to the outright level of rates, say. In other cases, the product development follows the usual path of creating structured notes with appealing payoff profiles.

Different interest rates constitute different financial observables for defining a volatility, hence one needs to be rather specific when defining a volatility-linked payoff.

5.16.1 Volatility Swaps

A *volatility swap* in interest rates is a contract that measures realized volatility (or a related quantity) of a given rate, although it is structured somewhat differently from volatility swaps in equity or FX markets. Let

$X_n(t)$ be the rate used for period n ; then the most common coupon of a volatility swap is given by

$$C_n = |X_{n+1}(T_{n+1}) - X_n(T_n)|,$$

or a capped version

$$C_n = \min(|X_{n+1}(T_{n+1}) - X_n(T_n)|, c).$$

The value of the (structured) leg of a volatility swap measures realized variation of the rate $X_n(\cdot)$,

$$\begin{aligned} V_{\text{volswap}}(t) &= \beta(t) E_t \left(\sum_{n=0}^{N-1} \tau_n \beta(T_{n+1})^{-1} \times |X_{n+1}(T_{n+1}) - X_n(T_n)| \right) \\ &\quad + V_{\text{floatleg}}(t), \end{aligned} \tag{5.26}$$

where

$$V_{\text{floatleg}}(t) \triangleq \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) L_n(t) = 1 - P(0, T_N).$$

There are two common choices for the rate X_n . One choice, a *fixed-tenor* volatility swap, involves a swap rate of the same tenor on each of the fixing dates. Technically speaking, *different* swap rates are therefore used for different periods,

$$X_n(t) = S_{n,m}(t)$$

with a fixed value of m , the number of periods in the swap rate (see the definition (5.14)). For example, a rolling 10 year CMS rate could be used. The other choice, a *fixed-expiry* volatility swap, specifies the swap rate to have a fixed expiry and tenor, i.e.

$$X_n(t) = S_{K,m}(t).$$

With this definition, the volatility swap pays the absolute variation of a rate with the fixing date T_K and spanning m periods of the tenor structure $\{T_n\}_{n=0}^{N+m}$. Often $K = N$, so that the variability of the rate $S_{N,m}$ is measured over the whole of its life.

Recently, volatility swaps on CMS spread rates have appeared. As the name implies, they measure the variation of the spread of two rates, e.g. $X_n = S_{n,m_1} - S_{n,m_2}$.

5.16.2 Volatility Swaps with a Shout

Sometimes, the investor in a volatility swap is given an option to *shout*, that is to choose when the fixing of the rate occurs for the purposes of calculating the coupon payoff. In particular, the payoff of the n -th coupon is then

$$C_n = |X_{n+1}(\eta_n) - X_n(T_n)|,$$

where the random stopping time $\eta_n \in [T_n, T_{n+1}]$ is chosen by the investor coupon-by-coupon. For the uncapped version of this payoff, it is intuitively clear that it is always optimal to postpone the shout until the end of the period, i.e. $\eta_n = T_{n+1}$. So, the option given to the investor is actually worthless, while designed to appear to have some value. Interestingly, for the capped version, it is optimal²¹ to shout at the lesser of T_{n+1} and the first time that $|X_{n+1}(\eta_n) - X_n(T_n)| \geq c$, where c is the cap level. As a consequence, a capped volatility swap with a shout option is equivalent to a volatility swap with a barrier on each coupon:

$$\begin{aligned} C_n &= c \times 1_{\left\{\max_{t \in [T_n, T_{n+1}]} |X_{n+1}(t) - X_n(T_n)| \geq c\right\}} \\ &\quad + |X_{n+1}(T_{n+1}) - X_n(T_n)| \times 1_{\left\{\max_{t \in [T_n, T_{n+1}]} |X_{n+1}(t) - X_n(T_n)| < c\right\}}. \end{aligned}$$

This decomposition follows from results in Broadie and Detemple [1995] and is discussed in more detail in Chapter 20. The fact that one can replace an optimal exercise feature with a known static barrier is quite convenient and allows for easy Monte Carlo valuation.

5.16.3 Min-Max Volatility Swaps

The structured coupon for a min-max volatility swap is given by

$$C_n = M_n - m_n,$$

where

$$\begin{aligned} M_n &= \max_{t \in [T_n, T_{n+1}]} X_n(t), \\ m_n &= \min_{t \in [T_n, T_{n+1}]} X_n(t). \end{aligned}$$

The coupon represents the spread between the maximum and the minimum values that a given rate achieves during a coupon period.

While the two products appear quite different at a first glance, there is an interesting connection between min-max and standard volatility swaps. We shall explore this further in Chapter 20.

5.16.4 Forward Starting Options and Other Forward Volatility Contracts

The value of a standard European swaption and a standard fixed-expiry volatility swap are both linked to the volatility of the swap rate over its

²¹Ignoring some small convexity effects and the difference between discrete vs. continuous shout option rights.

entire life, i.e. from the valuation date to the fixing date of the swap rate. Some clients, however, prefer securities that are linked to the volatility of a swap rate as measured over only a sub-period of this time; in effect, the clients want exposure to what is often known as *forward* volatility. Precise definitions and the importance of forward volatility are left for future chapters; for now we content ourselves by listing a few relevant varieties of forward volatility derivatives.

Midcurve swaptions are swaptions whose expiry T^e is strictly before the fixing date T_0 of the underlying swap rate. Their value depend on the volatility of the swap rate over the period $[t, T^e] \subset [t, T_0]$.

Given a swap rate $S(\cdot)$ with the fixing date T_0 , and a date $T^s < T_0$, a *forward-starting swaption straddle*²² is given by the payoff

$$A(T_0) \times |S(T_0) - S(T^s)|$$

paid at T_0 . Essentially, this contract is a combination of a receiver and a payer swaption, both of which will have their strikes fixed at time T^s to the then-prevailing level of the underlying swap rate. That is, the contract pays the value of the at-the-money straddle on the rate $S(\cdot)$ at time T^s with expiry T_0 . The value of a forward-starting swaption straddle is driven by the volatility of the swap rate over the period $[T^s, T_0]$.

Recall that European swaptions are typically quoted in terms of their implied volatilities. Due to this convention, some clients find the forward starting straddle too indirect and instead want to receive implied volatility itself. Particularly popular are contracts that pay implied Normal²³ volatility as defined on p. 204. Fortunately, the at-the-money Normal volatility has a direct relationship to the swaption price, and the payoff of an *implied Normal volatility contract* is

$$\sqrt{\frac{\pi}{2(T_0 - T^s)}} \times |S(T_0) - S(T^s)|$$

paid at T_0 . Apart from the factor $A(T_0)$ and a non-consequential scaling factor, the payoffs of a forward-starting straddle and the implied Normal volatility contract are identical. The differences in their prices are just a matter of a minor convexity correction, a topic we return to in Chapter 20.

²²The term *straddle* is used to denote the sum of a put and a call option with identical strikes.

²³Contracts paying implied log-normal, or Black, volatility are possible but less common, due to the common perception that interest rates are more Gaussian than log-normal, i.e. the implied Gaussian volatility is less sensitive to the changes in the level of interest rates than the implied Black volatility.

5.A Appendix: Day Counting Rules and Other Trivia

In this appendix, we very briefly cover some of the finer details of how schedules are constructed and how interest rate payments accrue under market conventions. We generally ignore these details in the main body of the book, and our treatment here only scratches the surface. For a full account, see Mayle [1993] or Stigum and Robinson [1996].

5.A.1 Libor Rate Definitions

Consider the 6 month Libor rate L fixing at time T . According to (4.2), we would compute this rate as simply

$$L(T) = L(T, T, T + 1/2) = (P(T, T + 1/2)^{-1} - 1) / \tau, \quad \tau = 1/2. \quad (5.27)$$

In reality, this computation ignores a number of quoting conventions. First, a 6 month USD Libor rate that fixes at time T , does not truly cover an *accrual period* of $[T, T + 1/2]$. Instead, the start date T^s of the accrual is set to be $T^s = T + \delta^s$, where δ^s is a delay of two business days²⁴. In other words, the quoted spot Libor rate is in actuality based on a *forward starting* CD that is entered into with time lag of δ^s after the quotation date T . As for the end date T^e of the accrual period, it is normally determined by counting 6 months ahead starting from T^s , adjusting the resulting date to ensure that it is a valid business day. The precise mechanism used to make such a business day adjustments of T^s is determined from a *date rolling convention*. For USD Libor, one always uses the so-called “Modified-Following” convention where weekend or holiday dates are rolled forward to the next business day, unless doing so would cause T^e to lie in the next calendar month, in which case the payment date is rolled to the previous business day. Other rolling conventions are discussed in Mayle [1993] and Stigum and Robinson [1996].

Once the correct accrual period $[T^s, T^e]$ has been determined, to compute rate accrual it remains to compute the proper *year fraction* (or *accrual factor* or *day count fraction*) τ representing how many whole years are spanned by $[T^s, T^e]$. For the purposes of our book, we normally write simply $\tau = T^e - T^s$, but a little thought shows that expressions like this are ambiguous when T^e and T^s are thought of as actual (discrete) calendar dates, rather than as arbitrary numbers on the real line. For instance, given the existence of leap years, how many days are there in a standard year? For quant purposes, it is common to assume that a calendar year has 365.25 years, such that $T^e - T^s$ is obtained by simply counting the number of days between T^s and T^e and then dividing this number by 365.25. This “convention” is sometimes known as Actual/365.25 (or sometimes just A365.25), and is rarely, if ever, used for actual market quoting purposes. Instead, for quotation of Libor

²⁴Libor rates in other currencies may have different delays. For instance, GBP Libor has zero business day delay.

rates the standard is to use an Actual/360 (A360) convention, where the number of days between T^s and T^e are converted to a year fraction by dividing by 360. As a consequence, the true value of τ used for 6 month Libor quotation purposes is typically slightly larger than 1/2. For additional year-count conventions (of which there are many), see Mayle [1993] and Stigum and Robinson [1996].

Due to the quoting standards used in real Libor markets, the relationship between discount bonds and quoted Libor rates is more complicated than (5.27). Specifically, if $L_{\text{mkt}}(T)$ represents the true quoted 6 month Libor rate at time T , we instead have

$$L_{\text{mkt}}(T) = \left((P(T, T^e)/P(T, T^s))^{-1} - 1 \right) \frac{360}{D(T^s, T^e)}, \quad (5.28)$$

where by $D(T^s, T^e)$ we denoted the number of days between T^s and T^e according to the convention used. Notice in particular how the formula now involves a forward starting zero-coupon bond $P(T, T^s, T^e) = P(T, T^e)/P(T, T^s)$, as a reflection of the settlement delay δ^s . Using existing “idealized” Libor rate notation (see (4.2)), we may write this expression as

$$L_{\text{mkt}}(T) = L(T, T^s, T^e) \times \frac{360}{365.25} \approx L(T, T^s, T^e) \times 0.986.$$

The difference between $L_{\text{mkt}}(T)$ and $L(T)$ is small enough for us to ignore it in most of this book, but any real system implementation obviously should use precise day counting rules when computing Libor fixings.

5.A.2 Swap Payments

The payments on swaps (and other instruments, such as CDs and FRAs) are subject to similar conventions as the Libor rate. Consider for instance a standard fixed-for-floating interest rate swap issued at time t (today). First, a²⁵ schedule $\{T_i\}_{i=0}^N$ for interest rate accrual must be constructed, starting from a given base frequency of the swap (e.g.: semi-annual). As was the case above, the schedule normally starts one or two business days after time t , i.e. $T_0 = t + \delta_0$ where δ_0 is some contractually specified delay. Date T_0 is known as the *effective date* of the swap. Given T_0 , the remaining T_i , $i = 1, \dots, N$, are computed by first laying out “unadjusted” dates according to the swap base frequency, and then applying a date rolling convention (typically Modified-Following) to each of the dates. As part of the swap contract, associated with each accrual period $[T_i, T_{i+1}]$ are then:

- A fixing date T_i^f : the date on which the floating leg index (Libor, most often) is observed. Typically T_i^f is two business days before time T_i .

²⁵We assume that the fixed and floating legs pay interest on the same schedule, but in reality, this may not be the case. For instance, in USD, the standard frequency for the fixed leg is six months, and three months for the floating leg.

- A payment date T_i^p : the date on which the swap payments are made. Normally $T_i^p = T_{i+1}$, but it is not uncommon to have payment delays of 1 or 2 business days after T_{i+1} .
- A fixed leg year fraction τ_i^{fix} : the year fraction used to determine the payment at time T_i^p on the fixed leg. In the US, the most common convention for the fixed leg is²⁶ 30/360.
- A floating leg year fraction τ_i^{flt} : the year fraction used to determine the payment at time T_i^p on the floating leg²⁷. In the US, the most common convention for the floating leg is Actual/360.

At time t , the value of a payer swap paying a fixed coupon c against Libor is therefore

$$V_{\text{swap}}(t) = \sum_{i=1}^N P(t, T_i^p) E^{T_i^p} \left(\tau_i^{flt} L_{\text{mkt}}(T_i^f) - \tau_i^{fix} c \right),$$

where we have used the T_i^p -forward measure to state the valuation, and where L_{mkt} is defined in (5.28). In the book, we normally simplify this to

$$V_{\text{swap}}(t) = \sum_{i=1}^N P(t, T_i) \tau_i E^{T_i} (L(T_{i-1}) - c), \quad \tau_i = T_i - T_{i-1}.$$

²⁶When counting days in the 30/360 convention, each month is assumed to have 30 days. The number of days used to determine interest rate accrual will therefore differ from the *actual* number of days (which distinguishes 30/360 from Actual/360).

²⁷Another, often ignored, complication with the floating leg is that the periods that define the (payment) year fractions τ_i^{flt} are sometimes slightly different from those that define forward Libor rates, due to certain conventions surrounding date adjustments. Again, see details in Mayle [1993] or Stigum and Robinson [1996].