OPEN PRESS TiU

J.M. Schumacher

INTRODUCTION TO FINANCIAL DERIVATIVES

Modeling, Pricing and Hedging

INTRODUCTION TO FINANCIAL DERIVATIVES

Modeling, Pricing and Hedging

by J.M. Schumacher

ISBN: 978-94-6240-611-7 (Interactive PDF)
https://digi-courses.com/openpresstiu-introduction-to-financial-derivatives/

ISBN: 978-94-6240-612-4 (Paperback)

Published by: Open Press TiU
Contact details: info@openpresstiu.edu
https://www.openpresstiu.org/

Cover Design by: Kaftwerk, Janine Hendriks
Layout Design by: J.M. Schumacher
Contact details: J.M.Schumacher@uva.nl

OPEN PRESS Tilburg University 2020

Preface

The material in this Open Press textbook originates from a course that I have taught at Tilburg University for more than ten years, until my retirement in 2016. The course was designed to provide students with an introduction to continuous-time models that are used to analyze derivative contracts in finance and insurance, as part of the MSc program in Quantitative Finance and Actuarial Science. Students in the QFAS master's program come in from the bachelor's program in Econometrics and Operations Research at Tilburg University, but also from comparable programs at universities elsewhere in the Netherlands as well as from abroad. The intended audience of the course therefore consists of students with a solid background in standard calculus, linear algebra, and probability, but not necessarily with prior exposure to stochastic calculus. The main ingredients in the course are:

- an introduction to stochastic calculus at a semi-rigorous level, without using measure-theoretic probability at the level of filtrations
- a discussion of financial modeling in continuous time, covering basic notions such as absence of arbitrage and market completeness
- an exposition of computational methods that are used in the field, analytical as well as numerical, with hands-on experience in the form of programming exercises
- somewhat more extensive coverage of a particular domain that is important in finance and insurance, namely the term structure of interest rates.

There is also a "hidden curriculum": enhancing students' appreciation of the subtlety and the richness of the interaction between mathematics and the real world.

Since my position at Tilburg University ended, time has not stood still, and the structure of the courses in the MSc program on Quantitative Finance and Actuarial Science has not remained the same. The material in the course as I taught it is still part of the program, but is included now partly in a concentrated course on stochastic calculus, and partly in a new course which also includes additional topics. The present text, based on the notes that I have written and expanded over the years, may still serve as support for students in the QFAS program, as well as

for students elsewhere who are are looking for an introduction to continuous-time financial modeling.

In the Open Press edition, the most recent version of course syllabus that I used has been expanded with material from several sources, including the set of slides that I developed for the course, as well as exam questions. I also reorganized the material somewhat and made various smaller changes, some motivated by things I have learned since retirement. The programming exercises in the original course were based on Matlab, since this was also used in the curriculum of the BSc program in Econometrics and Operations Research. I have chosen in the present textbook to keep the code examples in Matlab, while adding an appendix in which the meaning of the Matlab commands is explained to facilitate translation to other languages such as R, Julia, or Scilab.

Most of the material in the book falls in the category "general knowledge", but in Appendix A there are references for a few specific items. The following books contain source material and are excellent further reading for students who want to go beyond the introductory material that is presented here. Due in particular to the avoidance of filtrations, some of the theorem statements in this book are lacking in precision, and some of the proofs are lacking in rigor; for improvements in these respects as well, I would like to refer the reader to the sources below.

General:

Tomas Björk, *Arbitrage Theory in Continuous Time* (4th ed.), Oxford University Press, Oxford, UK, 2020.

Ioannis Karatzas and Steven E. Shreve, *Methods of Mathematical Finance*, Springer, New York, 1998.

Cornelis W. Oosterlee and Lech A. Grzelak, *Mathematical Modeling and Computation in Finance. With Exercises and Python and Matlab Computer Codes*, World Scientific, London, 2020.

Andrea Pascucci, *PDE and Martingale Methods in Option Pricing*, Springer, Milan, 2011.

Albert N. Shiryayev, *Essentials of Stochastic Finance. Facts, Models, Theory*, World Scientific, Singapore, 1999.

Chapter 1:

Peter L. Bernstein, *Capital Ideas*, The Free Press, New York, 1992.

Perry Mehrling, *Fischer Black and the Revolutionary Idea of Finance*, Wiley, Hoboken, NJ, 2005.

Chapter 2:

Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus* (2nd ed.), Springer, New York, 1991.

Fima C. Klebaner, *Introduction to Stochastic Calculus with Applications* (2nd ed.), Imperial College Press, London, 2005.

Philip Protter, *Stochastic Integration and Differential Equations. A New Approach*, Springer, Berlin, 1990.

Chapter 3:

Freddy Delbaen and Walter Schachermayer, *The Mathematics of Arbitrage*, Springer, Berlin, 2006.

Chapter 4:

Yue-Kuen Kwok, *Mathematical Models of Financial Derivatives*, Springer, Singapore, 1998.

Chapter 5:

Damiano Brigo and Fabio Mercurio, *Interest Rate Models—Theory and Practice. With Smile, Inflation and Credit* (2nd ed.), Springer, Berlin, 2006.

Chapter 6:

Daniel J. Duffy, *Finite Difference Methods in Financial Engineering. A Partial Differential Equation Approach*, Wiley, Chichester, UK, 2006.

You-lan Zhu, Xiaonan Wu, and I-Liang Chern, *Derivative Securities and Difference Methods*, Springer, New York, 2004.

Chapter 7:

Paul Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.

The literature is extensive and the above just represents a sample. In particular, there are many books covering application areas and extensions such as credit risk, transaction costs, portfolio management, and so on.

Over the years, I have received many comments on my course notes, from the TA's who worked with me, as well as from students who followed the course. I may not recall all exchanges, but let me at least mention Anton van Boxtel, Justinas Brazys, Renxiang Dai, Sebastian Gryglewicz, Fei Jia, Simon Polbennikov, Krzysztof Postek, Andreas Würth, Ran Xing, and Evren Yurtseven. I am grateful for their support. Also, I would like to thank my colleagues Bertrand Melenberg and Nikolaus Schweizer at Tilburg University who very competently responded to the task of teaching financial models to new generations of students, and who provided me with useful suggestions for the editing of the course notes. I am thankful as well to Daan Rutten for his suggestion to include the course notes in the Open Press series

of Tilburg University. My gratitude goes moreover to Wikipedia for making it easy to add some basic biographic notes on historical figures that are mentioned in the text.

The mathematical theory of derivatives is sometimes referred to as "rational option pricing". Indeed the theory could be compared to rational mechanics, the scientific discipline that speaks of point masses, weightless inextensible cords, and frictionless pulleys. A certain amount of idealization is involved; a large amount, perhaps. Models are confined to a certain domain of validity, and even within this domain they are not fully accurate. Nevertheless, the theory is meaningful, when applied with an understanding of its limitations. In the sometimes dazzling and overheated environment of finance, mathematical models provide much needed guidance. I hope the present text will help the reader to enjoy the cool world that has been created by the arbitrage theory of financial markets.

Hans Schumacher

Amsterdam, August 2020

Contents

This page intentionally left blank

Chapter 1

Introduction

1.1 The origins of the Black-Scholes formula

The Black-Scholes equation appears in a paper by Fischer Black and Myron Scholes that was published in 1973 in the *Journal of Political Economy*. Fischer Black has stated in a later publication that he had arrived at the equation already in 1969, but at the time was unable to solve it, even though he tried really hard. He writes: "I stared at the differential equation for many, many months. I made hundreds of silly mistakes that led me down blind alleys. Nothing worked."

Fischer Black had come into economics from an unusual angle. He entered Harvard University in 1955 as a physics student, but switched to applied mathematics for his graduate program. The PhD thesis that he completed in 1964 was on artificial intelligence, showing the design of a question answering machine. He subsequently joined the consulting firm Arthur D. Little, with the idea of helping businesses to make better use of their computers. It was there that he became interested in portfolio management and started reading the works of people such as Jack Treynor, one of the early proponents of the Capital Asset Pricing Model.

Treynor had published a paper in 1965 in the *Harvard Business Review*, in which he argued that there should be an adjustment for risk in assessing the performance of portfolio managers, since, due to the presence of a risk premium, more risky portfolios will on average have better returns than less risky portfolios. Fischer Black liked the "cruel truth", as he called it, that higher average return only comes at the expense of higher risk. He tried to apply the idea in several areas that interested him, such as monetary theory, business cycles, and the pricing of options and warrants.

Warrants are financial instruments that are similar to options: they give the right, during a certain period, to buy a given number of units of stock of a certain company at a stated price. The difference is that warrants are issued by the same company that also issues the underlying stock, whereas options are traded on an

exchange; for the purpose of pricing, however, this is inessential. During the 1960's warrants were more liquidly traded than options, so that papers discussing the pricing of such instruments were usually stated in terms of warrants rather than options. Among those who were interested in finding option pricing formulas was Paul Samuelson, one of the great minds of the 20th century, who in 1970 became the first American to receive the Nobel Prize in Economics.

Samuelson had done a bit of trading in warrants on a private account already since 1950, without making a lot of money though. Around 1952 he became aware of the work of the French trader and mathematician Louis Bachelier, who had connected the theory of Brownian motion with financial markets in his thesis presented at the Sorbonne in Paris in the year 1900. Even earlier, in 1880, the Danish actuary Thorvald Thiele published a paper on the least-squares method in which the stochastic process appears that we now call the "Wiener process" or "Brownian motion". Bachelier however was not aware of this work and developed the theory completely by himself, including the connection to partial differential equations which was to be rediscovered, again independently, in 1905 by none other than Albert Einstein. Options were traded at the Bourse at the time, and Bachelier derived an option pricing formula.

It was not only the option pricing formula that drew Samuelson's attention, but also the mathematical setting that Bachelier had used. Samuelson noted that the Brownian motion process as used by Bachelier (also known as arithmetic Brownian motion) would not be suitable as a model for stock prices, since it may well take negative values. Famously commenting that "a stock might double or halve at commensurable odds", Samuelson proposed a model in which the *logarithm* of the stock price follows a Brownian motion process, rather than the price itself. Thus appeared the geometric Brownian as a model for stock prices. Nowadays this model is usually referred to as the Black-Scholes model, since it serves as the basis for the Black-Scholes equation and the Black-Scholes formula for option prices, but it would actually be more appropriate to refer to it as the Bachelier-Samuelson model, since it arose as Samuelson's modification of Bachelier's original proposal for the modeling of stock prices. We can then still abbreviate it as the BS model.

The theory of Brownian motion was made mathematically rigorous in the 1930's by Norbert Wiener, and during the 1940's and 1950's the theory was expanded to a great extent by Kiyoshi Itô, who developed a stochastic calculus that could be used for instance to formulate stochastic differential equations. Samuelson, not feeling quite confident in the use of the new calculus himself, wrote a paper on the pricing of warrants in 1965 in collaboration with Henry McKean, his colleague from the MIT mathematics department who in the same year published a book on diffusion processes jointly with Itô. Despite the strong mathematical foundation, the pricing formula that Samuelson obtained in this paper was still not satisfactory, since

it contained some undetermined parameters. In the 1960's, several other pricing formulas were proposed, which however all suffered from the same problem.

Samuelson was well aware of the deficiencies of his formula. Looking for someone who could support him in the further mathematical developments that would be needed, he was happy to notice among the participants in his graduate course in 1967 a student who had just come in from California Institute of Technology as a result of a switch from applied mathematics to economics. In the spring, Samuelson hired the student, whose name was Robert C. Merton, as his research assistant, and in the summer he proposed that they would write a joint paper on the pricing of options. The paper appeared in 1969; it eliminated the undetermined parameters of Samuelson's earlier paper, but only at the expense of invoking an explicit description of the preferences of agents by means of utility functions. In October 1968, when Samuelson was announced to deliver the main lecture at the inaugural session of the MIT-Harvard Joint Seminar in Mathematical Economics, he surprised the assembled luminaries by instead giving the floor to his 24-year-old PhD student, in order to present their joint paper on option pricing. Merton later recalled that this experience at once cured him from any trepidation for audiences.

Myron Scholes arrived in the Boston area in the fall of 1968 as a starting assistant professor at MIT's Sloan School of Management, having just completed the PhD at the University of Chicago under the direction of Merton Miller. One of the people he made contact with in his new environment was Fischer Black, who was a regular visitor at Franco Modigliani's Tuesday night finance seminars at MIT, and whose office at Arthur D. Little was located close to the MIT campus. When Wells Fargo, one of the most innovative banks at the time, offered Scholes a consulting position, he suggested that they would hire Fischer Black as well. As a result Black and Scholes came to meet regularly, be it no longer at Arthur D. Little but rather at Black's own consulting practice which he had started after quitting from his job at ADL.

The two men talked about many things, but not about options at first. Then, some time in 1969, Black showed the equation he had derived to Scholes, and discussed with him the remarkable fact that the expected return on the underlying stock plays no role in it. From this observation, they concluded that candidate solutions to the equation might be found from simplified versions of the option pricing formulas that were already around in the literature. And indeed, working from a formula that was developed by a Yale University graduate student, they arrived at the solution. They had found an option pricing formula that, unlike its competitors, was stated directly in terms of observable quantities.

Fischer Black had arrived at his option pricing equation through an application of CAPM. When Bob Merton came to know about the equation, following a presentation by Scholes at the second Wells Fargo Conference on Capital Market Theory

in July 1970, he was skeptical. He couldn't believe that a static theory like CAPM could be reasonably combined with a theory of continuous or near-continuous trading. Thinking about it some more, he found a different argument leading to the same equation. On a Saturday afternoon in August, he made a phonecall to Scholes and said: "You're right."

As they say: the rest is history. Black and Scholes wrote their paper on the option pricing formula and submitted it to the *Journal of Political Economy* where it was promptly rejected, without even being sent out for review. Subsequently they sent their paper to the *Review of Economics and Statistics*, only to have it returned in the same way. At that point, Scholes' former PhD advisor Merton Miller and his colleague Eugene Fama stepped in; they convinced the editors of *JPE* that the paper might be worthwile after all. The paper was accepted subject to revision in August 1971, and it finally appeared in 1973, as it happened one month after the Chicago Board Options Exchange had opened for business. Soon, the *Wall Street Journal* would carry advertisements for calculators with the Black-Scholes formula built in.

The main argument presented for the Black-Scholes equation in the 1973 paper is the one that was provided by Merton. Black's original argument is given as an "alternative derivation". Merton provides yet another derivation in a paper published in 1977, which is only for the better, since the argument as used in the 1973 paper would be considered rather dubious by current standards. Major steps towards the completion of the theory were taken by Michael Harrison together with David Kreps in 1979 and together with Stanley Pliska in another paper published in 1981. In these papers one finds the notions of "self-financing strategy" and "equivalent martingale measure" that are lacking from the original option pricing papers, and that are essential for a full development of the theory even though Harrison and Kreps themselves refer to the EMM as a "somewhat abstruse concept". Other researchers have expanded the theory further, both strengthening its foundations and extending widely its domain of applications.

Fischer Black died of cancer in 1995. Myron Scholes and Robert Merton received the Nobel Prize in Economics in 1997. These three men have been pivotal in the development of a theory that has fundamentally transformed the world of finance.

1.2 Assets and self-financing strategies

1.2.1 Basic assumptions and notation

Money that is not needed for immediate consumption must be stored for later use. It may be kept in the form of cash, or in a savings account at a bank; it may be invested in government bonds, corporate bonds, stocks, gold, rare stamps, or in one

of the countless other investment opportunities that the world has to offer. Any item that can be used to store value will be referred to as an *asset*. Some assets are safe in the sense that their future value can be predicted quite accurately; other assets are risky and may bring large gains or severe losses. While the word "value" is often used in daily life for other things besides financial value, this book concentrates on the role of assets in finance. The value of an asset is therefore taken to be the price for which it can be bought or sold, and the terms "value" and "price" will be used interchangeably.

To facilitate the development of the theory, it is convenient to use the following assumptions.

(i) Assets are measured in units; the price of an asset refers to the price per unit. The price of c units is equal to c times the price of one unit. Prices are defined unambiguously at any point in time.

(ii) The value of a combination of assets (a *portfolio*) is the sum of the values of its constituent parts.

(iii) Assets can be traded freely, without transaction costs, at any time and in any quantity. The buying price is the same as the selling price.

(iv) From the point of view of an individual investor, the evolution of asset prices is an exogenous process which cannot be manipulated. In particular, the price process is not impacted by the investor's trades.

(v) Holding a fixed quantity of an asset brings no costs or dividends, other than gains or losses through value changes which are realized at the time at which the asset is sold.

The first four items are idealizing assumptions, which are quite helpful in the construction of mathematical models for the analysis of financial contracts. Of course it needs to be recognized that in reality trading takes place in a market environment which operates according to certain rules, that usually there is a bid-ask spread, that large trades in a given asset will impact its price, and so on. Researchers have constructed a variety of models that take these features into account; however, these models fall outside the scope of this book. Assumption (v) is of a different nature; one can make sure that this assumption is satisfied by incorporating any costs or dividends into the definition of the asset (see Section 4.3.4).

According to assumptions (i) and (ii) above, the *value* of a portfolio at any given time t is given by the formula

$$V_t = \sum_{i=1}^{m} \phi_t^i Y_t^i \tag{1.1}$$

where $i = 1, \ldots, m$ is an index used to distinguish different assets, V_t is the portfolio value at time t, Y_t^i is the price per unit of asset i at time t, and ϕ_t^i is the number of units of asset i that are present in the portfolio at time t. All prices are supposed to be expressed in a given unit of currency such as dollars or euros; portfolio value is then expressed in the same unit of currency. The numbers Y_t^i together form a vector of length m which will be written as Y_t. Likewise, we introduce an m-vector ϕ_t whose entries ϕ_t^i specify portfolio composition at time t. Both Y_t and ϕ_t are defined as column vectors. The expression (1.1) for portfolio value can then be rewritten as

$$V_t = \phi_t^\top Y_t \tag{1.2}$$

where the superscript $^\top$ denotes transposition. Vector notation will be used frequently throughout this book.

Under the idealizing assumptions above, investors have no control of the evolution of prices, but they can adjust their holdings (the numbers ϕ_t^i in the expression above) at any time. The evolution of the value of the portfolio depends both on the way that prices change in time and on the way in which the portfolio composition is modified in the course of time. The joint effect can be described in terms of formulas which will be reviewed in this section for the case in which portfolio composition is only changed at discrete points in time. Later on in this chapter, it will be argued that, for theoretical purposes, it is convenient to assume that portfolio composition can be changed *continuously*, even if in practice truly continuous trading is not possible. To describe the evolution of portfolio value that results from both continuously changing prices and continuously changing portfolio composition, some mathematical developments are needed. These are reviewed in Chapter 2.

In the continuous-time framework as used in this book, it will be assumed that prices do not experience instantaneous jumps, so that there is no ambiguity as to whether Y_t refers to a price before or after a jump has taken place at time t. With respect to portfolio composition, the situation is different. Instantaneous changes of portfolio composition will be allowed; these correspond to selling and/or buying a package of assets at a single point in time. In such cases, we need to be precise as to whether ϕ_t refers to portfolio composition before or after the trade at time t has been effectuated. By convention, the symbol ϕ_t is used to refer to portfolio composition *after* the trade, and ϕ_{t-} denotes portfolio composition before the trade; in other terms,

$$\phi_{t-} := \lim_{\tau \uparrow t} \phi_\tau$$

where the notation "$\tau \uparrow t$" indicates that the limit is taken from below.

1.2.2 Self-financing portfolios

Let us consider a fixed time interval during which a portfolio is held, possibly with changes in composition. It will be assumed that during this period no money is withdrawn from the portfolio (for instance for consumption), and neither are any funds added from outside, for instance from labor income or from other forms of income. As a consequence, all trading must take place under the *budget constraint* which states that, in every change of portfolio composition, the value of the assets sold must be equal to the value of the assets bought. Trading strategies that satisfy this condition are said to be *self-financing*. One also speaks of a "self-financing portfolio".

The restriction to self-financing strategies simplifies the presentation, but this is not the only reason to be especially interested in such strategies. Below we will often be concerned with the problem of determining the value of a *contingent claim*, i.e. a contract that will pay, at a time in the future, an amount that is determined by information that will be known at the time of payoff but that is not known now. Suppose it is possible to create a trading strategy that, starting from a given initial portfolio value V_0, causes the portfolio value at the time of payoff to be equal to the value of the contingent claim under *all* possible circumstances. The strategy is then said to *replicate* the claim. If the replicating strategy is self-financing, then the initial portfolio value V_0 can be viewed, and one might even say: must be viewed, as the "fair price" of the contract.

For convenience, the initial point of the time interval under consideration will be called $t = 0$, and the final point will be written as $t = T$. The value V_0 of the portfolio at time 0 may be considered given. We are interested in particular in getting expressions for the final portfolio value V_T as a function of decisions that are taken on the portfolio composition during the interval from 0 to T. If t is a time at which a change of portfolio composition takes place (a *rebalancing date*), then the asset holdings at that time are changed from old values to new values, so that $\phi_t^i \neq \phi_{t^-}^i$ for some or all of the asset indices $i = 1, \ldots, m$. The budget constraint, i.e. the condition that the total value of assets bought is equal to the total value of assets sold, is expressed in mathematical terms by

$$\sum_{i=1}^{m} \phi_{t^-}^i Y_t^i = \sum_{i=1}^{m} \phi_t^i Y_t^i \tag{1.3}$$

for each rebalancing date t. More specifically, let the rebalancing times be indicated by $t_1, \ldots, t_n$, with $0 < t_1 < \cdots < t_n < T$. Since by assumption there is no change in the portfolio composition between time t_j and time t_{j+1}, the equality $\phi_{t_{j+1}^-}^i = \phi_{t_j}^i$

holds and therefore the condition (1.3) may also be written as

$$\sum_{i=1}^{m} \phi_{t_{j+1}}^i Y_{t_{j+1}}^i = \sum_{i=1}^{m} \phi_{t_j}^i Y_{t_{j+1}}^i. \tag{1.4}$$

By subtracting $\sum_{i=1}^m \phi_{t_j}^i Y_{t_j}^i$ from both sides and using (1.1), we can alternatively write the condition as

$$V_{t_{j+1}} - V_{t_j} = \sum_{i=1}^{m} \phi_{t_j}^i (Y_{t_{j+1}}^i - Y_{t_j}^i) \tag{1.5}$$

which is the same as

$$V_{t_{j+1}} = V_{t_j} + \sum_{i=1}^{m} \phi_{t_j}^i (Y_{t_{j+1}}^i - Y_{t_j}^i). \tag{1.6}$$

In words, this says that the portfolio value at time t_{j+1} is equal to the value at time t_j plus the gains or losses that have been realized on the assets that constitute the portfolio. These gains or losses are computed as the changes in value of these assets, multiplied by the numbers of units of the assets that were selected in the rebalancing that took place at time t_j. This property is an alternative statement of what it means for a portfolio to be self-financing. Indeed, the rule (1.6) has been derived from the budget constraint (1.3), but vice versa it can be verified that (1.3) can be derived from (1.6) given that portfolio value is defined by (1.1), so that the two statements are in fact equivalent.

The notation can be simplified somewhat by switching to vector notation. Using the m-vector ϕ_t of asset holdings at time t and the m-vector Y_t of asset values at time t, we can write, instead of (1.5),

$$V_{t_{j+1}} - V_{t_j} = \phi_{t_j}^\top (Y_{t_{j+1}} - Y_{t_j}). \tag{1.7}$$

A further simplification can be made by introducing the *forward difference operator* Δ and writing the condition for a portfolio to be self-financing as

$$\boxed{\Delta V_{t_j} = \phi_{t_j}^\top \Delta Y_{t_j}} \tag{1.8}$$

where ΔV_{t_j} stands for $V_{t_{j+1}} - V_{t_j}$, and ΔY_{t_j} for $Y_{t_{j+1}} - Y_{t_j}$. To streamline the notation even more, let us set $t_0 = 0$ and $t_{n+1} = T$. We can then write

$$V_T - V_0 = \sum_{j=0}^{n} \Delta V_{t_j}$$

where use is made of the *telescope rule*.[1] This leads to the following expression for the portfolio value at time T:

$$V_T = V_0 + \sum_{j=0}^{n} \phi_{t_j}^\top \Delta Y_{t_j}. \tag{1.9}$$

The expression holds for self-financing portfolio strategies. In other words, if a strategy $\{\phi_t\}_t$ is defined that satisfies the budget constraint (1.3), then the portfolio value at time T can be computed on the basis of the formula above. Conversely, if the relation between portfolio value (as defined by (1.2)) at any two times τ_1 and τ_2 is given by (1.9) with 0 replaced by τ_1 and T by τ_2, taking the sum over all j such that t_j lies between τ_1 and τ_2, then the strategy $\{\phi_t\}_t$ is self-financing.

A portfolio that is rebalanced according to a well-defined self-financing strategy may itself be considered as an asset. Take for instance a simple financial product such as the zero-coupon bond which pays, for each invested euro, a given amount at a given future time.[2] A bank might construct a new product by the following strategy. Suppose that an initial capital is available. Use this capital at the initial time to buy five-year zero-coupon bonds. After one year, sell these bonds (which by then have become four-year bonds), and use the proceeds to buy five-year bonds. Do the same after two years, and so on. This strategy is self-financing, and it defines a new financial product which might be called a "perpetual five-year bond", or which might be sold under a more fancy name invented by the bank's marketing department. This product will have characteristics of its own (in particular it is sensitive to the variations of the five-year interest rate) which may make it attractive for some investors. The new product can be thought of as an asset by itself; it could be part of some portfolios which again may be subject to well-defined trading strategies, and so on. In this way, self-financing trading strategies can be thought of as devices which transform assets into new assets.

1.2.3 Use of a numéraire

Instead of using a unit of currency, such as euros or dollars, as a unit of account, we can also express prices in terms of a particular asset that has been chosen for this purpose. For instance, to make prices of assets at different times more comparable, one can express prices in terms of a number of units of a prescribed basket of commodities. When an asset is employed as a unit of account, we say that it is used

[1] The telescope rule states that the sum of the successive differences of a sequence of real numbers is equal to the last element of the sequence minus the first one. Formulawise, the rule can be written as $\sum_{i=1}^{n-1}(a_{i+1} - a_i) = a_n - a_1$.

[2] This product is sold to consumers under the name "deposit", and the amount to be received at the given future time is typically expressed in terms of an interest rate.

as a *numéraire*.[3] Since the number of units of one asset that can be traded against a given number of units of another asset is determined by the *relative* prices of the assets, no essential economic information is lost when prices are expressed relative to a numéraire rather than in terms of money. From a theoretical perspective, it may actually be preferable to avoid the indeterminacy that comes from choosing a particular currency.

Any asset can be used as a numéraire, as long as one can be sure that the value of the asset is never zero, since relative prices cannot be defined with respect to an asset that has zero value. Financial models typically contain many assets that always represent some value, in other words, whose price is always positive. Therefore one usually has a wide choice of possible numéraires; this may be used to advantage in the context of a particular pricing problem, much in the same way as one might choose a convenient coordinate system in a geometry problem. Numéraires will be used frequently in this book.

Some of the advantages of using a numéraire can already be seen when we discuss the evolution of portfolio value under the combined influence of changing asset prices and a self-financing trading strategy. Suppose that there are m assets to be traded which are numbered from 1 to m, and that asset m can be taken as a numéraire. To highlight the special role of this asset, we shall write its value at time t, rather than Y_t^m. Let the initial value of a portfolio be given. To specify a self-financing strategy, it is enough to specify the holdings of the first $m-1$ assets at the initial time and at the rebalancing times, because the number of units to be held of the numéraire asset is determined by the budget constraint.[4] Relative to the value of the numéraire at time t_j, the portfolio value at time t_j is given by

$$\frac{V_{t_j}}{N_{t_j}} = \sum_{i=1}^{m-1} \phi_{t_j}^i \frac{Y_{t_j}^i}{N_{t_j}} + \phi_{t_j}^m = \sum_{i=1}^{m-1} \phi_{t_{j-1}}^i \frac{Y_{t_j}^i}{N_{t_j}} + \phi_{t_{j-1}}^m \tag{1.10}$$

where the latter equality follows from the budget constraint (1.4). A similar expression can of course be written down at time t_{j-1}. Subtraction then leads to the following formula for the change of relative portfolio value between two successive rebalancing dates:

$$\frac{V_{t_j}}{N_{t_j}} - \frac{V_{t_{j-1}}}{N_{t_{j-1}}} = \sum_{i=1}^{m-1} \phi_{t_{j-1}}^i \left(\frac{Y_{t_j}^i}{N_{t_j}} - \frac{Y_{t_{j-1}}^i}{N_{t_{j-1}}} \right). \tag{1.11}$$

[3]The word *numéraire* is used in French to refer to coins and banknotes. The idea of using a traded asset as a unit of account, rather than some arbitrary currency, can be traced back to the works of the French engineer Achylle-Nicholas Isnard (1749–1803). Writing about economics in his spare time, Isnard was one of the earliest contributors to mathematical economics. The idea of expressing prices in terms of a numéraire is also used extensively in the work of the French-Swiss economist Léon Walras (1834–1910), who is known as the father of general equilibrium theory.

[4]Note that it is essential here that the value of the numéraire is never zero.

Using the forward difference operator again as well as the telescope rule, we can write

$$\frac{V_T}{N_T} = \frac{V_0}{N_0} + \sum_{j=0}^{n} \sum_{i=1}^{m-1} \phi^i_{t_{j-1}} \Delta \frac{Y^i_{t_j}}{N_{t_j}}. \tag{1.12}$$

All asset values at any time are now expressed relative to the value of the numéraire at the same time.

From the point of view of designing a trading strategy, it is of interest to note that both in (1.9) and in (1.12) the quantities $\phi^1_{t_j}, \ldots, \phi^{m-1}_{t_j}$ can be chosen freely. Comparing the two expressions (1.9) and (1.12) to each other, one notes that to compute the final portfolio value V_T by means of (1.9), the corresponding values of $\phi^m_{t_j}$ (holdings of the numéraire asset) must be computed at each rebalancing time t_j, which in turns requires calculating the portfolio value at each of these times. In contrast, the formula (1.12) gives the final portfolio value directly in terms of the free variables $\phi^i_{t_j}$ ($i = 1, \ldots, m-1$, $j = 0, \ldots, n$); however, the value is given in terms of the numéraire rather than directly in monetary terms. While the final result of a financial calculation is usually required in terms of a unit of money, it is often convenient to use a suitably chosen numéraire in intermediate steps. Examples of this will be seen at various occasions in later chapters.

As usual it is convenient to use vector notation. In vector form, the expression (1.12) becomes[5]

$$\boxed{\frac{V_T}{N_T} = \frac{V_0}{N_0} + \sum_{j=0}^{n} \phi^\top_{t_j} \Delta \frac{Y_{t_j}}{N_{t_j}}\,.} \tag{1.13}$$

This formula gives an expression for the final portfolio value V_T that results from the asset price process $Y_{t_0}, Y_{t_1}, \ldots$ and from the self-financing strategy whose first $m-1$ components are given by $\phi^1_{t_j}, \ldots, \phi^{m-1}_{t_j}$ ($j = 0, \ldots, n$). The last component $\phi^m_{t_j}$ is determined by the budget constraint which states that the value of the portfolio before and after rebalancing at time t_j must be the same.

1.3 Transition to continuous time

Now, let us consider what happens if the number n of trading times is large. In modern markets, positions in liquid assets can be revised and changed again in fractions of seconds, so that the number of rebalancings can indeed be very large. From a mathematical perspective it is then very attractive to allow ourselves to call

[5] In principle there is an ambiguity in (1.13) since the inner product that appears in the formula could be read as an inner product of the vectors $(\phi^1_{t_j}, \ldots, \phi^{m-1}_{t_j})$ and $(\Delta(Y^1_{t_j}/N_{t_j}), \ldots, \Delta(Y^{m-1}_{t_j}/N_{t_j}))$ or as an inner product of the vectors $(\phi^1_{t_j}, \ldots, \phi^m_{t_j})$ and $(\Delta(Y^1_{t_j}/N_{t_j}), \ldots, \Delta(Y^m_{t_j}/N_{t_j}))$. However the two inner products are the same, because $N_{t_j} = Y^m_{t_j}$ for all j so that $\Delta(Y^m_{t_j}/N_{t_j}) = 1 - 1 = 0$.

upon the power of differential and integral calculus and to think of asset holdings ϕ_t^j as general functions of continuous time, rather than to maintain the restriction that these functions must be piecewise constant. To make this approach successful, we should then be able to replace the expressions (1.9) and (1.13) by corresponding integral expressions

$$V_T = V_0 + \int_0^T \phi_t^\top \, dY_t \tag{1.14}$$

and, in terms of a numéraire,

$$\frac{V_T}{N_T} = \frac{V_0}{N_0} + \int_0^T \phi_t^\top \, d\frac{Y_t}{N_t} \,. \tag{1.15}$$

These are still tentative formulations, since there are issues to be addressed even in the definition of the integrals that appear in (1.14) and (1.15).

1.3.1 Riemann-Stieltjes integrals

Integrals of the form $\int_a^b f(x)\, dg(x)$, in which both the integrand $f(x)$ and the integrator $g(x)$ can be taken from some large class of functions, were already investigated in the 19th century. A typical approach is to look at sums of the form

$$S(f, g, \Pi, \xi) := \sum_{j=0}^{n} f(\xi_j)\big(g(x_{j+1}) - g(x_j)\big)$$

where $\Pi = (x_0, x_1, \ldots, x_{n+1})$ is a *partition* of $[a, b]$,[6] and where $\xi = (\xi_0, \xi_1, \ldots, \xi_n)$ is a corresponding sequence of intermediate points, i.e. $x_j \le \xi_j \le x_{j+1}$ for all $j = 0, \ldots, n$. The *mesh* of a partition $\Pi = (x_0, x_1, \ldots, x_n)$ is defined by

$$|\Pi| = \max_{j=0,\ldots,n} (x_{j+1} - x_j).$$

In order to achieve the transition to continuous time, one may think of applying the following theorem from Riemann-Stieltjes[7] integration theory. The theorem[8] refers to a particular property that is defined as follows: a function $g(x)$ defined on an interval $[a, b]$ is said to be *of bounded variation* if there exists a number M such that $\sum_{j=0}^{n} |g(x_{j+1}) - g(x_j)| \le M$ for all partitions $a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b$.

[6] A sequence of points $(x_0, x_1, \ldots, x_{n+1})$ is called a *partition* of the interval $[a, b]$ if $a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b$.

[7] Bernhard Riemann (1826-1866), German mathematician. Thomas Jan Stieltjes (1856-1894), Dutch mathematician.

[8] Integration theory can be built up in several ways, and therefore the theorem as stated here should be viewed as just a representative of various results in the same spirit, namely: the integral can be defined, i.e. the same limit is obtained irrespective of the sequence of refining partitions that is chosen, if the integrator and the integrand are sufficiently well-behaved.

The infimum of all numbers M that have this property is called the *total variation* of the function $g(x)$ on the interval $[a, b]$. Intuitively, a function of bounded variation has "finite length". It can be proved that a function is of bounded variation if and only if it can be written as the difference of two nondecreasing functions. A function of bounded variation need not be continuous; for instance, take $g(x)$ defined on $[0, 1]$ by $g(x) = 0$ for $0 \leq x < \frac{1}{2}$ and $g(x) = 1$ for $\frac{1}{2} \leq x \leq 1$. Conversely, there exist continuous functions that are not of bounded variation. For instance, consider the function $g(x)$ defined on $[0, 1]$ by $g(x) = x \sin(1/x)$ for $0 < x \leq 1$, and $g(0) = 0$.

Theorem 1.3.1 *Suppose that $f(x)$ is a continuous function defined on the interval $[a, b]$ and that $g(x)$ is a function of bounded variation defined on the same interval. In that case there exists a number, written as $\int_a^b f(x)\, dg(x)$, which has the property that for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\Big| \int_a^b f(x)\, dg(x) - \sum_{j=0}^{n} f(\xi_j) \Delta g(x_j) \Big| < \varepsilon$$

for all sequences of points $a = x_0 < x_1 \cdots x_n < x_{n+1} = b$ that satisfy $x_{j+1} - x_j < \delta$ for all $j = 0, \ldots, n$, and for all sequences of points $\xi_0, \ldots, \xi_n$ that satisfy $x_j \leq \xi_j \leq x_{j+1}$ for all $j = 0, \ldots, n$.

The number $\int_a^b f(x)\, dg(x)$ is called the *Riemann-Stieltjes integral* of f with respect to g. The theorem states that this number is defined by the functions f and g and by the integration interval $[a, b]$; in particular any choice of intermediate points will give rise to approximately the same value of the sum $\sum_{j=0}^{n} f(\xi_j) \Delta g(x_j)$, and as the intermediate points become more dense the approximation becomes more close. In this way there is no ambiguity about the value of the integral. One can show by examples that these properties need no longer hold if f is not continuous or g is not of bounded variation.

It may seem reasonable to assume that the trajectories of asset prices are of bounded variation. Certainly it is true that the total variation of actual stock prices in a given interval of time (i.e. the sum of the absolute values of the price changes that take place during that interval) is always finite, for the simple reason that the number of instants at which the price changes may be large, but must certainly be finite. This does not necessarily mean, however, that the assumption of bounded variation works well in an idealized model in which trading takes place continuously. In fact, some doubt on the applicability of Riemann-Stieltjes integration is raised by the experiment described below.

1.3.2 A trading experiment

One of the calculus rules of Riemann-Stieltjes integration states that, if g is continuous as well as of bounded variation and F is a continuously differentiable function, then

$$\int_a^b F'(g(x))\,dg(x) = F(g(b)) - F(g(a)). \tag{1.16}$$

This is a generalized form of the fundamental theorem of calculus (the standard form is obtained in the case that g is the identity function, i.e. $g(x) = x$). In particular, by taking $F(x) = \frac{1}{2}x^2$, we find

$$\int_a^b g(x)\,dg(x) = \tfrac{1}{2}g(b)^2 - \tfrac{1}{2}g(a)^2. \tag{1.17}$$

This rule might be used for the construction of trading strategies in a financial market. To simplify, suppose that there are only two assets to invest in, so that $m = 2$ in the derivations above. Write S_t ("stock") instead of Y_t^1 and B_t ("bond") instead of Y_t^2, and take the bond as a numéraire. The formula (1.13) then becomes

$$\frac{V_T}{B_T} = \frac{V_0}{B_0} + \sum_{j=0}^{n} \phi_{t_j}\,\Delta\,\frac{S_{t_j}}{B_{t_j}}\,. \tag{1.18}$$

Suppose now that we choose, at each time t_j $(j = 1, \ldots, n)$,

$$\phi_{t_j} = \frac{S_{t_j}}{B_{t_j}} - \frac{S_0}{B_0}. \tag{1.19}$$

This can indeed be done in practice; no "crystal ball" is required, since S_{t_j} and B_{t_j} are known quantities at time t_j. If the time intervals between rebalancings are sufficiently small, then, by the theorem above, the sum at the right hand side of (1.18) is close to the integral

$$\int_0^T \left(\frac{S_t}{B_t} - \frac{S_0}{B_0}\right) d\,\frac{S_t}{B_t}\,.$$

In this integral we can also write $d(S_t/B_t - S_0/B_0)$ instead of $d(S_t/B_t)$, and therefore by virtue of (1.17) the value of the integral is equal to

$$\frac{1}{2}\left(\frac{S_T}{B_T} - \frac{S_0}{B_0}\right)^2.$$

One remarkable observation here is that the increment of the relative portfolio value (i.e. relative to the numéraire) across the interval $[0, T]$ depends only on the increment of the relative value of the asset S. Moreover the dependence is quadratic. In particular the value of the integral is always nonnegative, and it is positive whenever

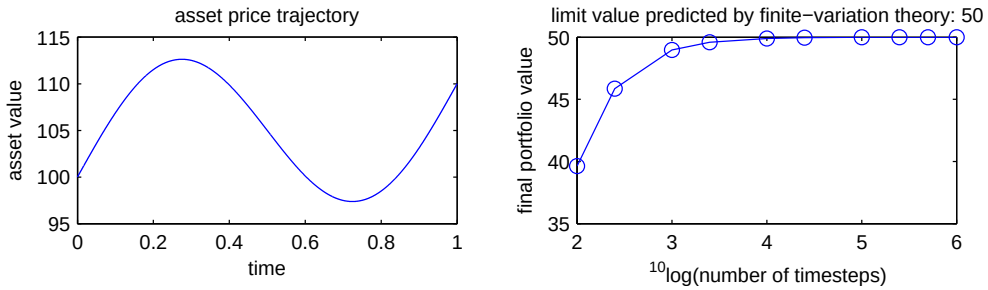

Figure 1.1: Test of money making scheme: $S_t = 100 + 10\sin(2\pi t) + 10t$; $B_t = 1$.

S_T/B_T is not equal to S_0/B_0. In particular we can use the strategy with *zero* initial capital ($V_0 = 0$), and obtain from (1.18)

$$\frac{V_T}{B_T} \approx \frac{1}{2}\left(\frac{S_T}{B_T} - \frac{S_0}{B_0}\right)^2$$

where the approximation should be better and better as we increase the frequency of portfolio rebalancings. If we assume that the assets S_t and B_t are really different assets in the sense that their values do not move in tandem, then it seems that the strategy (1.19) in general leads to a positive final portfolio value, while negative final portfolio values do not occur; moreover, *no* initial investment is required to achieve this.

Let us test this promising scheme. Figures 1.1 and 1.2 show cases in which the asset price is a smooth function. The results of the strategy lives up perfectly to the expectations; in the second case, where the asset price is quite oscillatory, convergence is only achieved when the partitioning is made rather fine, but it is achieved. These asset price trajectories are not terribly realistic, however. To get an asset price trajectory that is more like what we are used to seeing when looking at plots of stock prices, asset prices (on a fine grid) may be generated by a scheme of the following type:

$$S_{t_{j+1}} = S_{t_j} + \mu S_{t_j}\Delta t + \sigma S_{t_j}\sqrt{\Delta t}\, Z_j \tag{1.20}$$

where the Z_j's are independent standard normal variables, μ and σ are constants, and Δt is a very small time step (not larger than the length of the smallest interval between rebalancing times). Examples of the results are shown in Figures 1.3 and 1.4.

A rather different behavior is seen here. The final values of the portfolio strategy applied with increasingly higher frequencies to a given asset price trajectory do seem to converge, but not to the value predicted by the theory. As is seen from the graphs, negative results may well occur. Our scheme doesn't seem to work. Perhaps the prospect was too good to be true, but what is the mathematical explanation?

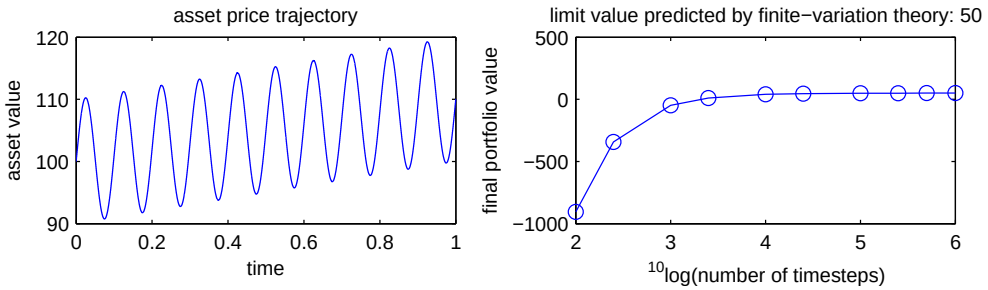

Figure 1.2: Test of money making scheme: $S_t = 100 + 10\sin(20\pi t) + 10t$; $B_t = 1$.

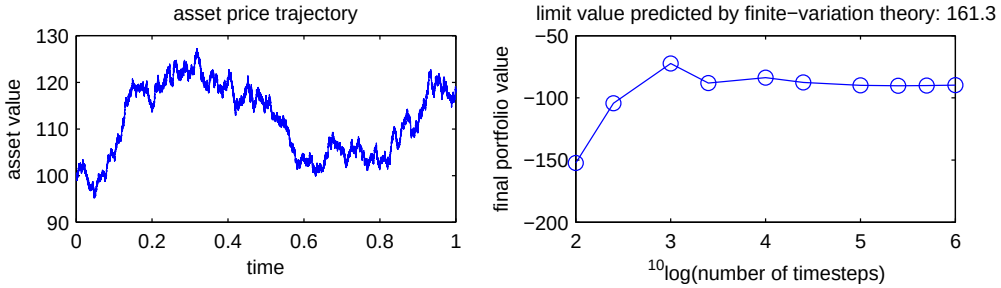

Figure 1.3: Test of money making scheme: S_t randomly generated as in (1.20) with $\mu = 0.08$ and $\sigma = 0.2$; $B_t = 1$.

After all, Theorem 1.3.1 above is a valid statement. The problem must be that the assumptions of the theorem are not satisfied — the trajectories of asset prices are *not* adequately described in continuous time as functions of bounded variation.

1.3.3 A new calculus

One response to the failed money making experiment might be to give up on the idea of replacing sums by integrals altogether. However, since in practice we can trade almost continuously and because calculus is such a convenient tool, it is preferable to develop a *generalized* calculus that can deal with trajectories that are not of bounded variation. Riemann-Stieltjes integration was developed in the 19th century; in the 20th century, mathematical tools have been constructed which enable us to deal

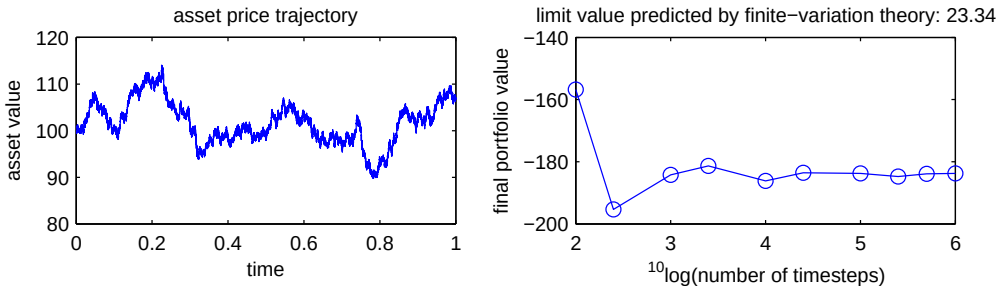

Figure 1.4: Test of money making scheme: S_t randomly generated as in (1.20) with $\mu = 0.08$ and $\sigma = 0.2$; $B_t = 1$.

with the irregularity of asset price trajectories. In the new calculus (known as *Itô calculus*)[9] we can still use rules of integration, and for instance devise strategies that make the portfolio value at time T depend in a particular way on the value of a particular asset at the same time. The calculus produces additional terms which do not appear in (1.17), and which preclude the development of money-making schemes such as the one discussed above. Stated in other words, these additional terms explain why such schemes do not work under the assumptions of the Itô calculus.

Nowadays, it is generally accepted that the additional terms produced by Itô's calculus have to be taken into account in the analysis of trading strategies in financial markets. Moreover, models based on Itô calculus are taken as guidelines to develop trading strategies that may not act as money machines but that still satisfy useful purposes, such as providing protection against liabilities that may arise ("hedging"), or, in investment management, optimizing the balance between risk and return according to a given criterion. The following chapters describe the new calculus and a number of applications in financial markets.

1.4 Exercises

The exercises in this chapter are somewhat atypical, in the sense that they require more extensive knowledge of real analysis than will be needed in exercises in other chapters.

1. Define a function g on $[0,1]$ by $g(0) = 0$ and $g(x) = x\sin(1/x)$ for $0 < x \leq 1$. Prove that (as claimed on p. 13) this function is continuous, but not of bounded variation on $[0,1]$.

2. **a**. Show that any continuous function on a closed and bounded interval is in fact uniformly continuous.[10]

b. Using part a., show that

$$\lim_{|\Pi|\to 0} \sum_{j=0}^{n} \big(g(x_{j+1}) - g(x_j)\big)^2 = 0$$

for any continuous function of bounded variation g defined on a closed and bounded interval $[a,b]$, where Π is the partition with partition points $a = x_0 < x_1 < \cdots <$

[9] Kiyoshi Itô (1915–2008), Japanese mathematician. Itô developed his calculus in the mid-1940s while working for the national statistical office of Japan.

[10] A real-valued function defined on a subset A of the real line is said to be uniformly continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all x and y in A such that $|x - y| < \delta$. The difference with ordinary continuity is that, for uniform continuity, it is required that the same δ can be used throughout the domain of definition.

$x_{n+1} = b$. In other (and more precise) words, show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sum_{j=0}^{n} \big(g(x_{j+1}) - g(x_j)\big)^2 < \varepsilon$$

for every partition $\Pi = (x_0, x_1, \ldots, x_{n+1})$ of $[a, b]$ that satisfies $|\Pi| < \delta$.

Chapter 2

Stochastic calculus

2.1 Brownian motion

2.1.1 Definition

Just as the normal distribution is in several senses the "nicest" of all continuous distributions that random variables can have, *Brownian motion*[1] (also known as the *Wiener process*)[2] is the continuous stochastic process that is most attractive in many ways. Most of the financial models that are used in practice are based on this process. The Wiener process[3] may be seen as the continuous version of the discrete-time *standard random walk*, which is the time series generated by the model

$$X_0 = 0, \quad X_{k+1} = X_k + Z_k, \quad Z_k \overset{\text{i.i.d.}}{\sim} N(0,1). \tag{2.1}$$

The definition of the Wiener process can be stated as follows.

Definition 2.1.1 A continuous-time process $\{W_t\}$ $(t \geq 0)$ is said to be a *Wiener process* or a *Brownian motion* if it satisfies the following properties.

(i) $W_0 = 0$.

(ii) If $t_1 < t_2 \leq t_3 < t_4$, then the increments $W_{t_2} - W_{t_1}$ and $W_{t_4} - W_{t_3}$ are independent.

(iii) For any given t_1 and t_2 with $t_2 > t_1$, the distribution of the increment $W_{t_2} - W_{t_1}$ is the normal distribution with mean 0 and variance $t_2 - t_1$.

The Wiener process has proven to be extremely useful in the modeling of financial markets. It is typically not used in pure form but rather processed by a stochastic differential equation, in a way that will be discussed below.

[1] Robert Brown (1773–1858), British biologist.

[2] Norbert Wiener (1894–1964), American mathematician.

[3] The terms "Wiener process" and "Brownian motion" are used interchangeably in this book.

Property (i) in the definition above is just a normalization. Property (ii) is called the *independent increments* property. Properties (ii) and (iii) together imply that the conditional distribution of W_{t_2} given W_t for $0 \leq t \leq t_1$, where $t_1 < t_2$, is the normal distribution with expectation W_{t_1} and variance $t_2 - t_1$. In particular, the conditional distribution of W_{t_2} given information up to time $t_1 < t_2$ depends only on W_{t_1} and not on any earlier values of W_t.

The definition as given above is a bit unusual in that it just lists a set of properties. In fact it is not at all trivial to show that it is indeed possible to define a collection of stochastic variables $\{W_t\}_{t\in[0,\infty)}$ in such a way that all conditions of the definition above are satisfied. Such a construction was carried out by Wiener, which is why the process bears his name. One of the key facts that make the construction possible is the following: if X_1 and X_2 are independent normal random variables with expectation 0 and with variance σ_1^2 and σ_2^2 respectively, then $X_1 + X_2$ is a normal random variable with expectation 0 and with variance $\sigma_1^2 + \sigma_2^2$. If this would not hold, then properties (ii) and (iii) in the definition of the Wiener process would not be compatible.

Some remarks on terminology need to be made. The process defined above is called by some authors a *standard* Wiener process. The term "Wiener process" without further qualification is then used for any process that satisfies conditions (i), (ii), and

(iii)′ There exists a constant $\sigma > 0$ such that, for any given t_1 and t_2 with $t_2 > t_1$, the distribution of the increment $W_{t_2} - W_{t_1}$ is the normal distribution with mean 0 and variance $\sigma^2(t_2 - t_1)$.

More specifically, such a process is called a *Wiener process with variance parameter* σ^2. If W_t is a Wiener process with variance parameter σ^2, then $\sigma^{-1}W_t$ is a standard Wiener process. In this book, the standard Wiener process is used so often that it is more convenient to refer to it simply as a "Wiener process" or "Brownian motion" without the specification "standard". So if mention is made below of a "Wiener process" or a "Brownian motion" without further qualification, then the standard Wiener process is meant.

2.1.2 Vector Brownian motions

It is often useful in financial market modeling to consider several Brownian motions at the same time. A *vector Brownian motion with variance-covariance matrix* Σ is a vector-valued stochastic process that satisfies the same properties as the Brownian motion defined above, except that the increments $W_{t_2} - W_{t_1}$ follow multivariate normal distributions with mean 0 and variance-covariance matrix $(t_2 - t_1)\Sigma$. The variance-covariance matrix describes correlation between increments of the components of a vector Brownian motion across the same interval of time; increments

corresponding to non-overlapping time intervals are independent, as in the case of the scalar Brownian motion. A *standard vector Brownian motion* is a vector Brownian motion whose variance-covariance matrix is the identity matrix. In other words, a k-vector standard Brownian motion is constructed from k independent scalar Brownian motions, taken together into a vector. Whenever several Brownian motions are discussed below, it will always be assumed that they together form a vector Brownian motion.[4]

A well known property of the normal distribution is that any linear combination of jointly normally distributed variables is again normally distributed. Likewise, one can show that any linear combination of (not necessarily standard) Brownian motions, which together form a vector Brownian motion, is again a (not necessarily standard) Brownian motion. For instance, if $W_{1,t}$ and $W_{2,t}$ are independent Brownian motions with variance parameters σ_1^2 and σ_2^2 respectively, then $aW_{1,t} + bW_{2,t}$ is a Brownian motion with variance parameter $\sigma^2 = a^2\sigma_1^2 + b^2\sigma_2^2$. In terms of standard Brownian motions, the addition rule can be stated as follows:

$$\sigma_1 W_{1,t} + \sigma_2 W_{2,t} = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}\, W_t \tag{2.2}$$

where $W_{1,t}$, $W_{2,t}$, and W_t are all standard Wiener processes, and ρ is the correlation coefficient of $W_{1,t}$ and $W_{2,t}$. More generally, if Z_t is an n-vector Brownian motion with variance-covariance matrix Σ and M is a matrix of size $k \times n$, then MZ_t is a vector Brownian motion with variance-covariance matrix $M\Sigma M^\top$.

These connections make it possible to express any (nonstandard) vector Brownian motion as a linear transformation of a standard vector Brownian motion. If for instance we have two Brownian motions $W_{1,t}$ and $W_{2,t}$ that are correlated with correlation coefficient ρ, then we can think of these two processes as being obtained from two *independent* Brownian motions $\hat{W}_{1,t}$ and $\hat{W}_{2,t}$ by the rules

$$\begin{aligned} W_{1,t} &= \hat{W}_{1,t} \\ W_{2,t} &= \rho\hat{W}_{1,t} + \sqrt{1-\rho^2}\,\hat{W}_{2,t}. \end{aligned}$$

In general, if W_t is a vector Brownian motion with variance-covariance matrix Σ, then we can think of W_t as being generated by

$$W_t = M\hat{W}_t$$

[4] A vector formed of normally distributed variables does not necessarily have a multivariate normal distribution; see Exc. 7. Likewise, when several Brownian motions are taken into a vector, the result is not necessarily a vector Brownian motion; however the examples that prove this point are somewhat artificial and not likely to be met in practice. Also, when a vector is formed of several *independent* Brownian motions, then the result is always a (standard) vector Brownian motion.

where $\hat{W}_t$ is a *standard* vector Brownian motion, and M is any matrix such that $MM^\top = \Sigma$. The decomposition of a positive definite matrix Σ in the form $\Sigma = MM^\top$ where M is lower triangular and has positive entries on the diagonal is known as the *Cholesky decomposition*.[5] As in the scalar case, when the term "vector Brownian motion" is used in this book, then a *standard* vector Brownian motion is meant.

2.2 Stochastic integrals

As discussed in Section 1.2, it is of interest for the analysis of trading strategies to be able to define integrals of the form $\int_0^T \phi_t \, dY_t$, even when Y_t is not of bounded variation. Such an integral should in some appropriate sense be a limit of expressions of the form

$$\sum_{j=0}^{n} \phi_{t_j}(Y_{t_{j+1}} - Y_{t_j})$$

where $0 = t_0 < t_1 < \cdots < t_{n+1} = T$ is a partitioning of the interval $[0, T]$; the limit should be approached more and more closely as the partitioning becomes finer and finer. However, the concept of Riemann-Stieltjes integration is not good enough when Y_t is not of bounded variation, because in this case one sequence of refining partitions may lead to a different limit than another sequence does, and the Riemann-Stieltjes integration theory doesn't provide a clue as to which limit is the "right" one. A more subtle notion of integral is required.

2.2.1 The idea of the stochastic integral

The purpose of this section is to discuss how to define an integral of the form $\int_0^T X_t \, dZ_t$ when X_t and Z_t are *stochastic processes* that satisfy certain conditions. The integral itself is in general also a stochastic variable. At first sight it may seem that integration theory would only become more complicated when it is applied to stochastic processes rather than to functions as in the Riemann-Stieltjes theory, but the stochastic context does have its advantages; in particular, it makes it possible to discard certain cases that occur with vanishing probability. Moreover, in applications to financial markets it is natural to think of prices as evolving in a stochastic way. As will be discussed below, the stochastic integral can be used not only to define results of trading strategies but also to develop models for the evolution of prices.

[5] André-Louis Cholesky (1875–1918), French military officer and mathematician. Cholesky developed the matrix decomposition named after him to facilitate the solution of the least-squares problems that he encountered in geodetic work. He died in battle a few months before the end of World War I.

So, let X_t and Z_t be stochastic processes defined on an interval $[0, T]$; suitable requirements for these processes will be specified in a moment. For any partitioning $0 = t_0 < t_1 < \cdots < t_{n+1} = T$ of the interval $[0, T]$, we can form the sum

$$\sum_{j=0}^{n} X_{t_j}(Z_{t_{j+1}} - Z_{t_j})$$

which defines a random number. One can ask oneself whether these random numbers converge to some random variable, which then may be called the stochastic integral on the interval $[0, T]$ of the process X_t (the *integrand*) with respect to the process Z_t (the *integrator*). Because we are discussing random numbers now, the notion of "convergence" has more flexibility than it has in the deterministic case. For instance we can make use of the notion of *convergence in probability*. Recall that a sequence of random variables $X_1, X_2, \ldots$ is said to converge in probability to a random variable X if for all $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

This means that, for any chosen positive number ε, cases in which the difference between X_n and X is larger than ε may occur, but such cases are increasingly rare as n becomes larger. Suppose now that the processes X_t and Z_t satisfy the following properties.

(i) The process X_t is *adapted* to the process Z_t. This means intuitively that, for any t, X_t can be written as a function of the values of Z_s for $s \leq t$.[6]

(ii) The process Z_t is a *martingale*. This means that $E|Z_t| < \infty$ for all t, and that the *martingale condition*

$$\boxed{E_s Z_t = Z_s} \tag{2.3}$$

holds for all s and t with $s < t$, where the notation "E_s" means "conditional expectation with respect to the information available at time s".[7]

It can be shown that, under suitable continuity and boundedness conditions, these

[6] A more precise definition would require material that is not included in this book. One has to take care in particular when the process Z_t is allowed to have jumps. However, within this book only integrators are used that have continuous paths.

[7] More precisely, the information available at time s from the process Z. More general definitions can be given in which information up to time s may also be taken from other processes.

properties guarantee that one can indeed define a stochastic integral by the formula

$$\boxed{\int_0^T X_t \, dZ_t = \lim_{\Delta t \downarrow 0} \sum_0^T X_{t_j}(Z_{t_{j+1}} - Z_{t_j})} \tag{2.4}$$

where the notation expresses that a limit is taken with respect to an arbitrary sequence of refining partitions on the interval $[0, T]$. The limit is understood in the stochastic sense of convergence in probability. The first version of this key fact was discovered by Itô, in the 1940s, and since then many extensions and refinements have been made.

Since one can integrate against martingales but also against processes that have paths of bounded variation (the latter on the basis of the classical Riemann-Stieltjes integral), one can also integrate against processes that are sums of martingales and bounded-variation processes, using the simple rule $\int X \, d(Y+Z) = \int X \, dY + \int X \, dZ$. Moreover, in the modern theory of stochastic integration it has turned out that martingales may be replaced by a closely related but somewhat more general type of processes called *local martingales*.[8] In this way, one ends up with "good integrators" which are sums of local martingales and bounded-variation processes. The processes obtained in this way are called *semimartingales*. It is definitely not true that every stochastic process is a semimartingale,[9] but it is generally accepted that continuous-time models for asset pricing should be based on semimartingales.[10] All processes that we consider in this book are indeed semimartingales.

2.2.2 Basic rules for stochastic integration

If a trader is passive and just keeps a constant holding of an asset, say one unit, then the result over a period from 0 and T is just the difference in the asset price per unit at time T and the price at time 0. In mathematical terms, this property is expressed by the continuous version of the *telescope rule*:

$$\boxed{\int_0^T dZ_t \;=\; Z_T - Z_0.} \tag{2.5}$$

[8] Note the contrast with normal usage: a black cat is a particular type of cat, but a local martingale is not a particular type of martingale. Rather, it is the other way around: martingales form a subclass of local martingales. Researchers sometimes use the term "true martingale" to emphasize that a process is a martingale, and not just a local martingale.

[9] For instance, it can be shown that the process $\{X_t\}$ defined by $X_t = \sqrt[3]{W_t}$, where W_t is a Brownian motion, is not a semimartingale.

[10] In fact, it was shown by Freddy Delbaen and Walter Schachermayer in 1994, in a paper in *Mathematische Annalen*, that commonly held requirements for pricing models, as will be discussed in Chapter 3, are violated by models allowing asset price processes that are not semimartingales.

One can easily verify that this property is indeed satisfied by the stochastic integral as defined in (2.4). It is a simple but essential and frequently used rule. Another basic property of the stochastic integral is *linearity*:

$$\int_0^T (aX_t + bY_t)\, dZ_t = a \int_0^T X_t\, dZ_t + b \int_0^T Y_t\, dZ_t \tag{2.6}$$

where a and b are constants. Linearity holds not only with respect to the integrand but also with respect to the integrator:

$$\int_0^T X_t\, d(aY_t + bZ_t) = a \int_0^T X_t\, dY_t + b \int_0^T X_t\, dZ_t. \tag{2.7}$$

2.2.3 Processes defined by stochastic integrals

In the above we have considered the stochastic integral on a given interval from 0 to T, but of course the end point T can be varied. In the trading interpretation, this means that the result is monitored continuously rather than just over a fixed period. In this way one defines, starting from a given process $\{Z_t\}$ and a process $\{X_t\}$ adapted to $\{Z_t\}$, a new process $\{Y_t\}$ by

$$Y_t = Y_0 + \int_0^t X_s\, dZ_s. \tag{2.8}$$

This relation is the basis of the definition of stochastic differential equations that will be given below. The process Y_t is said to be an *integral transform* of the process Z_t by means of the process X_t; in economic terms, Y_t can be viewed as the wealth process that is generated from a given asset price process Z_t by the application of a trading strategy X_t. Due to the telescope rule (2.5), we can also write the integral relation (2.8) in the differential form

$$dY_t = X_t\, dZ_t \tag{2.9}$$

where it is understood that this means that the integrals of both sides across any interval are equal.

Under suitable boundedness assumptions[11] relating to "admissible" trading strategies as discussed in the Section 3.2.1 below, the following key statement can be made.

Theorem 2.2.1 *An integral transform of a martingale is again a martingale.*

[11] For instance, a sufficient condition for the statement of Thm. 2.2.1 to hold, in the case in which the integrator Z_t is Brownian motion, is that the integral $\int_0^T E[X_t]^2\, dt$ is finite for all $T \geq 0$.

The boundedness assumptions referred to above are needed to justify the following computation, in which t and s are arbitrary points in time with $s \geq t$:

$$\begin{aligned} E_t Y_s - Y_t &= E_t(Y_t + \textstyle\int_t^s X_u \, dZ_u) - Y_t = E_t\big(\int_t^s X_u \, dZ_u\big) \\ &= E_t\big(\lim_{\Delta t \downarrow 0} \textstyle\sum X_{t_i}(Z_{t_{i+1}} - Z_{t_i})\big) = \lim_{\Delta t \downarrow 0} \textstyle\sum E_t\big(X_{t_i}(Z_{t_{i+1}} - Z_{t_i})\big) \\ &= \lim_{\Delta t \downarrow 0} \textstyle\sum E_t E_{t_i}\big(X_{t_i}(Z_{t_{i+1}} - Z_{t_i})\big) = \lim_{\Delta t \downarrow 0} \textstyle\sum E_t X_{t_i} E_{t_i}\big(Z_{t_{i+1}} - Z_{t_i}\big) \\ &= \lim_{\Delta t \downarrow 0} \textstyle\sum E_t X_{t_i} \cdot 0 = 0. \end{aligned}$$

The computation makes use of the *tower law* of conditional expectations: if $s \geq t$, then $E_t E_s X = E_t X$.

In terms of trading, the theorem above means that if the price of an asset follows a martingale, then the expected result of any trading in this asset is zero. In particular, it is not possible to come up with a trading strategy that always produces a nonnegative result and that leads to a positive result with zero probability, because such a strategy would have a positive expected value. In other words, arbitrage is not possible with respect to a martingale. The actual application to financial markets takes into account that a probability measure may be used that is different from (but equivalent to) the "real-world" measure, and that prices must be taken relative to a numéraire in order to get the martingale property; see the discussion in Subsection 3.2.1. The martingale concept is a key notion from the mathematical point of view; it turns out to be central in financial applications as well.

The following implication of Thm. 2.2.1 is important enough to be stated separately. Appropriate boundedness conditions on integrand and integrator are again tacitly assumed.

Theorem 2.2.2 *The expected value of a stochastic integral with respect to a martingale is zero.*

In particular, we have

$$\boxed{E \int_0^T X_t \, dW_t = 0} \tag{2.10}$$

where W_t denotes a Brwonian motion and X_t can be any process that is adapted with respect to W_t and that satisfies some mild boundedness conditions. The statement of Thm. 2.2.2 can be phrased briefly as "you can't beat the system" or, as Bachelier wrote in the year 1900, "L'espérance mathématique du spéculateur est nulle".[12]

[12] "The mathematical expectation of the speculator is nil." Louis Bachelier (1870–1946) was a

2.3 Stochastic differential equations

2.3.1 Definition

Before Itô's work, stochastic processes were typically specified through their associated conditional densities (the distribution of X_{t+h} given X_t, for $h > 0$). The introduction of the stochastic integral made it possible to develop a theory of *stochastic differential equations* (SDEs). Whereas in the older approach the point of view is "collective" (looking at the population as a whole), stochastic differential equations represent an "individual" viewpoint in that they produce scenarios. This latter viewpoint is often more convenient for modeling and analysis.[13]

Let W_t denote a Wiener process. An expression of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \tag{2.11}$$

is called a *stochastic differential equation driven by Brownian motion*. This means that SDE uses a Brownian motion process as its source of uncertainty. One can also define SDEs that are driven by other processes than Brownian motion, but these will not be used in this book. The fluctuations of the process W_t and the passage of time are transferred to the process X_t in a way that is determined by the two functions μ and σ, which represent *drift* and *volatility* respectively.[14]

The formulation (2.11) still needs to be given a meaning, because neither the left hand side nor the right hand side has been defined as such. Adding integral signs to both sides leads to expressions that have already been defined, and this fact is used in the following definition.

Definition 2.3.1 A stochastic process X_t that is adapted to the Wiener process W_t is said to be a *solution* of the stochastic differential equation (2.11) if for all $t \geq 0$ we have

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s. \tag{2.12}$$

French mathematician who made early contributions to mathematical finance; see Section 1.1. The quote is from his PhD thesis, entitled *Théorie de la spéculation* (The Theory of Speculation).

[13] A framework built similarly on individual scenarios was actually already developed in 1940 by the French mathematician Vincent Doblin (1915–1940). This was discovered only in 2000, when his relatives gave permission to open a sealed letter that Doblin, while serving as a soldier in the French army, had sent to the Académie des Sciences in Paris. Doblin's construction generates scenarios as modifications of Brownian motion by means of a transformation of the time parameter. This is in general somewhat less convenient for purposes in finance than Itô's construction, although the idea of time change itself does have a financial interpretation ("business time") and has been incorporated into some financial models.

[14] More precisely, the function $\sigma(t, X_t)$ indicates how shocks of the driving Brownian motion W_t translate to shocks of the state process X_t. Therefore, σ should in fact be viewed as a correlation parameter. For instance, its value might well be negative.

Similar definitions apply to models that are driven by other processes than Brownian motion.

Intuitively, a stochastic differential equation represents a situation in which one process (the "driving process", which in (2.11) is the Brownian motion W_t) generates another process (the process X_t in the equation above). In such situations, the relation between the processes W_t and X_t is *nonanticipative*, which means that X_t at a given time t depends only on W_s for $s \leq t$. Therefore it is natural to require that the process $\{X_t\}$ should be adapted to the process $\{W_t\}$. There is also a technical reason for this requirement. If $\{X_t\}$ is adapted to $\{W_t\}$, then the processes $\{\mu(t, X_t)\}$ and $\{\sigma(t, X_t)\}$, which are just instantaneous transformations of $\{X_t\}$, are adapted to $\{W_t\}$ as well. This fact is needed to make sure that the integrals appearing in (2.12) are well-defined, since adaptedness of the integrand to the integrator is one of the requirements in the definition of the Itô integral.

According to the definition, to verify that a given process X_t is a solution of the SDE (2.11), one should do the following: compute the integral transforms $Y_t := \int_0^t \mu(s, X_s)\, ds$ and $Z_t := \int_0^t \sigma(s, X_s)\, dW_s$, and verify that $X_t = X_0 + Y_t + Z_t$ for all t. In practice this is not a very attractive method however. Usually, it is much more convenient to use the *Itô rule* that will be discussed in Chapter 2. There are a few cases in which it is possible to obtain explicit solutions for stochastic differential equations. These SDEs lead to tractable models which play an important role in the theory. Section 2.6 below presents the best known cases in which an explicit solution can be found.

When the drift term in (2.11) is zero, the solution is of the form of a constant plus an integral transform of the Wiener process:

$$X_t = X_0 + \int_0^t \sigma(s, X_s)\, dW_s.$$

It follows from Thm. 2.2.1 that in this case the solution X_t is a martingale, if the volatility function $\sigma(t, X_t)$ satisfies certain conditions.[15] Models in mathematical finance are usually constructed in such a way that these conditions are satisfied. Exceptions can be found for instance in models that are designed to describe stock market bubbles.

2.3.2 Euler discretization

If in (2.11) the infinitesimals are replaced by finite forward differences which we write as Δ, so that ΔX_t means by definition $X_{t+\Delta t} - X_t$ where X can be any

[15] For instance, the process generated by the SDE $dX_t = X_t^{1+\alpha}\, dW_t$, with $\alpha > 0$, is not a martingale. It is still a local martingale.

time-dependent quantity,[16] then we obtain

$$\Delta X_t = \mu(t, X_t)\Delta t + \sigma(t, X_t)\Delta W_t.$$

We may write this as

$$X_{t_{k+1}} = X_{t_k} + \mu(t_k, X_{t_k})\Delta t + \sigma(t_k, X_{t_k})\sqrt{\Delta t}\, Z_k \tag{2.13}$$

where $t_k = k\Delta t$ and the Z_k's form a sequence of independent standard normal variables. The time series model (2.13) is called the *Euler discretization*[17] after a similar method for ordinary differential equations. It can be shown that, under mild conditions, the solutions of (2.13) converge (in an appropriate sense for stochastic evolutions) to the solutions of (2.11) when the time step Δt tends to zero. Euler discretization can be applied to vector equations as well as to scalar equations. The method can be used for instance to compute approximations to quantities relating to the process X_t such as EX_T, EX_T^2, $E[X_{T_1}X_{T_2}]$, and so on.

The Euler discretization can be motivated by noting that the exact expression that links $X_{t_{k+1}}$ and X_{t_k} in the SDE (2.11) is

$$X_{t_{k+1}} \;=\; X_{t_k} + \int_{t_k}^{t_{k+1}} \mu(X_t)\, dt + \int_{t_k}^{t_{k+1}} \sigma(X_t)\, dW_t.$$

The scheme (2.13) is obtained when the integrand in the first integral is approximated by $\mu(X_{t_k})$ and in the second integral by $\sigma(X_{t_k})$.

Euler discretization is the workhorse of mathematical finance. It is by far the most frequently used method to combine the convenience of continuous-time models with the requirements of computation. While the corresponding method for ordinary (i.e. non-stochastic) differential equations, as originally devised by Euler, is nowadays considered to be mainly of historical interest, it turns out that the convergence properties of the Euler scheme benefit from the smoothing effect that is brought about by stochasticity.[18]

Implementation of the Euler method is in many cases straightforward; basically, the recipe is to replace the differential d by the discrete forward difference Δ everywhere. Nevertheless, there are some issues that may arise. One of these concerns

[16] Note that here we do *not* follow the convention that the use of subscript t indicates that the corresponding quantity is known at time t; according to this convention we should write $\Delta X_{t+\Delta t}$ rather than ΔX_t, and we should then interpret Δ as a *backward* difference operator.

[17] Leonhard Euler (1707–1783), Swiss mathematician.

[18] The relatively poor performance of the Euler scheme in the deterministic case is reflected in the stochastic case by the fact that the accuracy of the scheme tends to deteriorate when the volatility is small relative to the drift.

the preservation of theoretical properties. For instance, consider the SDE

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t, \qquad X_0 > 0. \tag{2.14}$$

It can be proved (see Section 2.6.1) that this generates a stochastic process that is always positive. However, the scenarios generated by the discretized version

$$X_{t_{k+1}} = X_{t_k} + \mu X_{t_k}\Delta t + \sigma X_{t_k}\sqrt{\Delta t}\, Z_k, \qquad Z_k \overset{\text{i.i.d.}}{\sim} N(0,1) \tag{2.15}$$

are not guaranteed to remain positive. In fact, since Z_k follows a normal distribution, there is at every step a positive probability that $X_{t_{k+1}}$ will be negative; once this occurs, it is quite likely that subsequent simulated values of X_t will be negative as well. For moderate values of the volatility σ, positive values of the drift parameter μ, and small values of the time step Δt, the probability of a breach of positivity is small, even when a large number of scenarios is simulated. To be sure, however, the basic scheme (2.15) might be amended, for instance by changing the sign of $X_{t_{k+1}}$ if in a particular scenario this quantity becomes negative. Note that replacing $X_{t_{k+1}}$ in such a case by the closest nonnegative value, namely 0, would not be suitable since this would cause all following values $X_{t_{k+i}}$ to be 0 as well. A way to avoid the problem altogether is to derive an SDE for $\log X_t$ (it is discussed in Section 2.4.3 how to do this) and to simulate using that SDE, rather than from the SDE for X_t itself. Afterwards, X_t can be recovered by taking the exponential, and positivity is guaranteed.

Another potential issue is related to the preservation of financial properties. Suppose for instance that one wants to generate scenarios for the evolution of portfolio value under a certain trading strategy. To be specific, suppose that a model is considered in which there are two traded assets whose time-t prices are denoted by S_t and B_t, and that a trading strategy is defined in terms of t and S_t. The trading strategy might prescribe to hold $\phi(t, S_t)$ units of assets S_t and $\psi(t, S_t)$ units of asset B_t at every time t, where the functions ϕ and ψ are defined in such a way that the strategy is self-financing (conditions for this property in continuous time are discussed in Section 3.1.2). A discretized scenario for portfolio value V_t could then be generated as follows (where expressions for $S_{t_{k+1}}$ and $B_{t_{k+1}}$ should be generated from a given model for the evolution of these prices):

$$\begin{aligned} S_{t_{k+1}} &= \cdots \\ B_{t_{k+1}} &= \cdots \\ V_{t_{k+1}} &= V_{t_k} + \phi(t_k, S_{t_k})(S_{t_{k+1}} - S_{t_k}) + \psi(t_k, S_{t_k})(B_{t_{k+1}} - B_{t_k}). \end{aligned} \tag{2.16}$$

However, the self-financing property that holds for the strategy (ϕ, ψ) in continuous

```
mu    = @(t,x) 0.05*x;  % drift
sigma = @(t,x) 0.2*x;   % volatility
X0    = 100;            % initial value
T     = 1;              % length of simulation interval
N     = 100;            % number of time steps
dt    = T/N;            % time step
Xs    = zeros(1,N+1);   % reservation of memory space
% --- initialization ---
t     = 0;           % initial time
Xs(1) = X0;
% --- time stepping ---
for k = 1:N
    X = Xs(k);       % read state value at time t(k)
    dW = sqrt(dt)*randn;
    dX = mu(t,X)*dt + sigma(t,X)*dW;
    X  = X+dX;
    t  = t+dt;
    Xs(k+1) = X;     % write state value at time t(k+1)
end
% --- output ---
plot(0:dt:T,Xs)
```

Code Example 2.1: Euler discretization for plotting of a single scenario.

time is only approximately true in the discretized version. The cumulative effect of small errors in successive time steps may lead to an undesirable perturbation of results. To avoid the numerical appearance or disappearance of money, it is better to write the approximation scheme in such a way that the budget constraint is enforced also in the discretized version. This can be ensured by replacing (2.16) by

$$V_{t_{k+1}} = V_{t_k} + \phi(t_k, S_{t_k})(S_{t_{k+1}} - S_{t_k}) + \frac{V_{t_k} - \phi(t_k, S_{t_k})S_{t_k}}{B_{t_k}}(B_{t_{k+1}} - B_{t_k}). \quad (2.17)$$

The above equation describes *exactly* the portfolio value at time t_{k+1} when the portfolio at time t_k has $\phi(t_k, S_{t_k})$ units of asset S_t and the rest of the value in asset B_t, assuming that these portfolio holdings are not changed between time t_k and t_{k+1}. This is still only an approximation to the continuous-time strategy (ϕ, ψ), but at least we are sure that no money is artificially created or lost through the discretization. In general, it is recommended, whenever possible, to construct discretizations in such a way that the discretized version has a sensible meaning by itself, in addition to being an approximation of a continuous-time model.

An example of Euler discretization is shown in Code Example 2.1. The script that is shown aims at plotting a single scenario. Plotting scenarios can be useful for instance as a reality check, but for computational purposes typically one needs many scenarios to get approximate values for instance of expectation and variance of quantities of interest at some given future time. One way to generate many

```
mu    = @(t,x) 0.05*x;  % drift
sigma = @(t,x) 0.2*x;   % volatility
X0    = 100;            % initial value
T     = 1;              % length of simulation interval
M     = 10^5;           % number of scenarios
dt    = 0.01;           % time step
% --- initialization ---
t = 0;
X = X0;
% --- time stepping ---
while t < T-0.5*dt   % subtract 0.5*dt for robustness
    dW = sqrt(dt)*randn(M,1);
    dX = mu(t,X)*dt + sigma(t,X).*dW;
    X  = X+dX;
    t  = t+dt;
end
% --- output ---
EX = mean(X);
sX = std(X);
disp(['calculated expected value: ' num2str(EX)])
disp(['estimated 95% confidence interval: [' ...
  num2str(EX-1.96*sX/sqrt(M)) ', ' num2str(EX+1.96*sX/sqrt(M)) ']' ])
```

Code Example 2.2: Euler discretization for computing the mean of a variable that is generated by a stochastic process. An estimated confidence interval is provided as well; this does not take into account the discretization error.

scenarios would be simply to take the code as shown in Code Example 2.1) and to write a loop around it in which the generation of a scenario is repeated as many times as needed. However, many programming languages make it possible to use *vector coding*, in which random variables can be represented as vectors whose entries correspond to possible outcomes. An example of code written in this way is shown in Code Example 2.2. Here it is assumed that the aim of the computation is to compute an approximation of $E[X_T]$ where X_t for $0 \leq t \leq T$ is described by a given SDE. The intermediate values of X_t are not needed for this purpose; therefore they are not stored, which saves memory space. Note also that the time stepping loop is handled in a bit different way than in Code Example 2.1. The calculation also provides an estimated 95% confidence interval, based on an application of the central limit theorem.[19]

[19]The astute reader might remark that, in the code, `1.96` should be replaced by `norminv(0.975)`, where `norminv` is the Matlab command for the inverse of the standard normal cumulative distribution function, and that `sqrt(M-1)` should be used instead of `sqrt(M)`. In view of the typical size of errors in calculation of expected values by means of simulations, such subtleties are usually ignored.

2.4 The univariate Itô rule

2.4.1 The chain rule for Riemann-Stieltjes integrals

The following statement is sometimes referred to as "the fundamental theorem of calculus": if f is a differentiable function, then for all t

$$f(t) = f(0) + \int_0^t f'(s)\,ds. \tag{2.18}$$

This is indeed a fundamental theorem because it states that the two basic operations of calculus, integration and differentiation, are each other's inverse. In the spirit of the notation that we have used for stochastic differential equations, the rule above may also be written in the form

$$df(t) = f'(t)\,dt \tag{2.19}$$

since, if integral signs are placed on both sides, one obtains

$$\int_0^t f'(s)\,ds = \int_0^t df(s) = f(t) - f(0)$$

where the final equality follows from the telescope rule that holds for Riemann-Stieltjes integrals as well as for Itô integrals (cf. (2.5)).

The integral in (2.18) is defined as a limit of sums, and so what is actually stated in the fundamental theorem of calculus is:

$$\lim_{\Delta t \downarrow 0} \sum_0^T f'(t_i)(t_{i+1} - t_i) = f(T) - f(0) \tag{2.20}$$

where the limit is taken in the sense of refining partitions on the interval $[0, T]$. The argument of f' is taken to be t_i in the above to preserve the parallel with the stochastic integral, although in the deterministic case the limit actually doesn't depend on the choice of the argument as long as it is in the interval from t_i to t_{i+1}.

The fundamental theorem of calculus may be extended in the context of Riemann-Stieltjes integrals. Here one integrates the given function $f(t)$ (the *integrand*) not against time itself but against a given function of time $g(t)$ (the *integrator*). In the theory of the Riemann-Stieltjes integral, it is required that the integrator g is of bounded variation, which means that there is a constant M such that $\sum |g(t_{i+1}) - g(t_i)| \leq M$ for all partitions $0 = t_1 < \cdots < t_{k+1} = T$. The Riemann-Stieltjes integral is defined by

$$\int_0^T f(t)\,dg(t) = \lim_{\Delta t \downarrow 0} \sum_0^T f(t_i)(g(t_{i+1}) - g(t_i)) \tag{2.21}$$

A special case arises when the integrand $f(t)$ is of the form $f(t) = \phi'(g(t))$ where ϕ is a smooth function. The function ϕ will be sometimes referred to as a *transformation* because it does not depend directly on time but rather operates on another function which does depend directly on time. If g is a continuous function of bounded variation, then one can prove that the following formula holds:

$$\lim_{\Delta t \downarrow 0} \sum_{0}^{T} \phi'(g(t_i))(g(t_{i+1}) - g(t_i)) = \phi(g(T)) - \phi(g(0)). \tag{2.22}$$

This may also be written as

$$\int_0^T \phi'(g(t))\, dg(t) = \phi(g(T)) - \phi(g(0)). \tag{2.23}$$

By the "telescope rule" (2.5), the right hand side can be written as $\int_0^T d(\phi(g(t)))$. Therefore, analogously to (2.19) the rule (2.23) may also as a way of notation be expressed by

$$d\phi(g(t)) = \phi'(g(t))\, dg(t). \tag{2.24}$$

This is a way of writing the *chain rule*, another very basic theorem of calculus. Since (2.24) is a generalization of (2.19), it incorporates the fundamental theorem of calculus as well.

2.4.2 Integrators of bounded quadratic variation

There exist continuous functions that are not of bounded variation on any interval. Such functions are highly irregular, and it is not easy to write down examples in explicit form. However, if one looks at the trajectories of Brownian motion, it turns out that (with probability 1) they fall in this peculiar class of functions that are not of bounded variation on any interval. To get an idea of how this comes about, note that the expected value of the absolute value of the increment of Brownian motion on an interval of length Δt between successive partition points is on average of order $\sqrt{\Delta t}$, since $E(\Delta W_t)^2 = \Delta t$. If the number of partition points is increased by a factor of 100, then the distance Δt between successive partition points is reduced by the same factor; the absolute values of the increments of the Brownian motion are on average only reduced by a factor of 10 however, and so the sum of the absolute values of the increments is increased by about a factor of 10. Therefore, there is no bound to the sum of the absolute values of the increments when the partitions are made finer and finer, and so the paths of Brownian motion are not of bounded variation.

That may be bad news, but there is also some good news, since the same reasoning suggests that the paths of Brownian motion are still of bounded *quadratic*

variation, that is, the sum of the squares of the increments corresponding to arbitrary sequences of partitions is bounded. This can indeed be proved,[20] and moreover it turns out that the property of bounded quadratic variation can be used to obtain a version of the chain rule (2.24) which is valid for Brownian motion and for a range of other stochastic processes. Let us see how this works.

The relation (2.22) states that, if g is a continuous function of bounded variation, then the long-term increment $\phi(g(T)) - \phi(g(0))$ can be approximated to arbitrarily high accuracy (by taking step sizes smaller) in terms of the short-term increments $g(t_{i+1}) - g(t_i)$ by means of the formula

$$\phi(g(T)) - \phi(g(0)) \approx \sum_{i=0}^{k} \phi'(g(t_i))(g(t_{i+1}) - g(t_i)). \tag{2.25}$$

To get an idea of why this is true and what can be done when g is only of bounded quadratic variation, use the discrete-time version of the telescope rule and write down a Taylor series expansion:

$$\begin{aligned} \phi(g(T)) - \phi(g(0)) &= \sum \phi(g(t_{i+1})) - \phi(g(t_i)) \\ &= \sum \phi'(g(t_i))(g(t_{i+1}) - g(t_i)) + \sum \tfrac{1}{2}\phi''(g(t_i))(g(t_{i+1}) - g(t_i))^2 + \cdots . \end{aligned} \tag{2.26}$$

The first term on the right is the one that also appears on the right hand side of (2.25), so the assumption that is made in the approximation (2.25) is that the other terms in (2.26) can be ignored. It can be shown that, if g is a continuous function of bounded variation, its quadratic variation $\sup \sum (g(t_{i+1} - g(t_i))^2$ (supremum over all partitions) is zero. This fact is what causes the right hand side of (2.25) to be equal to the left hand side in the limit. If the function g is not of bounded variation but is still of bounded quadratic variation, then it can be shown that the "cubic variation" $\sup \sum (g(t_{i+1} - g(t_i))^3$ is zero, which means that the terms of order three in (2.26) can be neglected. The same argument applies to terms of higher order than three. Consequently, the approximation formula is still valid for functions of bounded quadratic variation if we include the quadratic term in the approximation. There is no need to include any terms of higher order.

Suppose the function g is of bounded quadratic variation on $[0, T]$; then it is also of bounded quadratic variation on $[0, t]$ for any t such that $0 < t < T$. Introduce the function

$$[g, g](t) := \lim_{\Delta t \downarrow 0} \sum_{0}^{t} \big(g(t_{i+1}) - g(t_i)\big)^2. \tag{2.27}$$

[20] More precisely, the statement holds true with probability 1 with respect to the paths of Brownian motion, for any given sequence of partitions.

The function $[g, g]$ is called the *quadratic variation function* of g. It is a nondecreasing function of t. If the transformation ϕ is twice continuously differentiable and g is continuous and of bounded quadratic variation, then the approximation

$$\phi(g(T)) - \phi(g(0)) \approx \sum_{i=0}^{k} \left(\phi'(g(t_i))(g(t_{i+1}) - g(t_i)) + \tfrac{1}{2}\phi''(g(t_i))(g(t_{i+1}) - g(t_i))^2\right) \tag{2.28}$$

becomes exact in the limit when the number of partition points is increased, no matter in which way this is done, as long as the maximum distance between two successive partition points tends to zero. Moreover it can be proved that the same result is obtained if the squares of increments that appear in (2.28) are replaced by increments of the quadratic variation function. Then we are justified in writing

$$\phi(g(T)) - \phi(g(0)) = \int_0^T \left(\phi'(g(t))\, dg(t) + \tfrac{1}{2}\phi''(g(t))\, d[g, g](t)\right) \tag{2.29}$$

even though the usual conditions for Riemann-Stieltjes integration are not fulfilled. On the basis of the continuous telescope rule, we can also write differential version of the same formula:

$$d\phi(g(t)) = \phi'(g(t))\, dg(t) + \tfrac{1}{2}\phi''(g(t))\, d[g, g](t). \tag{2.30}$$

The differential version is shorter and is often more convenient to use than the integral form. The rule (2.30) is in fact a *generalization* of (2.24): if g is of bounded variation, the rule above is still valid and in fact reduces to (2.24), because for such a function the quadratic variation function $[g, g](t)$ is zero.

What is the meaning of these considerations in practice? Financial data are monitored at high frequencies but not truly continuously, and so one might say that the whole discussion about bounded variation simply does not apply to for instance stock prices. However, it is possible to ask whether the approximation in (2.25), at the highest frequencies that can be obtained in practice, is noticeably improved when it is replaced by (2.28). There is general agreement that it is indeed essential to include the second-order term (but not a third-order term), so that modeling in terms of the Brownian motion with its paths of bounded quadratic variation is appropriate, and the formula (2.30) should be used instead of the formula (2.24) which applies only to functions g of bounded variation.

2.4.3 First rules of stochastic calculus

It can be shown that the story above applies to stochastic processes as well. In this case, of course, the deterministic Riemann-Stieltjes integral is replaced by the stochastic integral. Instead of a quadratic variation function one defines a *quadratic*

variation process:

$$[X, X]_t = \lim_{\Delta t \downarrow 0} \sum_{0}^{t} (X_{t_{i+1}} - X_{t_i})^2. \tag{2.31}$$

We arrive at the following basic rule of stochastic calculus, which holds for transformations ϕ that are twice continuously differentiable and for continuous processes X_t of finite quadratic variation:

$$\boxed{d\phi(X) = \phi'(X)\, dX + \tfrac{1}{2}\, \phi''(X)\, d[X, X].} \tag{2.32}$$

For brevity, subscripts have been dropped in the formula above. The same notational convention will often be used below.

The formula (2.32) is known as the *Itô rule*. It is the key rule of stochastic calculus.

The rule (2.30) is not very useful unless one has the quadratic variation function in hand. The same holds for the stochastic version (2.32). It is therefore very fortunate that the quadratic variation process of Brownian motion is known and is in fact quite simple.

Theorem 2.4.1 *The value at time t of the quadratic variation process of Brownian motion is t.*

In general the value of the quadratic variation of a given process is a stochastic variable (in other words, the quadratic variation process is itself a stochastic process), but the theorem above states that in the case of Brownian motion the quadratic variation is a deterministic function of time and indeed a very simple one. The theorem may be written as a formula:

$$\boxed{[W, W]_t = t.} \tag{2.33}$$

This rule is often written in differential form:

$$\boxed{d[W, W]_t = dt.} \tag{2.34}$$

The quadratic variation process of more general process X_t can be determined by the rule below, which assumes that an expression is available for the stochastic differential of X_t. This happens in particular when X_t is given by a stochastic differential equation such as (2.11).

Theorem 2.4.2 *If $\{X_t\}$ satisfies $dX_t = Y_t\,dt + Z_t\,dW_t$ where both $\{Y_t\}$ and $\{Z_t\}$ are adapted to $\{W_t\}$, then the quadratic variation process of $\{X_t\}$ is given by*

$$\boxed{d[X,X]_t = Z_t^2\,dt.} \tag{2.35}$$

In particular, if X_t satisfies (2.11), then

$$\boxed{d[X,X]_t = \sigma^2(t,X_t)\,dt.} \tag{2.36}$$

2.4.4 Examples

Below are some simple examples of applications of the Itô rule. More general versions of the Itô rule and additional applications are discussed in the following section.

Example 2.4.3 Let f be the function $f(x) = x^2$, and let W_t denote Brownian motion as usual. According to the Itô rule, we have

$$d(W_t^2) = 2W_t\,dW_t + dt. \tag{2.37}$$

Taking integrals on both sides, we find $W_T^2 = 2\int_0^T W_t\,dW_t + T$ so that

$$\int_0^T W_t\,dW_t = \tfrac{1}{2}W_T^2 - \tfrac{1}{2}T. \tag{2.38}$$

Without much effort, we explicitly computed a stochastic integral. Note the difference with (1.17).

Example 2.4.4 Applying the Itô rule to the fourth power of Brownian motion, one finds

$$d(W_t^4) = 4W_t^3\,dW_t + 6W_t^2\,dt \tag{2.39}$$

so that

$$W_T^4 = 4\int_0^T W_t^3\,dW_t + 6\int_0^T W_t^2\,dt. \tag{2.40}$$

In particular, the expectations on both sides must be equal. The expectation of the stochastic integral is zero by Thm. 2.2.2, whereas

$$E\int_0^T W_t^2\,dt = \int_0^T EW_t^2\,dt = \int_0^T t\,dt = \tfrac{1}{2}T^2.$$

Taking $T = 1$, we find using (2.40) that $EW_1^4 = 3$. In other words, the fourth

moment of a standard normal variable is equal to 3. This well known fact can also be obtained by standard integral calculus.

Example 2.4.5 Consider the stochastic differential equation

$$dX_t = \mu X_t\,dt + \sigma X_t\,dW_t, \quad X_0 \text{ given} \tag{2.41}$$

where μ and σ are constants. Define Y_t by $Y_t = \log X_t$. According to the Itô rule, we have

$$dY = \frac{1}{X}\,dX - \frac{1}{2X^2}\,d[X,X]. \tag{2.42}$$

Since from (2.41) we have $dX = \mu X\,dt + \sigma X\,dW$ and $d[X,X] = \sigma^2 X^2\,dt$, the equation above can be rewritten as

$$dY = (\mu - \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dW_t. \tag{2.43}$$

This equation can be solved by direct integration, since the right hand side does not depend on Y. The solution is

$$Y_t = Y_0 + (\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t. \tag{2.44}$$

Because $Y_t = \log X_t$ we have conversely $X_t = \exp Y_t$, and it can be concluded that the process X_t defined by

$$X_t = X_0\,\exp[(\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t] \tag{2.45}$$

solves the stochastic differential equation (2.41). The process defined by (2.41) or equivalently by (2.45) is called the *geometric Brownian motion*. As is seen from (2.45), the process only takes positive values, and its distribution at any given time t is lognormal. The geometric Brownian motion is the most popular model of mathematical finance. The use of this model for the evolution of stock prices was initiated by Paul Samuelson.[21]

2.4.5 Variance of the stochastic integral

Let X_t be an integral transform of Brownian motion:

$$X_t = \int_0^t Z_s\,dW_s \tag{2.46}$$

where the process $\{Z_t\}$ is adapted to $\{W_t\}$ and satisfies suitable boundedness assumptions. We already know that $EX_t = 0$ for all t (Theorem 2.2.2, "you can't

[21] Paul A. Samuelson (1915–2009), American economist; Nobel prize 1970.

beat the system"). Using the Itô calculus, we can also compute the variance of X_t. The process $\{X_t\}$ satisfies $dX_t = Z_t\,dW_t$. According to the Itô rule and Thm. 2.4.2, we have $dX^2 = 2X\,dX + d[X,X] = 2ZX\,dW + Z^2\,dt$, so that

$$EX_T^2 = E\int_0^T dX_t^2 = 2E\int_0^T Z_t X_t\,dW_t + E\int_0^T Z_t^2\,dt = \int_0^T EZ_t^2\,dt. \tag{2.47}$$

In this way we find:

$$\boxed{\operatorname{var}\Big(\int_0^T Z_t\,dW_t\Big) \;=\; \int_0^T EZ_t^2\,dt.} \tag{2.48}$$

Suppose now that the process Z_t is in fact a deterministic function $g(t)$. The stochastic integral $\int_0^T g(s)\,dW_s$ is the limit of stochastic variables of the form $\sum g(t_i)(W_{t_{i+1}} - W_{t_i})$; because the coefficients $g(t_i)$ are deterministic and the Brownian increments $W_{t_{i+1}} - W_{t_i}$ are normally distributed, each of these variables is a linear combination of normal variables and therefore is itself normally distributed. As a result, the limit $\int_0^T g(s)\,dW_s$ also follows a normal distribution. We know that its expectation is zero, and its variance is given by (2.48). Therefore the distribution of the stochastic integral with deterministic integrand $\int_0^T g(s)\,dW_s$ is completely determined. We have:

$$\boxed{\int_0^T g(t)\,dW_t \;\sim\; N(0,\sigma^2) \quad \text{with} \quad \sigma^2 = \int_0^T g^2(t)\,dt.} \tag{2.49}$$

The stochastic variable $\int_0^T Z_t\,dW_t$ is in general not normally distributed when Z is not deterministic. For instance, it is seen from (2.38) that $\int_0^1 W_t\,dW_t$ follows a shifted χ^2 distribution with one degree of freedom.

2.5 The multivariate Itô rule

The standard chain rule has a multivariate version, which makes use of partial derivatives. The notation is a bit more complicated than in the single-variate case. Suppose that $\phi(x_1,\ldots,x_n)$ is a smooth real-valued function of n variables, and let its partial dervatives be denoted by $(\partial\phi/\partial x_i)(x_1,\ldots,x_n)$. If $g_i(t)$ $(i = 1,\ldots,n)$ are functions of bounded variation, then the composite function $f(t) := \phi(g_1(t),\ldots,g_n(t))$ is a real-valued function of t, and the following version of the chain rule holds:

$$df(t) = \sum_{i=1}^n \frac{\partial\phi}{\partial x_i}(g_1(t),\ldots,g_n(t))\,dg_i(t). \tag{2.50}$$

The formula may look more easily digestible when it is written in shorthand notation, as follows:

$$df = \sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i}\, dg_i. \tag{2.51}$$

If the functions g_i are only assumed to be of finite quadratic variation, then the second-order terms have to be taken into account. The resulting formula involves not only quadratic variation functions but also quadratic *covariations* between the functions g_i. In general, the quadratic covariation of two functions g_1 and g_2 is defined by

$$[g_1, g_2](t) = \lim_{\Delta t \downarrow 0} \sum_{0}^{t} (g_1(t_{i+1}) - g_1(t_i))(g_2(t_{i+1}) - g_2(t_i)). \tag{2.52}$$

It should be noted that the notation is consistent; that is, the quadratic variation as defined in (2.27) is recovered in the case that $g_1 = g_2 = g$. The multivariate chain rule in the case of functions of finite quadratic variation reads as follows: if $f(t) = \phi(g_1(t), \ldots, g_n(t))$, then (in shorthand notation)

$$df = \sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i}\, dg_i + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 \phi}{\partial x_i \partial x_j}\, d[g_i, g_j](t). \tag{2.53}$$

Essentially the same equation holds for processes of finite quadratic variation. The *multivariate Itô rule* is stated as follows, again in shorthand notation: if $Y = \phi(X_1, \ldots, X_n)$, then

$$\boxed{dY = \sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i}\, dX_i + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 \phi}{\partial x_i \partial x_j}\, d[X_i, X_j].} \tag{2.54}$$

In longhand notation, this would read as follows:

$$dY_t = \sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i}(X_{1,t}, \ldots, X_{n,t})\, dX_{i,t} + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 \phi}{\partial x_i \partial x_j}(X_{1,t}, \ldots, X_{n,t})\, d[X_i, X_j]_t. \tag{2.55}$$

2.5.1 Nine rules for computing quadratic covariations

In order to work with the multivariate Itô rule, one needs to be able to compute quadratic covariations. This section provides a number of calculus rules for this purpose. In the rules given below, the term "process" is always understood as "semimartingale" (see Section 2.2.1). It can be shown that semimartingales do have well-defined quadratic variation and covariation processes.

(i) For all X and Y, one has $[X, Y] = [Y, X]$.

(ii) For all X, Y, and Z, one has $[X, Y + Z] = [X, Y] + [X, Z]$.

(iii) If a is a constant, then $[aX, Y] = a[X, Y]$.

(iv) The quadratic covariation of any semimartingale with a continuous process of bounded variation is zero. In particular, $[X, g] = 0$ whenever g is a deterministic continuous function of bounded variation.

(v) If W_1 and W_2 are Wiener processes with correlation coefficient ρ, then

$$d[W_1, W_2] = \rho\, dt. \tag{2.56}$$

In particular, if W_1 and W_2 are independent Brownian motions, then $d[W_1, W_2] = 0$.

(vi) If $dX = Y dZ$, then

$$d[X, V] = Y d[Z, V]. \tag{2.57}$$

(vii) If $dX_1 = \sum_{i=1}^{k} Y_i^{(1)} dZ_i^{(1)}$ and $dX_2 = \sum_{i=1}^{m} Y_i^{(2)} dZ_i^{(2)}$ then

$$d[X_1, X_2] = \sum_{i=1}^{k} \sum_{j=1}^{m} Y_i^{(1)} Y_j^{(2)} d[Z_i^{(1)}, Z_j^{(2)}]. \tag{2.58}$$

This follows by repeated application of (vi) and the linearity rule (ii). In particular, if $dX_1 = Y_1\, dt + Z_1\, dW_1$ and $dX_2 = Y_2\, dt + Z_2\, dW_2$, then

$$d[X_1, X_2] = Z_1 Z_2 \rho\, dt \tag{2.59}$$

where ρ is the correlation coefficient of the Wiener processes W_1 and W_2.

(viii) The quadratic covariation process of two semimartingales is itself a process of bounded variation. If X is continuous and Y is continuous, then $[X, Y]$ is continuous as well. In particular, we have

$$[X, [Y, Z]] = 0 \tag{2.60}$$

when X is a semimartingale and Y and Z are continuous semimartingales.[22]

[22] This property does not necessarily hold when the semimartingales involved are not continuous. For instance, if X_t is the standard Poisson process (with jumps of size 1), then $[X, X]_t = X_t$, and consequently $[X, [X, X]]_t = X_t$.

(ix) If $\phi(z_1, \ldots, z_n)$ is a smooth function of n variables and the processes X and $Z_1, \ldots, Z_n$ are continuous semimartingales, then

$$d[X, \phi(Z_1, \ldots, Z_n)] = \sum_{i=1}^{n} \frac{\partial \phi}{\partial z_i}(Z_1, \ldots, Z_n)\, d[X, Z_i]. \tag{2.61}$$

This follows from the Itô rule and from rules (vi) and (viii). In particular, we have

$$d[X, YZ] = Y\, d[X, Z] + Z\, d[X, Y]. \tag{2.62}$$

As a way of remembering the rules above, one can use the following informal expressions, in which $d[X, Y]$ is written as $dX\, dY$:

(i) $(dt)^2 = 0$

(ii) $dt\, dW = 0$

(iii) $(dW)^2 = dt$

(iv) $dW_1\, dW_2 = \rho\, dt$.

Applying these expressions together with linearity, one finds for instance

$$(\mu_X\, dt + \sigma_X\, dW_1)(\mu_Y\, dt + \sigma_Y\, dW_2) = \sigma_X \sigma_Y \rho\, dt$$

which is an alternative way of writing (2.59).

2.5.2 More examples

Example 2.5.1 The rule that governs differentials of products of processes can be obtained as a special case of the multivariate Itô rule. Indeed, the product function $\phi(x, y) = xy$ is a function of two variables, and its second partial derivatives are given by

$$\frac{\partial^2 f}{\partial x^2}(x, y) = \frac{\partial^2 f}{\partial y^2}(x, y) = 0, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = 1.$$

From the general formula (2.54), it follows that

$$\boxed{d(XY) = Y\, dX + X\, dY + d[X, Y].} \tag{2.63}$$

This is known as the *stochastic product rule*. Compare this rule to the product rule of standard calculus, which can be written in the form $d(fg) = f\, dg + g\, df$. Similarly, one can derive for instance also a *stochastic quotient rule*.

Example 2.5.2 Suppose that we have a scalar process X_t and we form a new process Y_t by $Y_t = \phi(t, X_t)$ where $\phi(t, x)$ is a smooth function of two variables. We can then find dY by applying the multivariate Itô rule. Since t is a deterministic function of bounded variation, the quadratic variation $[t, t]$ and the quadratic covariation $[X, t]$ are both zero. The only Itô correction term that remains is therefore the one that involves the second derivative of ϕ with respect to x. As a result, we have the formula

$$\boxed{dY = \frac{\partial \phi}{\partial t}\, dt + \frac{\partial \phi}{\partial x}\, dX + \frac{1}{2}\frac{\partial^2 \phi}{\partial x^2}\, d[X, X]} \tag{2.64}$$

which is called the *time-dependent Itô rule*.

Example 2.5.3 Similarly to how (2.48) was obtained, one can derive the following formula:

$$\operatorname{cov}\Big(\int_0^T X_t\, dW_t, \int_0^T Y_t\, dW_t\Big) = \int_0^T E(X_t Y_t)\, dt. \tag{2.65}$$

In particular, for two stochastic integrals with deterministic integrands:

$$\operatorname{cov}\Big(\int_0^T f(t)\, dW_t, \int_0^T g(t)\, dW_t\Big) = \int_0^T f(t) g(t)\, dt. \tag{2.66}$$

This generalizes the variance rule as stated in (2.49). More generally, one can obtain in a similar way the distribution of a random vector of the form $\int_0^T F(t)\, dW_t$ where $F(t)$ is a deterministic $n \times k$ matrix function, and W_t is a k-dimensional Brownian motion.

2.6 Explicitly solvable SDEs

As is also the case for ordinary differential equations, only some special types of stochastic differential equations allow explicit solutions. In the case of SDEs, one usually says that a solution is "explicit" if it is obtained by a direct operation on a (vector) Brownian motion.

2.6.1 The geometric Brownian motion

It was already found in Example 2.4.5 that the solution of the stochastic differential equation

$$dX_t \;=\; \mu X_t\, dt + \sigma X_t\, dW_t, \qquad X_0 \text{ given} \tag{2.67}$$

is provided by the geometric Brownian motion process:

$$X_t = X_0 \exp\big((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t\big). \tag{2.68}$$

In the example, the solution was derived by a log transformation of the SDE, which resulted in an easily solvable linear equation. However, using the time-dependent Itô rule (2.64), one can also verify directly that (2.68) solves (2.67). Indeed, given (2.68), one can write (think of X_t as $X_t = f(t, W_t)$ with $f(t, x) = X_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma x)$):

$$dX_t = (\mu - \tfrac{1}{2}\sigma^2)X_t\,dt + \sigma X_t\,dW_t + \tfrac{1}{2}\sigma^2 X_t\,dt = \mu X_t\,dt + \sigma X_t\,dW_t.$$

The expectation and variance of the geometric Brownian motion can be obtained from the standard formulas for the expectation and variance of a lognormal variable, which are

$$E\big[e^{a+bZ}\big] = e^{a+\frac{1}{2}b^2}, \quad \operatorname{var}\big(e^{a+bZ}\big) = \big(e^{b^2} - 1\big)e^{2a+b^2} \qquad (Z \sim N(0,1)). \tag{2.69}$$

For the process given by (2.45), this implies

$$E[X_t] = e^{\mu t}X_0, \quad \operatorname{var}(X_t) = \big(e^{\sigma^2 t} - 1\big)e^{2\mu t}X_0^2. \tag{2.70}$$

The expression for the expectation can alternatively be found by noting that (2.67) implies

$$X_t - X_0 = \int_0^t \mu X_s\,ds + \int_0^t \sigma X_s\,dW_s.$$

Taking expectations on both sides, one gets

$$E[X_t] - X_0 = E\int_0^t \mu X_s\,ds = \mu\int_0^t E[X_s]\,ds.$$

By differentiation, one obtains an ordinary (i.e. non-stochastic) differential equation for the deterministic function $E[X_t]$:

$$\frac{d}{dt}E[X_t] = \mu E[X_t].$$

The solution of this is the exponential function shown in (2.70). The expression for the variance can be obtained in a similar way, without using the explicit form of the solution. Given that an expression for the expectation is already available, it is enough to find an expression for the second moment. From (2.67), it follows that

$$dX_t^2 = 2X_t\,dX_t + d[X,X]_t = (2\mu + \sigma^2)X_t^2\,dt + 2\sigma X_t^2\,dW_t.$$

This leads to the ODE

$$\frac{d}{dt}E[X_t^2] = (2\mu + \sigma^2)E[X_t^2]$$

from which it follows that $E[X_t^2] = e^{(2\mu+\sigma^2)t}X_0^2$. Subtracting the square of the expectation, one finds the expression for the variance in (2.70).

2.6.2 The Ornstein-Uhlenbeck process

If in a stochastic differential equation for X_t (possibly a vector) the drift is a linear function of X_t and the volatility does not depend on X_t, then one speaks of a *linear* SDE. An example of a stochastic differential equation of this type is provided by the one-dimensional SDE

$$dX_t = -aX_t\,dt + \sigma\,dW_t. \tag{2.71}$$

The minus sign is used here because typically in applications the coefficient multiplying X_t in the drift term is negative. As is usually the case with differential equations, solving (2.71) requires a little trick: write

$$\begin{aligned} & dX_t + aX_t\,dt = \sigma\,dW_t \\ \Leftrightarrow \quad & e^{at}(dX_t + aX_t\,dt) = e^{at}\sigma\,dW_t \\ \Leftrightarrow \quad & d\big(e^{at}X_t\big) = e^{at}\sigma\,dW_t \end{aligned}$$

Now apply the telescope rule. One finds

$$X_t = e^{-at}X_0 + \int_0^t e^{-a(t-s)}\sigma\,dW_s. \tag{2.72}$$

This is the sum of a deterministic function and a stochastic integral with deterministic integrand. The distribution of X_t for any given t is therefore *normal* with expectation $EX_t = e^{-at}X_0$ and variance

$$\operatorname{var}(X_t) = \int_0^t e^{-2a(t-s)}\sigma^2\,ds = \frac{1-e^{-2at}}{2a}\,\sigma^2.$$

In particular, if the coefficient a is positive, then

$$\lim_{t\to\infty} \operatorname{var}(X_t) = \frac{\sigma^2}{2a}\,.$$

Expressions for the first and second moment could also have been derived, without solving the SDE, by the technique of constructing ordinary differential equations for these quantities.

The process (2.72), with a positive, is called the *Ornstein-Uhlenbeck process*[23] (OU process) with parameters a and σ. When the coefficient a is positive, the drift term in (2.71) is negative when X_t is positive and positive when X_t is negative, so that the process has the tendency to go down when it is above zero and to go up

[23] Leonard S. Ornstein (1880–1941), Dutch physicist. George E. Uhlenbeck (1900–1988), Dutch/US physicist.

when it is below zero. The process is said to be *mean-reverting*. The term "mean" refers here to the value taken by EX_t as t tends to infinity, which is 0 as noted above. It is possible to construct a process that reverts to a nonzero mean simply by adding a constant to the process described by (2.71). The stochastic differential equation is in this case replaced by

$$dX_t \;=\; a(c - X_t)\,dt + \sigma\,dW_t \tag{2.73}$$

where c is a constant representing the mean that the process reverts to. This generalized version is also called an OU process.

A process $\{X_t\}$ is said to be *stationary* if, for any set of time indices $t_1, \ldots, t_k$ and any $h > 0$, the joint distribution of $X_{t_1}, \ldots, X_{t_k}$ is the same as the joint distribution of $X_{t_1+h}, \ldots, X_{t_k+h}$. Intuitively, this means that the process looks the same everywhere; in other words, it is not possible, by looking at the trajectories within a fixed time window, to tell where on the time axis this window is located. The OU process is not stationary, but it is asymptotically stationary, meaning that, for increasingly larger values of t, the process satisfies the stationarity conditions increasingly closely. The reason why the process is not stationary is that, for small values of t, the variance is smaller than the limit value, because we let the process start from a deterministic initial condition; also, the expected value may deviate from the asymptotic value 0 if the initial condition is not 0. One can get a strictly stationary version of the OU process by taking the initial condition not as a deterministic number, but as a draw from a normal distribution with mean 0 and variance $\frac{\sigma^2}{2a}$.

2.6.3 Higher-dimensional linear SDEs

Solution formulas similar to (2.72) can be given for the broad class of linear stochastic differential equations of the form

$$dX_t = (AX_t + g(t))\,dt + B(t)\,dW_t \tag{2.74}$$

where A is a constant matrix, $g(t)$ is a vector function that depends deterministically on time, and $B(t)$ is a matrix depending deterministically on time. As usual, W_t is a vector Brownian motion. This equation may be compared to the deterministic differential equation

$$\frac{dx}{dt}(t) = Ax(t) + g(t)$$

where $x(t)$ is a deterministic vector function of time. In the scalar case, an equation of the form

$$\frac{dx}{dt}(t) = ax(t) + g(t)$$

can be solved by writing it as

$$\frac{d}{dt}(e^{-at}x(t)) = e^{-at}g(t)$$

which leads to

$$x(t) = e^{at}x(0) + \int_0^t e^{a(t-s)}g(s)\,ds. \tag{2.75}$$

Essentially the same method is applied in the vector case, also in the situation (2.74) in which there is a stochastic forcing term.[24] To write the solution in a compact form we need the notion of the *matrix exponential function*, which is defined by

$$e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \tag{2.76}$$

where A is a square matrix. It can be verified that the sum converges for any matrix A, and that the solution of the vector differential equation

$$\frac{dx}{dt}(t) = Ax(t) \tag{2.77}$$

is given by

$$x(t) = e^{At}x(0).$$

Using this notation, the equation (2.74) can now be solved by essentially the same method as was used in the case of the OU process. Rewrite (2.74) as

$$d(e^{-At}X_t) = e^{-At}g(t)\,dt + e^{-At}B(t)\,dW_t. \tag{2.78}$$

On the left hand side we apply the stochastic product rule, which however in this case is the same as the deterministic product rule because the entries of the matrix e^{At} are continuous functions of bounded variation. We obtain the following solution of (2.74):

$$X_t = e^{At}X_0 + \int_0^t e^{A(t-s)}g(s)\,ds + \int_0^t e^{A(t-s)}B(s)\,dW_s. \tag{2.79}$$

The expression (2.79) indicates that the solution of the SDE (2.74) can be written as the sum of a deterministic function and a stochastic integral with deterministic integrand, which defines a normally distributed random variable. Therefore we can conclude that at any time t, *the value X_t of the solution of the stochastic differential equation (2.74) is normally distributed*. The same conclusion still holds if the initial condition X_0 is stochastic rather than deterministic, as long as X_0 is drawn from a normal distribution. Also, not only can we say that X_t follows a normal distribution,

[24] A "forcing term" in a differential equation is a function of time that appears in the equations and that is determined exogenously, i.e. not as part of the solution of the differential equation.

but also that any sequence of values $(X_{t_1}, \ldots, X_{t_k})$, where $t_1, \ldots, t_k$ are given time points, forms a jointly normally distributed set of random variables. A process that has this property is called a *Gaussian process*.[25] Such processes can be described completely in terms of expectations and covariances.

Let us see what we can say, for a given time point t, about the expectation and the variance of the random variable X_t defined by (2.74). Since the expectation of the stochastic integral in (2.79) is zero (indeed, it is an integral with respect to a martingale), we have

$$EX_t = e^{At}EX_0 + \int_0^t e^{A(t-s)}g(s)\,ds. \tag{2.80}$$

This shows that $m(t) := EX_t$ as a function of t satisfies the deterministic linear differential equation

$$\frac{dm}{dt}(t) = Am(t) + g(t) \tag{2.81}$$

which could also have been obtained directly from (2.74) in a similar way as discussed in the case of the GBM.

To determine the variance of X_t, introduce $Z_t := X_t - m(t)$ and note that Z_t satisfies the stochastic differential equation

$$dZ_t = AZ_t\,dt + B(t)\,dW_t. \tag{2.82}$$

Let the variance of X_t (the variance-covariance matrix if X_t is multivariate) be denoted by $H(t) = EZ_tZ_t^\top$. Since it follows from (2.82) by the product rule that

$$d(Z_tZ_t^\top) = \big(AZ_tZ_t^\top + Z_tZ_t^\top A^\top + B(t)B^\top(t)\big)dt + \big(\cdots\big)dW_t \tag{2.83}$$

(where the precise form of the volatility term is unimportant for our present purposes), the following deterministic differential equation holds for $H(t)$:

$$\frac{dH}{dt}(t) = AH(t) + H(t)A^\top + B(t)B^\top(t). \tag{2.84}$$

It can be verified by direct calculation that the solution of this equation is given by

$$H(t) = e^{At}H(0)e^{A^\top t} + \int_0^t e^{A(t-s)}B(s)B^\top(s)e^{A^\top(t-s)}\,ds. \tag{2.85}$$

This could also have been obtained from the solution formula (2.79) by making use of the rule (2.49) for the variance of a stochastic integral with deterministic integrand.

[25] Carl Friedrich Gauss (1777–1855), German mathematician. "Gaussian distribution" is an alternative name for the normal distribution.

From the explicit solution formula for X_t, one may compute what is called the *autocovariance* of the process, namely

$$\operatorname{cov}(X_t, X_s) = e^{A(t-s)}H(s) \quad (t \geq s). \tag{2.86a}$$

The expression above holds for $s \leq t$. Using the fact that in general $\operatorname{cov}(Z_1, Z_2) = \operatorname{cov}(Z_2, Z_1)^\top$, one can derive that

$$\operatorname{cov}(X_t, X_s) = H(t)e^{A^\top(s-t)} \quad (t \leq s). \tag{2.86b}$$

Suppose now that the initial condition $X(0)$ is itself a random variable which follows a normal distribution with expectation 0 and variance-covariance matrix H, where H is a solution of the matrix equation

$$AH + HA^\top + BB^\top = 0. \tag{2.87}$$

Then the differential equation (2.84) is solved by the constant function $H(t) = H$. From (2.86), it follows that in this case the covariance matrix of X_t and X_s depends only on the time difference $t - s$; in other words, the covariance of X_t and X_s is the same as the covariance of X_{t+h} and X_{s+h}, for any $h \geq 0$. It follows that the process defined by

$$dX_t = AX_t\,dt + B\,dW_t, \qquad X_0 \sim N(0, H) \tag{2.88}$$

where H satisfies the matrix equation (2.87) is an example of a stationary process (its statistical properties do not change over time).

The matrix equation (2.87) is known as the *Lyapunov equation*.[26] The left hand side is a linear transformation of the unknown H and so (2.87) is a linear equation. The equation does not always have a nonnegative definite solution. The conditions under which a nonnegative definite solution exists are closely related to the conditions under which the solutions of the deterministic equation (2.77) are bounded as t tends to infinity; the matrix A needs to be *stable*, which implies that it should not have any eigenvalues with positive real part.[27]

2.7 Girsanov's theorem

Taking expectations under different measures is an important notion in finance. It is shown in financial theory that risk can be taken into account by a pricing rule based on expectation under a measure that is different from the real-world measure.

[26] Aleksandr Michailovich Lyapunov (1857–1918), Russian mathematician.

[27] Boundedness of the solutions of the equation (2.77) was the original purpose for which Lyapunov developed the equation named after him.

Changes of measure can be expressed by the so-called *Radon-Nikodym derivative.*[28] If P and Q are given measures, and if there exists a random variable θ such that

$$E^Q X = E^P \theta X \tag{2.89}$$

for all random variables X, then θ is said to be the Radon-Nikodym derivative of Q with respect to P, and one also writes $\frac{dQ}{dP}$ instead of θ. For example, in the case of random variables that take only finitely many different values, the Radon-Nikodym derivative gives, for each possible outcome, the ratio of the Q-probability versus the P-probability. The expectation under P of the Radon-Nikodym derivative θ must be equal to 1, because $E^P\theta = E^Q 1 = 1$; so the average value of θ (in the sense of P) is equal to 1. This means that if θ is smaller than 1 in some regions of the outcome space, it must be larger than 1 in other regions; the change of measure is a shift of probability mass.

The idea of a change of measure can be applied not only for random variables but also for random processes. The change of measure is then not indicated by a random variable but by a stochastic process, called the *Radon-Nikodym process.* The Radon-Nikodym process $\{\theta_t\}$ is an adapted process such that, for all $0 \le s \le t$,

$$E_s^Q X_t = E_s^P \frac{\theta_t}{\theta_s} X_t \tag{2.90}$$

where X_t is a stochastic variable whose value is known at time t, and the subscript s indicates that expectation is taken conditional on information available at time s. The change of measure between time points s and t depends in general on both time points, as indicated above. The quotient form θ_t/θ_s ensures that the law of iterated expectations holds:

$$E_{t_1}^Q E_{t_2}^Q X_{t_3} = E_{t_1}^P \frac{\theta_{t_2}}{\theta_{t_1}} E_{t_2}^P \frac{\theta_{t_3}}{\theta_{t_2}} X_{t_3} = E_{t_1}^P \frac{\theta_{t_3}}{\theta_{t_1}} X_{t_3} = E_{t_1}^Q X_{t_3} \tag{2.91}$$

for $0 \le t_1 \le t_2 \le t_3$. In analogy with the fact that the P-expectation of the Radon-Nikodym derivative must be equal to 1, the Radon-Nikodym process satisfies $E_s^P \frac{\theta_t}{\theta_s} = E_s^Q 1 = 1$, or in other words $E_s^P \theta_t = \theta_s$; that is to say, the Radon-Nikodym process is a P-martingale. Moreover, just as in the case of random variables, the process must be positive for the two measures P and Q to be equivalent.

Stochastic differential equations are used as models for the evolution of stock prices and other variables of interest in financial markets, but they also offer a convenient way to generate positive martingales, and hence to describe a change of measure. Specifically, given a k-vector continuous semimartingale λ_t, one can

[28] Johann Radon (1887–1956), Austrian mathematician. Otton M. Nikodym (1887–1974), Polish mathematician.

generate from this a new process θ_t by means of the stochastic differential equation

$$d\theta_t = -\theta_t \lambda_t^\top dW_t, \qquad \theta_0 = 1 \tag{2.92}$$

where W_t is a k-vector Brownian motion. The absence of a drift term in (2.92) implies that the solution θ_t is a local martingale; if the process λ_t satisfies suitable boundedness conditions, then θ_t is a true martingale. Moreover, since the process starts at 1 and its volatility becomes small as θ_t approaches zero, the process θ_t remains positive for all time t, again assuming certain boundedness conditions on λ_t. In fact, one can show that, in a model in which the stream of incoming information is given by the vector process W_t, every continuous positive martingale can be written as the solution of an SDE of the form (2.92), so that in fact there is no loss in generality in considering changes of measure that are generated by equations of the form (2.92). In fact one may think of (2.92) as providing an alternative parametrization of a change of measure, in terms of the process λ_t. There are advantages to using λ_t rather than θ_t, because θ_t is subject to rather strict conditions (it has to be a positive martingale), whereas λ_t can be quite a general process.[29] As will be seen in the following chapters, there are also quite good reasons from the perspective of financial modeling to work with λ_t rather than with θ_t directly.

Since in this book we work with models that are driven by Brownian motion, the effect of a change of measure on these models is captured completely if we can describe its effect on the Brownian motion process. The characteristics of Brownian motion are described in terms of its increments in Def. 2.1.1. The properties listed there are stated in terms of expectation and variance, and so they are sensitive to the choice of the probability measure. To be more precise about this, let us look at what happens to the distribution of increments of a Brownian motion process W_t when a change of measure is applied (from the original measure P to the new measure Q) that is induced by a process λ_t, as described in (2.92). Assume for simplicity that the process W_t is scalar ($k = 1$); the vector case is not essentially different.

Generally speaking, the distribution of many random variables can be investigated by means of the *moment generating function*, which is defined by $M_X(\alpha) = E[e^{\alpha X}]$ (if the expectation is finite), where $\alpha \in \mathbb{R}$ and X is a given random variable. For instance, if one finds that the moment generating function of a random variable is of the form $\exp(\alpha\mu + \frac{1}{2}\alpha^2\sigma^2)$, then that proves that this variable follows a normal distribution with expectation μ and variance σ^2. To study the distribution of the increments of W_t under Q, one can therefore look at the function $E^Q[\exp(\alpha \Delta W_t)]$,

[29] In particular, there are no sign conditions on λ_t. The minus sign on the right hand side in (2.92) is consequently redundant, since λ_t could be replaced by $-\lambda_t$. It is however conventional to use a minus sign here (see Thm. 2.7.1).

where $\Delta W_t = W_{t+\Delta t} - W_t$, t is a fixed point in time, and Δt is a (small) time step. By definition, we have

$$\begin{aligned} E_t^Q\big[\exp(\alpha \Delta W_t)\big] &= E_t^P\Big[\frac{\theta_{t+\Delta t}}{\theta_t}\,\exp(\alpha \Delta W_t)\Big] \\ &= E_t^P\big[\exp(\log\theta_{t+\Delta t} - \log\theta_t + \alpha\Delta W_t)\big]. \end{aligned}$$

Since the increment of the process $\log\theta_t$ appears here, it is natural to employ the SDE for this process that is obtained by applying Itô's formula to (2.92):

$$d(\log\theta_t) = -\tfrac{1}{2}\lambda_t^2\,dt - \lambda_t\,dW_t.$$

Using the assumption that the process λ_t is continuous, one can write

$$\log\theta_{t+\Delta t} - \log\theta_t \approx -\tfrac{1}{2}\lambda_t^2\Delta t - \lambda_t \Delta W_t$$

for small Δt. By the standard formula $E[e^{aZ}] = e^{\frac{1}{2}a^2}$ for $Z \sim N(0,1)$, one finds

$$\begin{aligned} E_t^P \exp(-\tfrac{1}{2}\lambda_t^2\Delta t - \lambda_t\Delta W_t + \alpha\Delta W_t) &= \exp(-\tfrac{1}{2}\lambda_t^2\Delta t + \tfrac{1}{2}(\alpha-\lambda_t)^2\Delta t) \\ &= \exp(-\alpha\lambda_t\Delta t + \tfrac{1}{2}\alpha^2\Delta t). \end{aligned}$$

The conclusion from the calculations is that

$$E_t^Q\big[\exp(\alpha\Delta W_t)\big] \approx \exp\big(-\alpha\lambda_t\Delta t + \tfrac{1}{2}\alpha^2\Delta t\big)$$

where the approximation is increasingly accurate as Δt becomes smaller. In other words, for small increments, the distribution of the increment ΔW_t under the new measure Q that is described by (2.92), as seen from time t, is approximately normal with expectation $-\lambda_t\Delta t$ and variance Δt. If we add $\lambda_t\Delta t$ to the increment, we obtain approximately a normal variable with expectation 0 and variance Δt. These are the properties of a Brownian motion. The famous *theorem of Girsanov*[30] states that, in the infinitesimal limit, one indeed gets a Brownian motion in this way.

Theorem 2.7.1 (Girsanov, 1960) *Let W_t be a k-vector Brownian motion and let λ_t be a k-vector process adapted to W_t. If the process λ_t satisfies mild boundedness conditions, then the scalar process θ_t defined by (2.92) is a positive P-martingale and we may take it as a Radon-Nikodym process that defines a change of measure from the original measure P to a new measure Q. Under this new measure, the process $\widetilde{W}_t$ defined by*

$$d\widetilde{W}_t = \lambda_t\,dt + dW_t, \qquad \widetilde{W}_0 = 0 \tag{2.93}$$

is a k-vector Brownian motion.

[30] Igor Vladimirovich Girsanov (1934–1967), Russian mathematician. The actual theorem as proved by Girsanov is stated in a much more general setting than is shown here.

Recall that the term "vector Brownian motion" is used in this book to refer to standard vector Brownian motion. In other words, it is assumed in the theorem that the components of the process W_t are independent under P, and part of the conclusion is that the components of the process $\widetilde{W}_t$ are independent under Q.

The theorem implies that stochastic differential equations respond in a quite simple way to a change of measure, when this change of measure is given by a process λ_t in the way of (2.92). Suppose that a process X_t satisfies a stochastic differential equation of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \tag{2.94}$$

where W_t is a Brownian motion under P. It follows from Girsanov's theorem that, when a new measure Q is defined through (2.92) and (2.90), the process X_t also satisfies the stochastic differential equation

$$dX_t = (\mu(t, X_t) - \sigma(t, X_t)\lambda_t)dt + \sigma(t, X_t)\, d\widetilde{W}_t \tag{2.95}$$

where $\widetilde{W}_t$ is a Brownian motion under Q. Compared to (2.94), the volatility term is the same but the drift has changed; this is sometimes summarized as "change of measure is change of drift." The converse statement, "change of drift is change of measure", is true if the volatility $\sigma(t, X_t)$, which is in general a matrix of size $n \times k$, is square and invertible.

The fact that the process X_t can also be described by (2.95) is convenient for instance for computation of expressions such as $E^Q[f(X_T)]$, where f is a given function. All of the techniques discussed in this chapter can be used, starting from the SDE (2.95). Alternatively, starting from λ_t one might also solve for θ_t from (2.92), and then compute $E^Q[f(X_T)]$ as $E^P[\theta_T f(X_T)]$, but that is usually more work.

It will be seen in the coming chapters that the "Girsanov parameter" λ_t is not only convenient from a mathematical point of view, but also has an important financial meaning in cases where the change of measure is from "real-world measure" to "pricing measure". The popularity of writing financial models in continuous time, rather than in discrete time, is explained to a large extent by the fact that there is no Girsanov's theorem in the discrete-time setting.

Example 2.7.2 The simplest choice for the process λ_t is to make it constant. In this case, the SDE (2.92) becomes (for the case $k = 1$)

$$d\theta_t = -\lambda\theta_t\, dW_t, \qquad \theta_0 = 1$$

which is a special case of the stochastic differential equation for the geometric Brow-

nian motion. From the general expression (2.45), it is seen that the solution is given by

$$\theta_t = \exp(-\tfrac{1}{2}\lambda^2 t - \lambda W_t). \tag{2.96}$$

Suppose for instance that we want to compute $E^Q X_T$ for a process X_t that is given by the Ornstein-Uhlenbeck SDE

$$dX_t = -aX_t\,dt + \sigma\,dW_t, \qquad X_0 \text{ given}$$

where a and σ are given numbers and where W_t is a Brownian motion under P. First, let us try the method using the Radon-Nikodym process θ_t:

$$E^Q[X_T] = E^P[\theta_T X_T] = e^{-\frac{1}{2}\lambda^2 T} E^P[e^{-\lambda W_T} X_T].$$

We know from Section 2.6.2 that the random variable X_T follows a normal distribution with mean and variance given, under the measure P, by (writing just E instead of E^P, and var instead of var^P)

$$E[X_T] = e^{-aT} X_0, \qquad \mathrm{var}(X_T) = \frac{1 - e^{-2aT}}{2a}\,\sigma^2.$$

Introduce an auxiliary variable Y by

$$Y = X_T - \frac{\mathrm{cov}(X_T, W_T)}{\mathrm{var}(W_T)}\,W_T.$$

The two variables Y and W_T are uncorrelated, and since they are jointly normal, this means that they are independent. We have

$$E[e^{-\lambda W_T} X_T] = E\big[e^{-\lambda W_T} Y\big] + \frac{\mathrm{cov}(X_T, W_T)}{\mathrm{var}(W_T)}\,E[e^{-\lambda W_T} W_T]. \tag{2.97}$$

From the fact that Y and W_T are independent, it follows that Y and $e^{-\lambda W_T}$ are also independent, so that

$$E\big[e^{-\lambda W_T} Y\big] = E\big[e^{-\lambda W_T}\big] E[Y] = e^{\frac{1}{2}\lambda^2 T} e^{-aT} X_0.$$

Concerning the second term on the right in (2.97), note that, from the formula (2.66) for the covariance of two stochastic integrals with deterministic integrands:

$$\begin{aligned}\mathrm{cov}(X_T, W_T) &= \mathrm{cov}\Big(\int_0^T e^{-a(T-t)}\sigma\,dW_t, \int_0^T dW_t\Big) \\ &= \int_0^T e^{-a(T-t)}\sigma\,dt = \frac{1 - e^{-aT}}{a}\,\sigma.\end{aligned}$$

The expectation of $Xe^{\lambda X}$, where $X \sim N(0,1)$, can now be computed as follows:[31]

$$\begin{aligned} E[Xe^{\lambda X}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\lambda x - \frac{1}{2}x^2}\, dx \\ &= \frac{\lambda}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda x - \frac{1}{2}x^2}\, dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\lambda - x) e^{\lambda x - \frac{1}{2}x^2}\, dx \\ &= \frac{\lambda e^{\frac{1}{2}\lambda^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2}(x-\lambda)^2}\, dx - \frac{1}{\sqrt{2\pi}} e^{\lambda x - \frac{1}{2}x^2}\Big|_{-\infty}^{\infty} \\ &= \lambda e^{\frac{1}{2}\lambda^2}. \end{aligned}$$

This implies that

$$E[e^{-\lambda W_T} W_T] = -\lambda T e^{\frac{1}{2}\lambda^2 T}.$$

Putting everything together, one finds

$$\begin{aligned} E^Q[X_T] &= e^{-\frac{1}{2}\lambda^2 T}\Big[e^{\frac{1}{2}\lambda^2 T} e^{-aT} X_0 - \frac{1 - e^{-aT}}{aT}\, \sigma \cdot \lambda T e^{\frac{1}{2}\lambda^2 T}\Big] \\ &= e^{-aT} X_0 - \frac{1 - e^{-aT}}{aT}\, \sigma\lambda = e^{-aT}\Big(X_0 + \frac{\sigma\lambda}{a}\Big) - \frac{\sigma\lambda}{a}\,. \end{aligned} \tag{2.98}$$

Now, let us do the computation with the aid of Girsanov's theorem. According to the theorem, we can write

$$dX_t = -aX_t\, dt + \sigma\, dW_t = (-aX_t - \sigma\lambda)\, dt + \sigma\, d\widetilde{W}_t$$

where $\widetilde{W}_t$ is a Brownian motion under Q. The equation above can also be written as

$$d\Big(X_t + \frac{\sigma\lambda}{a}\Big) = -a\Big(X_t + \frac{\sigma\lambda}{a}\Big)\, dt + \sigma\, d\widetilde{W}_t$$

which implies in particular that

$$E^Q\Big[X_T + \frac{\sigma\lambda}{a}\Big] = e^{-aT}\Big(X_0 + \frac{\sigma\lambda}{a}\Big).$$

In other words,

$$E^Q[X_T] = e^{-aT}\Big(X_0 + \frac{\sigma\lambda}{a}\Big) - \frac{\sigma\lambda}{a}$$

just as we found from the method via the Radon-Nikodym process. It appears that, in this case, Girsanov's theorem provides a substantial shortcut with respect to the Radon-Nikodym method.

[31] An alternative calculation is $E[Xe^{\lambda X}] = \frac{d}{d\lambda} E[e^{\lambda X}] = \lambda e^{\frac{1}{2}\lambda^2}$. This requires an argument (which can be given) to motivate the interchange of expectation and differentiation.

2.8 Exercises

1. Plot two sample paths of the Wiener process on the interval $[0, 1]$, discretized on a grid of equidistant time points. Use 100 steps.

2. Let W_t be a Wiener process. Show that $\mathrm{cov}(W_{t_1}, W_{t_2}) = \min(t_1, t_2)$. For every step in your reasoning, indicate which property of the Wiener process you use.

3. Let W_t be a Wiener process, and let α be a positive constant. Define a process $\widetilde{W}_t$ by

$$\widetilde{W}_t = \alpha^{-1/2} W_{\alpha t}.$$

Show that $\widetilde{W}_t$ is a Wiener process.

4. Let W_t be a Wiener process. Show that the process X_t defined by

$$X_t = \tfrac{1}{3}W_t^3 - tW_t$$

is a martingale. Give a proof that does not use the Itô formula, and also one that does. [*Hint*: For the proof without the Itô formula, expand $((W_{t+h} - W_t))^3$ and take conditional expectations given information up to time t.]

5. Let the process X_t be defined by $X_t = tW_t$, where W_t is a Brownian motion. Determine $[X, X]_t$ for $t \geq 0$.

6. Let W_t be a Wiener process.

a. Using the Itô formula, derive a deterministic differential equation for the function $x(t) := E[\cos W_t]$. Use the result to show that

$$E[\cos aZ] = e^{-\frac{1}{2}a^2} \qquad (Z \sim N(0,1)). \tag{2.99}$$

b. Using the Itô formula, derive a deterministic differential equation that relates the function $y_k(t) := E[Z_t^{2k}]$, for $k \geq 1$, to the function $y_{k-1}(t)$. Use the result to show that

$$E\big[Z^{2k}\big] = (2k-1)E\big[Z^{2k-2}\big] \qquad (k \geq 1,\ Z \sim N(0,1)). \tag{2.100}$$

7. Let X be a standard normal variable, and let Y be the discrete variable that takes the values 1 and -1 each with probability $\frac{1}{2}$. Assume that X and Y are independent. Let Z be defined by $Z = XY$.

a. Prove that Z follows a standard normal distribution.

b. Prove that $X + Z$ does *not* follow a normal distribution.

8. **a.** Using Euler discretization, generate three approximate trajectories of the solution of the stochastic differential equation

$$dX_t \;=\; \mu X_t\,dt + \sigma X_t\,dW_t \tag{2.101}$$

on the interval $[0, 20]$, with initial condition $X_0 = 1$. Use the parameter values $\mu = 0.08$ and $\sigma = 0.2$, and set the time step Δt equal to 0.1.

b. Generate 10^4 approximate sample paths under the same conditions. Plot the average of the simulated values of X_t as a function of t. Also, plot the variance of the simulated values of X_t as a function of t. Relate the results that you get to the theory of geometric Brownian motion in Section 2.6.3.

9. Same questions as in Exc. 8 for the stochastic differential equation

$$dX_t \;=\; \mu X_t\,dt + \sigma\,dW_t \tag{2.102}$$

with parameter values $\mu = -0.2$ and $\sigma = 0.5$. Relate the results that you get to the theory of linear SDEs in Section 2.6.3.

10. Same questions as in Exc. 8 for the stochastic differential equation

$$dX_t \;=\; (\alpha X_t + \beta)\,dt + \sigma\sqrt{X_t}\,dW_t \tag{2.103}$$

with parameter values $\alpha = -2$, $\beta = 1$ and $\sigma = 2$.

11. The plots in Fig. 2.1 show sample paths of processes $\{X_t\}$ described by the following stochastic differential equations:

a. $dX_t = -0.9X_t\,dt + 0.4\,dW_t$

b. $dX_t = 0.8X_t\,dt + 1.5X_t\,dW_t$

c. $dX_t = 0.8(1 - \sqrt{|X_t|})\,dt + 0.2X_t\,dW_t$

d. $dX_t = 0.8\,dt + 0.1X_t\,dW_t$.

Find out which plot belongs probably to which process. Motivate your answer.

12. The plots in Fig. 2.2 show sample paths of processes $\{X_t\}$ described by the following stochastic differential equations:

a. $dX_t = -0.5(1 - \exp(1 - X_t))\,dt + 0.2X_t\,dW_t$

b. $dX_t = 0.5X_t\,dt + 0.1\,dW_t$

c. $dX_t = 0.4X_t\,dt + X_t\,dW_t$

d. $dX_t = 0.5\,dt + 0.1(1 - X_t)\,dW_t$.

Find out which plot belongs probably to which process. Motivate your answer.

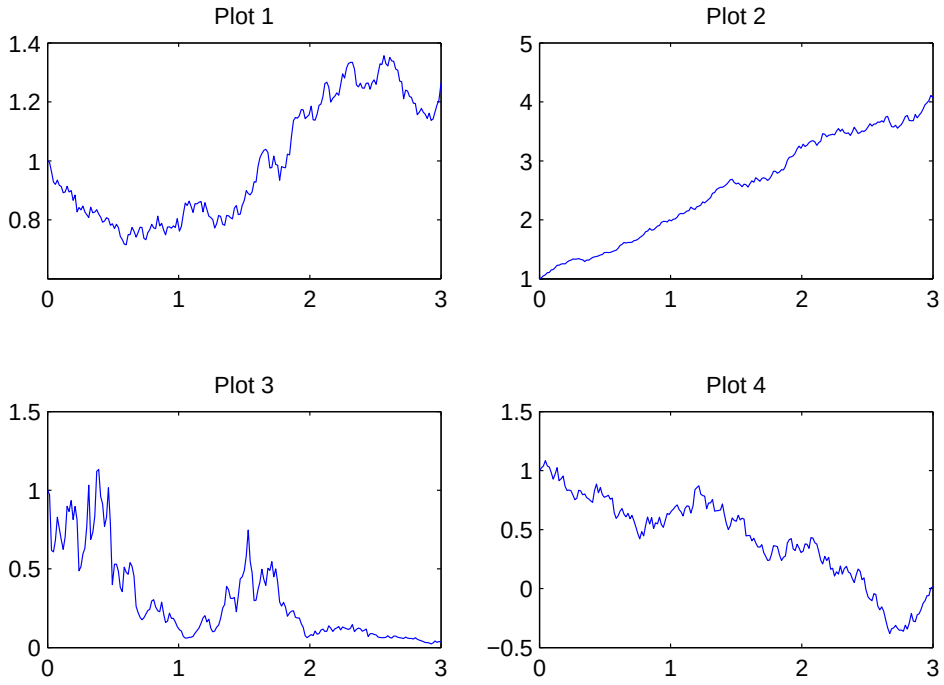

Figure 2.1: Sample paths obtained from the four SDEs in Exc. 11

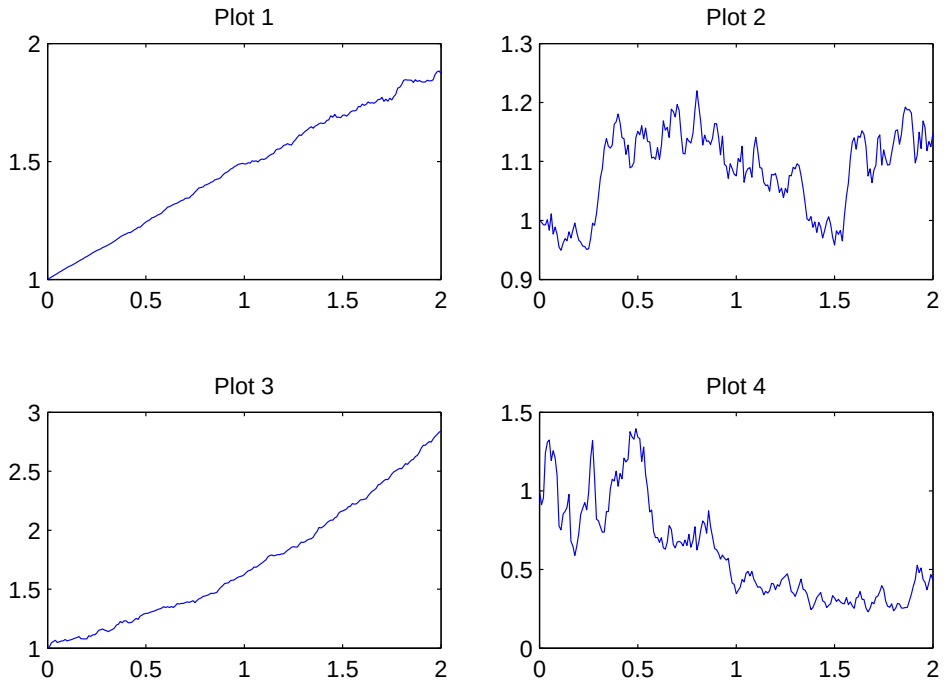

Figure 2.2: Sample paths obtained from the four SDEs in Exc. 12

13. Suppose that the stochastic processes X_t and Y_t satisfy the stochastic differential equations

$$dX_t = -\tfrac{1}{2}X_t\,dt - Y_t\,dW_t \tag{2.104a}$$

$$dY_t = -\tfrac{1}{2}Y_t\,dt + X_t\,dW_t \tag{2.104b}$$

where W_t is Brownian motion.

a. Define a stochastic process Z_t by $Z_t = X_t^2 + Y_t^2$. Compute dZ_t.

b. Show that, for any constants a and b, the stochastic processes defined by

$$X_t = a\cos(W_t + b), \qquad Y_t = a\sin(W_t + b)$$

form a solution of (2.104).

c. Apply Euler discretization to the equation system (2.104). Use the discretization to generate a trajectory of the joint variables (X_t, Y_t) on an interval of length 5 with initial conditions $X_0 = 1$ and $Y_0 = 0$, and plot the trajectory in the (X, Y)-plane. Use time step $\Delta t = 0.01$. Does the plot satisfy the conditions that are suggested by parts a. and b.? Repeat with $\Delta t = 0.001$.

14. Suppose that the vector $(W_0, W_1, W_2, \ldots, W_N)$ has been generated as a sampling of a trajectory of Brownian motion on a grid of $N+1$ equidistant points on the interval $[0, T]$. Define $\Delta t = T/N$. Construct a new vector $(\hat{W}_0, \hat{W}_1, \hat{W}_2, \ldots, \hat{W}_{2N})$ as follows:

$$\hat{W}_{2i} = W_i \text{ for } i = 0, \ldots, N$$

$$\hat{W}_{2i+1} = \tfrac{1}{2}(W_i + W_{i+1}) + V_i \text{ for } i = 0, \ldots, N$$

where the V_i's are drawn from a normal distribution with mean zero and variance $\frac{1}{4}\Delta t$ independently of each other and of the W_i's. Show that this procedure is a valid way to generate a sampling of Brownian motion on the interval $[0, T]$ with $2N$ steps. (It will be general enough if you verify the following properties for the increments $\hat{W}_1 - \hat{W}_0$, $\hat{W}_2 - \hat{W}_1$, and $\hat{W}_3 - \hat{W}_2$: (i) they are jointly normally distributed, (ii) the expected value of each increment is 0, (iii) the variance of each increment is $\frac{1}{2}\Delta t$, and (iv) the increments are pairwise independent.)

15. **a**. Define a random variable X by

$$X = \int_0^1 W_t\,dt$$

where W_t is a Wiener process. Show that X is normally distributed, and determine the expectation and variance of X. [*Hint*: compute the differential of tW_t, and

use the telescope rule to write X as the difference of two stochastic integrals with deterministic integrands.[32]]

b. Generalize to the case in which X is defined by $X = \int_0^T W_t\, dt$, with $T > 0$.

16. **a**. Generate a sample path of Brownian motion on the interval $[0, 1]$, using $N = 10^6$ steps. Extract from this sample path the corresponding path with time steps of length 0.1. Compute the sum of absolute values of increments

$$\sum_{j=0}^{9} |W_{0.1(j+1)} - W_{0.1j}| .$$

Repeat with time steps 0.01, 0.001, up to 10^{-6}. Does the sequence of numbers that you get in this way appear to converge to a finite value?

b. Repeat part a., but now compute the sum of *squares* of increments instead of the sum of absolute values of increments. Again comment on whether the sequence of numbers that you obtain in this way appears to converge to a finite value.

17. Generate two trajectories of Brownian motion on the time interval $[0, 1]$, with 100 time steps. Plot both trajectories. Also plot the corresponding trajectories of the cumulative sum of squared increments. Repeat the experiment with 10^5 time steps rather than 100. Relate the results to the property (2.33).

18. **a**. Generate trajectories of two independent Brownian motions on the time interval $[0, 1]$, with 100 time steps. Plot the corresponding trajectory of the cumulative sum of products of increments.

b. Repeat the experiment of part a., but now using 10^5 time steps rather than 100. Compare the vertical scale in the plot that you obtain to the one you found in part a. Relate the results to the property (2.56).

19. **a**. Using the definition (2.76), show that the solution of the vector differential equation $(dx/dt)(t) = Ax(t)$ with $x(0) = x_0$ is given by $x(t) = \exp(At)x_0$. (In case the matrix A is triangular, this can be used to compute $\exp(At)$ by solving a series of scalar differential equations.)

b. Let A be a square matrix and let S be an invertible matrix of the same size. Use the definition (2.76) to prove that $\exp(SAS^{-1}t) = S\exp(At)S^{-1}$, where S is any invertible matrix of the same size as A. (Remember that any matrix can be brought into triangular form by a similarity transformation. Part a. and part b. can therefore be used in conjunction to find $\exp(At)$ for any square matrix A.)

[32] This technique follows the method of integration by parts from deterministic calculus.

c. Using either the method of part a. or the definition (2.76) directly, derive that

$$\exp\left(\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} t\right) = \begin{bmatrix} e^{a_1 t} & 0 \\ 0 & e^{a_2 t} \end{bmatrix}$$

and that

$$\exp\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t\right) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

20. Generate 10^4 approximate trajectories of (2.101) on an interval of length 10, with 100 steps. Use the parameter values $\mu = 0.01$, $\sigma = 0.2$, and initial condition $X_0 = 1$. On the basis of these trajectories, generate plots of the following functions:

(i) the logarithm of the average of the values of X_t, as a function of time;

(ii) the average of the logarithms of the values of X_t, as a function of time.

Relate the results that you get to the theory of the geometric Brownian motion in Section 2.6.1.

21. **a**. If S_t is given by the geometric Brownian motion model

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$$

where μ and σ are constants, what is the quadratic variation process of $\log S_t$?

b. Generate a trajectory of geometric Brownian motion S_t using Euler approximation, with parameter values $\mu = 0.08$ and $\sigma = 0.15$, and compute the corresponding cumulative sum of squared increments of $\log S_t$. Use $T = 10$ years and take 250 time steps per year. Plot both S_t and the cumulative sum of squared increments of $\log S_t$ as functions of time. Repeat the experiment twice. Does the cumulative sum of squared increments match the quadratic variation process that you computed?

c. Take your favorite stock or stock index. Collect daily price data over a reasonably long period (10 or more years) and plot the cumulative sum of squared increments of $\log Y(t)$ where $Y(t)$ represents the series of prices. Comment on the validity of the Black-Scholes model for your data.

22. The purpose of this exercise is to take a look at the time discretization error that is introduced by the Euler approximation method. To make it possible to compute the error, a case is taken in which the exact solution is available.

a. Generate 10^5 trajectories of geometric Brownian motion (2.41) by means of the Euler approximation method with 20 steps.[33] Use $X_0 = 100$, $\mu = 0.04$, $\sigma = 0.20$,

[33] We need a sufficiently high number of trajectories to prevent that the discretization error will be swamped by the Monte Carlo error.

$T = 10$. Each of these trajectories corresponds to a particular trajectory of Brownian motion. Also compute, for each trajectory, the exact value of X_T according to formula (2.45). In this way you obtain 10^5 realizations of the variable X_T^{ap} that is produced by the Euler approximation, and 10^5 corresponding realizations of the exact random variable X_T^{ex}. On the basis of these draws, compute estimates of

(i) the relative *root mean squared error* (RMSE), i.e.

$$\frac{\sqrt{E[(X_T^{\text{ap}} - X_T^{\text{ex}})^2]}}{E[X_T^{\text{ex}}]}$$

(ii) the relative error in the second moment, i.e.

$$\frac{\left|E[(X_T^{\text{ap}})^2] - E[(X_T^{\text{ex}})^2]\right|}{E[(X_T^{\text{ex}})^2]}.$$

b. Repeat part a., but now using N steps where N is of the form $N = 2^k$ with $k = 1, \ldots, 6$. Generate a loglog plot of the relative RMSE against N, and do this also for the relative error in the second moment.

c. In numerical analysis, an approximation method is said to have *order of convergence* α if the approximation error behaves asymptotically as a constant times $N^{-\alpha}$, where N is a measure of computation time (for instance the number of time steps used, as in part b.). On the basis of the six data points gathered in part b., carry out a linear regression of log RMSE with respect to $\log N$ and a vector of ones; in other words, estimate the parameters β_1 and β_2 in the relation

$$\log \text{RMSE} = \beta_1 + \beta_2 \log N + \varepsilon.$$

Do this also for the error in the second moment. What would you say is the order of convergence of the Euler approximation method in the two cases?[34]

d. To summarize, by how much should the step size be decreased to reduce the discretization error by a factor 2 in the case of (i) RMSE, (ii) second moment?

23. Let the process X_t be given by the stochastic differential equation

$$dX_t = \mu X_t \, dt + \sigma X_t \, dW_t \tag{2.105}$$

[34] In the context of the analysis of numerical methods for approximating random variables, the order of convergence for RMSE is called the *strong* order of convergence, and the order of convergence for quantities of the form $E[f(X)]$ is called the *weak* order of convergence. We used the function $f(x) = x^2$ here, but typically one finds the same weak order of convergence for practically any function f.

where W_t is a Brownian motion under a given probability measure P. Consider the process θ_t given by

$$d\theta_t = -\lambda\theta_t\, dW_t, \qquad \theta_0 = 1 \tag{2.106}$$

which, by the solution formula for geometric Brownian motion, has the explicit solution $\theta_t = \exp(-\frac{1}{2}\lambda^2 t - \lambda W_t)$. The process θ_t is a positive martingale under P, and therefore it can be used to define a change of measure from P to a new measure Q. Suppose that $X_0 = 100$, $\mu = 0.04$, $\sigma = 0.2$, $\lambda = 0.25$, $T = 5$.

a. Indicate which of the following statements are true, and motivate your choices.

1. Under the new measure defined by the RN process θ_t, there is a shift of probability weight on the outcomes of W_T such that

   (i) high (positive) values become more likely

   (ii) low (negative) values become more likely

   (iii) neither of the two above.

2. Judging from the differential equation (2.105), the sign of the correlation between X_T and W_T is

   (i) positive

   (ii) negative

   (iii) zero.

3. On the basis of items 1. and 2., the following relation is expected to hold:

   (i) $E^Q X_T > E^P X_T$

   (ii) $E^Q X_T < E^P X_T$

   (iii) $E^Q X_T = E^P X_T$

   (iv) no conclusion can be drawn.

b. The exact value of $E^Q X_T$ can be computed by the Radon-Nikodym formula $E^Q X_T = E^P[\theta_T X_T]$, or alternatively by making use of Girsanov's theorem. Compute the value of $E^Q X_T$, for any given $T \geq 0$, by both methods. What is the value of $E^P X_T$?

c. An approximate value of $E^Q X_T$ can also be computed via simulation. Generate 1000 trajectories of the stochastic differential equation (2.105) with timestep $\Delta t = 0.1$ on the interval $[0, T]$ with $T = 5$, and compute both the mean and the standard deviation of the results to obtain an estimated 95% confidence interval for $E^Q X_T$

(i) using the Radon-Nikodym formula and the exact value of θ_T

(ii) using the Radon-Nikodym formula and the simulated value of θ_T (i.e. simulate also (2.106))

(iii) using Girsanov's theorem (i.e. simulate under Q).

d. Compare the confidence intervals obtained from the three methods above. Which one gives the best results?

e. Repeat part c. with $\lambda = 0.5$ instead of $\lambda = 0.25$. Which method now gives the smallest confidence interval?

24. **a**. Let X_t be a continuous martingale, and let α be a constant. Prove that the process Y_t defined by

$$Y_t = \exp(\alpha X_t - \tfrac{1}{2}\alpha^2 [X, X]_t)$$

is a martingale. You may assume that the boundedness conditions mentioned in connection with Thm. 2.2.1 are fulfilled. [*Hint*: compute dY_t.]

b. Let X_t be a martingale whose quadratic variation process is a deterministic function of time; say

$$[X, X]_t = g(t).$$

Prove that, for any t_1 and t_2 with $t_2 > t_1$, the distribution of the increment $X_{t_2} - X_{t_1}$, conditional on information up to time t_1, is normal with mean 0 and variance $g(t_2) - g(t_1)$. [*Hint*: use part a. together with the standard fact that a random variable follows a normal distribution with mean 0 and variance σ^2 if and only if its moment generating function $f(\alpha) = E\exp(\alpha Z)$ satisfies $f(\alpha) = e^{\frac{1}{2}\alpha^2\sigma^2}$.]

c. Prove Lévy's theorem:[35] if X_t is a continuous martingale with $[X, X]_t = t$, then X_t is a Brownian motion.

25. This exercise provides a proof of Thm. 2.7.1 in the case in which only one Brownian motion is involved. The proof for the case of a vector Brownian motion is analogous. It will be assumed that the boundedness assumptions required for the application of Thm. 2.2.1 are satisfied in all cases where needed. Consider the stochastic differential equations

$$d\widetilde{W}_t = \lambda_t\, dt + dW_t, \qquad \widetilde{W}_0 = 0$$

and

$$d\theta_t = -\lambda_t \theta_t\, dW_t, \qquad \theta_0 = 1.$$

[35] Paul Lévy (1886–1971), French mathematician. Lévy proved the theorem in the 1930s, before the Itô calculus was invented.

Under the assumptions mentioned, the process θ_t is a positive P-martingale and so it defines a change of measure from the original measure P to a new measure Q.

a. Show that $[\widetilde{W}, \widetilde{W}]_t = t$. [*Hint*: use the calculus rules for quadratic (co)variation in Subsection 2.5.1.]

b. Show that $\theta_t \widetilde{W}_t$ is a martingale with respect to P. [*Hint*: use the Itô calculus and Thm. 2.2.1.]

c. Show that $\widetilde{W}_t$ is a martingale with respect to Q. [*Hint*: use (2.90).]

d. Prove that $\widetilde{W}_t$ is a Brownian motion with respect to Q. This is the claim of Girsanov's theorem. [*Hint*: use Exc. 24.]

Chapter 3

Financial models

For the purposes of this textbook, a mathematical model for a financial market consists of a specification of the joint evolution of prices of a number of given assets. Further information that may be important in practice, such as trading restrictions, are abstracted away by the assumption of perfect liquidity. This chapter provides a generic way to set up a financial model, using Brownian motions as the source of stochasticity. It is shown that key properties such as absence of arbitrage and completeness can be verified effectively by means of linear algebra.

3.1 The generic state space model

3.1.1 Formulation of the model

We start from a model of the following form:

$$\boxed{\begin{aligned} dX_t &= \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t \\ Y_t &= \pi_Y(t, X_t). \end{aligned}} \tag{3.1}$$

This is the general continuous-time state space model based on stochastic differential equations driven by Brownian motion. The following notational conventions will be in use. The symbol W_t denotes a vector-valued Brownian motion; unless mentioned otherwise, W_t is a standard vector Brownian motion (i.e. there is no correlation between the components of W_t). The symbol X_t is reserved for the state variable, whereas the vector of variables of interest, which typically is a vector of asset prices, is denoted by Y_t. The letters k, n, and m are used to denote the dimensions of

W_t, X_t, and Y_t respectively.[1] Then $\mu_X(t,x)$ is an n-vector-valued function of $1+n$ variables, $\sigma_X(t,x)$ is an $n \times k$-matrix-valued function of $1+n$ variables, and $\pi_Y(t,x)$ is an m-vector-valued function also of $1+n$ variables.[2]

In the above model, all stochasticity derives from the driving process W_t. The state X_t is a collection of variables that have been chosen in such a way that all other relevant quantities can be expressed as functions of these variables. The vector Y_t denotes quantities of interest that depend on the state variables; in the models of financial markets that will be considered below, Y_t is always a vector of prices of tradable assets. The components of the state vector may be themselves prices of assets, but they can be other indicators as well, such as the 6-month Libor rate or the temperature in Rio.

The state process is a *Markovian* process.[3] Intuitively, this means that all information from the past that is relevant to the future is summarized in the current value of the state variable. In terms of the state, it only matters where you are, not how you got there. In more formal terms, a stochastic process $\{X_t\}$ is said to be a *Markov process* if, for any sequence of time points $t_1 < t_2 < \cdots < t_k < t_{k+1} < \cdots < t_{k+l}$, the conditional distribution of $(X_{t_{k+1}}, \ldots, X_{t_{k+l}})$ given $(X_{t_1}, \ldots, X_{t_k})$ is the same as the conditional distribution of $(X_{t_{k+1}}, \ldots, X_{t_{k+l}})$ given only X_{t_k}.

In the model (3.1), the process $\{Y_t\}$ may not be a Markovian process itself, but we can say that all information from the past of $\{Y_t\}$ that is relevant to the future of $\{Y_t\}$ is summarized in the current value of the state variable X_t. Not all stochastic processes have the property that they can be described as instantaneous functions of a finite-dimensional Markovian process (meaning that Y_t is a function of t and X_t only, X_t has finitely many components, and the process $\{X_t\}$ is Markovian). However, processes that do not have this property are in practice rarely used as models for asset prices, since they are hard to simulate.

Part of the state space modeling philosophy is to assume that cashflows that are generated by the contract at a given time T are determined completely by the value of the state variable at that point in time. In practice it is always possible to ensure that this property holds, by extending the number of state variables if necessary. A theoretical example of a situation in which an infinity of state variables would be needed is provided by the *continuously-sampled sliding-window American-Asian option*, which is a contract that entitles the holder to a payoff that depends on

[1] Indeed, n is the $n$umber of state variables, k is the number of ris$k$ factors, and m refers to the $m$enu of investment opportunities.

[2] Shorthand notation will be used at a number of occasions below, to keep formulas more compact. In this style of notation, arguments of functions and time subscripts are dropped, so that for instance we write $dX = \mu_X\, dt + \sigma_X\, dW$. This should not lead to confusion. While in this style of notation the second of line of (3.1) becomes $Y = \pi_Y$, it should be remembered that the notation Y refers to $(Y_t)_t$, which is a stochastic process, whereas π_Y is a function from $\mathbb{R}^{1+n}$ to $\mathbb{R}^m$.

[3] Andrey Andreyevich Markov (1856–1922), Russian mathematician.

the integral of the price of a given asset across the interval $[\tau - c, \tau]$, where c is a constant and τ is the time, to be chosen by the holder, at which the contract pays off. To make it possible to write the payoff as a function of the value of the state variable at the time the payoff takes place, the state variable at time t must include a representation of the full trajectory of the asset price from time $t - c$ up to time t. This information cannot be stored into finitely many variables. In such cases one might think of extending the model (3.1) in such a way that the state variable takes values in an infinite-dimensional space instead of $\mathbb{R}^n$. Infinite-dimensional state space models might also be formulated, for the purpose of generality. In this book we stick to the finite-dimensional model (3.1).

In the model (3.1), it is assumed that the driving Brownian motions are independent. In practice, however, it is often convenient to write a model of the form (3.1) in which the components of the vector W_t are Brownian motions that may be *dependent*. As discussed in Section 2.1, it is always possible to write a vector Brownian motion W_t with possibly dependent entries in the form $W_t = F\hat{W}_t$ where F is a suitable constant matrix and $\hat{W}_t$ is a standard Brownian motion. This means that a model of the form

$$dX_t = \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t \tag{3.2}$$

can be written as

$$dX_t = \mu_X(t, X_t)\, dt + \sigma_X(t, X_t) F\, d\hat{W}_t$$

which in turn can be written as

$$dX_t = \mu_X(t, X_t)\, dt + \hat{\sigma}_X(t, X_t)\, d\hat{W}_t$$

where

$$\hat{\sigma}_X := \sigma_X F. \tag{3.3}$$

In this way, any model of the form (3.1) in which the entries of the driving Wiener process W_t are dependent can be replaced by a model of the same form in which the driving process is a standard Brownian motion.

Therefore, to allow the driving process in (3.1) to be a nonstandard vector Brownian motion does not make the model more general. The rule (3.3) can be used to transform expressions based on independent Brownian motions to expressions based on dependent Brownian motions. Specifically, if one starts with a model (3.2) in which the process W_t is a vector Brownian motion with variance-covariance matrix Σ, then all expressions below (which are derived under the assumption that W_t is a standard vector Brownian motion) are still valid provided that all instances of σ_X are replaced by $\sigma_X F$, where F is any matrix such that $FF^\top = \Sigma$, and W is

accordingly replaced by $\hat{W}$. Note that $\sigma_X \sigma_X^\top$ is replaced according to this rule by $\sigma_X F F^\top \sigma_X^\top = \sigma_X \Sigma \sigma_X^\top$, which doesn't depend on the particular choice of the matrix F. Also note that the combination $\sigma_X \, dW$ is replaced by $\sigma_X F \, d\hat{W} = \sigma_X \, dW$ because $W = F\hat{W}$; so this combination actually doesn't require adaptation.

The state process X_t may be restricted both in time and in values; for instance, we may write our model only for a certain period, such as the interval $[0, T]$, and some components of X_t may be constrained to take only positive values. The letter $\mathcal{D}$ (for *domain*) indicates the collection of time-value pairs (t, x) taken into account in our model; so $\mathcal{D}$ is a subset of $\mathbb{R} \times \mathbb{R}^n$. Although in a rigorous model description the domain $\mathcal{D}$ needs to be specified explicitly, the "right" choice of $\mathcal{D}$ is usually obvious and it will often be omitted in model specifications below. The symbols μ_X and σ_X denote functions from $\mathcal{D}$ to $\mathbb{R}^n$ and to $\mathbb{R}^{n \times k}$, respectively. They should be such that the stochastic differential equation for the state variable

$$dX_t = \mu_X(t, X_t) \, dt + \sigma_X(t, X_t) \, dW_t \tag{3.4}$$

has a unique solution in the domain $\mathcal{D}$ given an initial condition X_0. The function $\pi_Y(t, x)$, which is used to express the asset price vector Y_t in terms of the current time and the current state, is defined on $\mathcal{D}$ and takes values in $\mathbb{R}^m$. The following is a standing assumption.

Assumption 3.1.1 In the model (3.1), the vector of pricing functions $\pi_Y(t, x)$ is nonzero for all (t, x) in the domain $\mathcal{D}$.

In other words, it cannot happen that all asset prices are zero simultaneously. If this assumption would not hold, the model would not admit any numéraires. It will also always be assumed that the assets Y_i are "pure" assets, net of all costs and dividends; so the assets Y_i are self-financing quantities.

From the model (3.1), one can obtain formulas for the drift and the volatility of asset prices. On the basis of Itô's lemma, it is possible to write

$$dY_t \; = \; \mu_Y(t, X_t) \, dt + \sigma_Y(t, X_t) \, dW_t \tag{3.5}$$

where the functions $\mu_Y(t, x)$ and $\sigma_Y(t, x)$ can be expressed in terms of the data in (3.1). The function μ_Y takes values in $\mathbb{R}^m$, and σ_Y is a matrix of size $m \times k$. Writing down explicit expressions in vector form for μ_Y and σ_Y is a bit unwieldy though; therefore, let us write the equations in a componentwise way. Take a component of Y_t (i.e., a single asset), and let it be denoted by C_t.[4] By the multivariate Itô rule

[4] The letter C is used here to stand for "claim" or "contract".

(2.54), the drift function and the volatility function of C can be written as[5]

$$\mu_C = \frac{\partial \pi_C}{\partial t} + \frac{\partial \pi_C}{\partial x}\,\mu_X + \tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial^2 \pi_C}{\partial x^2}\,\sigma_X \sigma_X^\top \tag{3.6a}$$

$$\sigma_C = \frac{\partial \pi_C}{\partial x}\,\sigma_X. \tag{3.6b}$$

Shorthand notation is used here, in which the arguments t and x are suppressed. The function $\mu_C(t,x)$ is of scalar type, whereas $\sigma_C(t,x)$ has size $1 \times k$. The gradient vector $\partial \pi_C / \partial x$ is defined as a row vector. The symbol $\partial^2 \pi_C / \partial x^2$ is used to denote the Hessian[6] matrix of the second partial derivatives of π_C. The trace operator is employed in (3.6a) as a device to avoid summation symbols. More explicitly, the expression for the drift can be written as

$$\begin{aligned}\mu_C &= \frac{\partial \pi_C}{\partial t} + \sum_{i=1}^{n} \frac{\partial \pi_C}{\partial x_i}(\mu_X)_i + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \pi_C}{\partial x_i \partial x_j}(\sigma_X \sigma_X^\top)_{ij} \\ &= \frac{\partial \pi_C}{\partial t} + \sum_{i=1}^{n} \frac{\partial \pi_C}{\partial x_i}(\mu_X)_i + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\ell=1}^{k} \frac{\partial^2 \pi_C}{\partial x_i \partial x_j}(\sigma_X)_{i\ell}(\sigma_X)_{j\ell}\,.\end{aligned} \tag{3.7}$$

In shorthand notation, not only arguments of functions but also subscripts t for stochastic processes are omitted, so that for instance (3.5) can be written as $dY = \mu_Y dt + \sigma_Y dW$ where it is understood that Y in dY and W in dW should have subscripts t, and that μ_Y and σ_Y should be evaluated at (t, X_t).

The generic notation X_t has been used above for the state vector, but in applications the state variables often have particular meanings and associated letters with mnemonic value are used, such as S_t for "stock value", B_t for "bond value", r_t for "interest rate", and so on. In a model which has for instance S and B as state variables, we then write $\mu_X(t,S,B)$ instead of $\mu_X(t,x)$, still keeping the subscript X to indicate that the symbol refers to the drift of the state variables. Correspondingly, instead of expressions such as $\partial \pi_C / \partial x_i$, we write $\partial \pi_C / \partial S$ and so on. In a deviation from mathematical purity that is motivated by shorthand notation (see footnote 2), even $\partial C / \partial S$ will be used instead of $\partial \pi_C / \partial S$, in line with what is often seen in the literature.

3.1.2 Portfolio strategies

In the context of a market model in state space form, a *Markovian portfolio strategy* is defined as a function that determines a vector of asset holdings at time t just on

[5] The symbol "tr" denotes the trace operator, which assigns to a square matrix the sum of its diagonal elements. The matrix on which the operator acts in (3.6a) is the product of the three matrices that follow the tr symbol. A more explicit form is shown in (3.7).

[6] Otto Hesse (1811–1874), German mathematician.

the basis of the current value of the state variable and the calendar time t. More formally, a Markovian portfolio strategy is given by a function that is defined on the domain $\mathcal{D} \subset \mathbb{R} \times \mathbb{R}^n$ and that takes values in $\mathbb{R}^m$. It would not be reasonable to allow arbitrary functions as portfolio strategies, so some regularity conditions are usually imposed. In particular, it will be required that portfolio processes have continuous paths, as already anticipated in the discussion of stochastic calculus rules in the previous chapter. The portfolio holdings process generated by a strategy function ϕ is $\phi_t = \phi(t, X_t)$.[7]

From a financial point of view, a particularly important class of portfolio strategies is formed by the *self-financing strategies* which define portfolios that satisfy the budget constraint of no funds added and no funds extracted. Given a portfolio holdings process ϕ_t, the value of the portfolio defined by this strategy at time t and in state X_t is

$$V_t = \phi_t^\top Y_t = \phi_t^\top \pi_Y(t, X_t). \tag{3.8}$$

The process $\{V_t\}$ is called the *value process* defined by ϕ. In case of a Markovian portfolio strategy, the portfolio value can also be written as an instantaneous function of the state variables: define a function π_V by

$$\pi_V(t, x) = \phi^\top(t, x)\pi_Y(t, x) \tag{3.9}$$

then

$$V_t = \pi_V(t, X_t). \tag{3.10}$$

The portfolio holdings process ϕ is self-financing if (cf. (1.8))

$$\boxed{dV_t = \phi_t^\top \, dY_t} \tag{3.11}$$

or in more detail for Markovian strategies:

$$dV_t = \phi^\top(t, X_t)\, dY_t = \phi^\top(t, X_t)\mu_Y(t, X_t)\, dt + \phi^\top(t, X_t)\sigma_Y(t, X_t)\, dW_t. \tag{3.12}$$

For a self-financing portfolio strategy $\{\phi_t\}$, we have

$$\boxed{V_T = V_0 + \int_0^T \phi_t^\top \, dY_t} \tag{3.13}$$

by the telescope rule. This achieves the goal that was formulated in Section 1.3. It

[7] Note that the symbol ϕ is "overloaded" here; it is used both to denote an m-vector-valued function of $1 + n$ variables and an m-vector-valued stochastic process. The distinction is made by the use of either subscript or brackets.

is the concept of the stochastic integral that allows us to write down this expression even though the process $\{Y_t\}$ is not assumed to be of bounded variation.

The formula (3.13) is valid for self-financing strategies. As already discussed in Section 1.2, it is possible to write a similar formula in which the self-financing constraint no longer appears, if asset prices are expressed relative to a *numéraire*. Any asset whose price is always positive may serve as a numéraire. More generally, any fixed linear combination of assets or, even more generally, any self-financing portfolio of assets may be taken as a numéraire, as long as the price of the portfolio is always positive. In the context of the state space model (3.1), a numéraire is therefore defined by any m-vector function $\nu = \nu(t,x)$[8] having th following properties:

(i) $\nu(t,x)^\top \pi_Y(t,x) > 0$ for all $(t,x) \in \mathcal{D}$;

(ii) the process N_t defined by $N_t = \nu(t,X_t)^\top Y_t$ satisfies $dN = \nu^\top dY$.

The second condition expresses that the portfolio strategy ν should be self-financing; the condition can be written in a more explicit form by means of Itô's rule. Now, consider a not necessarily self-financing trading strategy ϕ_t. Define a process V_t by

$$\frac{V_t}{N_t} = \frac{V_0}{N_0} + \int_0^t \phi_t^\top \, d\frac{Y_s}{N_s} \tag{3.14}$$

and another process ψ_t by

$$\psi_t = \frac{V_t - \phi^\top Y_t}{N_t}\,. \tag{3.15}$$

We then have

$$V_t = \phi_t^\top Y_t + \psi_t N_t = \hat{\phi}_t^\top Y_t$$

where $\hat{\phi}_t = \phi_t + \psi_t \nu_t$, and $\nu_t := \nu(t,X_t)$. Moreover, the strategy $\hat{\phi}_t$ is self-financing. To verify that claim, note that by the product rule of stochastic calculus we can write

$$dV = N\,d\frac{V}{N} + \frac{V}{N}\,dN + d[N,V/N], \qquad dY = N\,d\frac{Y}{N} + \frac{Y}{N}\,dN + d[N,Y/N].$$

Moreover, from (3.14) it follows that $d(V/N) = \phi^\top d(Y/N)$; by calculus rules (i) and (vi) in Section 2.5.1, this implies in particular that $d[N,V/N] = \phi^\top d[N,Y/N]$. From the fact that the numéraire is self-financing it follows that

$$dN = \nu^\top dY = N\nu^\top d\frac{Y}{N} + \nu^\top \frac{Y}{N}\,dN + \nu^\top d[N,Y/N] = N\nu^\top d\frac{Y}{N} + dN + \nu^\top d[N,Y/N]$$

so that

$$N\nu^\top d\frac{Y}{N} + \nu^\top d[N,Y/N] = 0.$$

[8] The symbol used here is the Greek letter ν (nu).

We can now compute

$$\begin{aligned} dV - \hat{\phi}^\top dY &= N\left(d\frac{V}{N} - \phi^\top d\frac{Y}{N} - \psi\nu^\top d\frac{Y}{N}\right) + \left(\frac{V}{N} - \hat{\phi}^\top \frac{Y}{N}\right) dN \\ &\qquad + d[N, V/N] - \phi^\top d[N, Y/N] - \psi\nu^\top d[N, Y/N] \\ &= -\psi\left(N\nu^\top d\frac{Y}{N} + \nu^\top d[N, Y/N]\right) = 0. \end{aligned}$$

Since $V_t = \hat{\phi}^\top Y_t$ and $dV_t = \hat{\phi}^\top dY_t$, it follows that indeed $\hat{\phi}_t$ is a self-financing strategy, and V_t is the corresponding portfolio value process. The equation (3.14) implies that the portfolio value at time T, relative to the value of the numéraire at time T, is given by

$$\boxed{\frac{V_T}{N_T} = \frac{V_0}{N_0} + \int_0^T \phi_t^\top \, d\frac{Y_t}{N_t}\,.} \tag{3.16}$$

This relates to (3.13) in the same way as (1.13) relates to (1.9). The formula (3.16) is used below in conjunction with Thm. 2.2.2 ("you can't beat the system") to derive a fundamental characterization of conditions under which a continuous-time market is free of arbitrage: see Thm. 3.2.1.

The numéraire has been defined above in quite general terms, as a portfolio that is constructed from the available assets in such a way that its value is always positive. In practical applications, the numéraire is often simply a component of the vector of assets. In that case, the function $\nu(t, x)$ is constant and equal to a unit vector. The trading strategy ϕ_t can then be viewed as a strategy of trading in the non-numéraire assets, and the process ψ_t adds the corresponding process of holdings in the numéraire so that the trading strategy in all assets is self-financing.

3.1.3 Examples

Undoubtedly the most popular model in mathematical finance is the Black-Scholes[9] model, which can be specified as follows:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t \tag{3.17a}$$

$$dB_t = r B_t\, dt. \tag{3.17b}$$

[9] Fischer Black (1938–1995), American economist. Myron S. Scholes (1941), Canadian/US economist, Nobel prize 1997.

Here, W_t is a Brownian motion, and the quantities μ, σ, and r are constants. The symbol S refers to "stock", and B refers to "bond".[10] The model can be solved explicitly; S_t follows a geometric Brownian motion

$$S_t = S_0 \exp\left((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t\right)$$

and B_t behaves as an exponential function:

$$B_t = e^{rt} B_0.$$

We can take S and B as state variables. At the same time S and B are also the traded assets in the model, so that the BS model is an example of a model in which the state variables are the same as the traded assets. In terms of the generic state space model (3.1), the functions that specify the model are

$$\mu_X(t,S,B) = \begin{bmatrix} \mu S \\ rB \end{bmatrix}, \quad \sigma_X(t,S,B) = \begin{bmatrix} \sigma S \\ 0 \end{bmatrix}, \quad \pi_Y(t,S,B) = \begin{bmatrix} S \\ B \end{bmatrix}. \tag{3.18}$$

Variants of the Black-Scholes model can be created by replacing any of the three constants μ, σ, and r by a stochastic process. For instance, one can replace the drift coefficient μ by a variable μ_t that follows an Ornstein-Uhlenbeck process:

$$dS_t = \mu_t S_t\, dt + \sigma S_t\, dW_{1,t} \tag{3.19a}$$
$$d\mu_t = a(b - \mu_t)\, dt + c\, dW_{2,t} \tag{3.19b}$$
$$dB_t = rB_t\, dt \tag{3.19c}$$

where a, b, and c are constants. The parameter a should be positive to ensure that μ_t is mean-reverting rather than mean-fleeing. The model (3.19) can be viewed as an example of a *business cycle model*, because the variable μ_t can be thought of as a business cycle indicator: "boom" when μ_t is larger than its long-term average b, "bust" when it is below that value. For the purposes of such an interpretation, the Brownian motions $W_{1,t}$ and $W_{2,t}$ are usually assumed to be correlated. The traded assets in the model are S and B, and as state variables we can take S, B, and μ. The model can be written in standard space form by defining

$$\mu_X(t,S,B,\mu) = \begin{bmatrix} \mu S \\ rB \\ a(b-\mu) \end{bmatrix}, \quad \sigma_X(t,S,B,\mu) = \begin{bmatrix} \sigma S & 0 \\ 0 & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix},$$

[10] The term "bond" is customary here, but, unlike what is usually the case for bonds, the asset with price B_t does not have a time of maturity, The asset is more properly described as a savings account, or as a money market account.

$$\pi_Y(t, S, B, \mu) = \begin{bmatrix} S \\ B \end{bmatrix}. \tag{3.20}$$

A criticism of the Black-Scholes model that is frequently expressed is that volatility in this model is constant, whereas in reality there are times in which markets are more nervous, and prices are more volatile. One may attempt to capture such changes in volatility by replacing the constant parameter σ in the BS model by a stochastic process $\{\sigma_t\}$. For instance, the following model is sometimes used:

$$dS_t = \mu S_t\, dt + \sqrt{\nu_t}\, S_t\, dW_{1,t} \tag{3.21a}$$

$$dB_t = rB_t\, dt \tag{3.21b}$$

$$d\nu_t = \kappa(\theta - \nu_t)\, dt + \xi\sqrt{\nu_t}\, dW_{2,t} \tag{3.21c}$$

where μ, r, κ, and ξ are constants, and the Brownian motions $W_{1,t}$ and $W_{2,t}$ are allowed to be correlated. The constant ξ is called the *volatility of volatility*. The model (3.21a–3.21c) for the evolution of stock prices is known as the *Heston model* after Steven Heston who proposed it in 1993. The standard state space representation can be written down with

$$\mu_X(t, S, B, \nu) = \begin{bmatrix} \mu S \\ rB \\ \kappa(\theta - \nu) \end{bmatrix}, \quad \sigma_X(t, S, B, \nu) = \begin{bmatrix} \sqrt{\nu}\, S & 0 \\ 0 & 0 \\ 0 & \xi\sqrt{\nu} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix},$$

$$\pi_Y(t, S, B, \nu) = \begin{bmatrix} S \\ B \end{bmatrix}. \tag{3.22}$$

The Black-Scholes model also assumes that the interest rate is constant, and in fact that the same discount rate r applies to all maturities. In practice, the three-month interest rate may well be different from the one-year rate, which again may be different from the ten-year rate; moreover, all of these rates tend to change in time. Modeling of interest rates will be discussed more extensively in Chapter 5 of this book. A simple way to introduce variable interest rates in the BS model is as follows:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_{1,t} \tag{3.23a}$$

$$dB_t = r_t B_t\, dt \tag{3.23b}$$

$$dr_t = a(b - r_t)\, dt + c\, dW_{2,t} \tag{3.23c}$$

where μ, a, b, and c are constants, and where $W_{1,t}$ and $W_{2,t}$ are Brownian motions. The two Brownian motions could be correlated. The model can be written

in standard state space form by defining

$$\mu_X(t,S,B,r) = \begin{bmatrix} \mu S \\ rB \\ a(b-r) \end{bmatrix}, \quad \sigma_X(t,S,B,r) = \begin{bmatrix} \sigma S & 0 \\ 0 & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix},$$
$$\pi_Y(t,S,B,r) = \begin{bmatrix} S \\ B \end{bmatrix}. \tag{3.24}$$

Given the equation (3.23b), the variable r_t should be interpreted as the interest that is paid on short-maturity loans (the *short rate*). The model (3.23c) for the evolution of the short rate is called the *Vasicek model*.[11] The model (3.23) as a whole is sometimes referred to as the *Black-Scholes-Vasicek model*.

Many more models for financial markets have been developed. For instance, such models may include prices of several stocks driven by different (but correlated) Brownian motions, or they may attempt to describe the behavior of prices of commodities such as oil or gold. There are also the *term structure models* which concentrate on the important financial products that depend on the evolution of interest rates. Most of the models used in practice are driven by Brownian motion, although sometimes other sources of stochasticity (such as for instance jump processes) are used as well. Typically, models can be written in the form (3.1). Starting from the state space model (3.1) we therefore reach a level of generality that is sufficient for most practical purposes. The model is specified in terms of the three vector/matrix functions μ_X, σ_X, and π_Y, and so it must be possible to discuss all notions associated to financial models, in particular absence of arbitrage and completeness, in terms of these three functions. This is what will be done in the sections below.

3.2 Absence of arbitrage

3.2.1 The fundamental theorem of asset pricing

An *arbitrage* is a self-financing trading strategy which, starting with zero initial portfolio value, creates a portfolio value at a later time that is nonnegative with probability 1 and positive with positive probability.[12] In other words, an arbitrage strategy is one that starts with nothing, never makes a loss, and realizes a gain at

[11] Oldrich A. Vasicek (1942), Czech/US mathematician. Vasicek proposed the model that is named after him in 1977.

[12] The term "arbitration" is used in historical texts in English to describe the process of choosing among various ways to settle a payment in a foreign currency, either directly or via a third currency. The equivalent French term "arbitrage" has come in use to refer to the possibility of making a riskless profit that arises when the difference between the pathways is large enough, and the more expensive one can be reversed.

least in some cases. Models of financial markets are usually constructed in such a way that they do not allow arbitrage, at least by strategies that are "admissible" in the sense that they satisfy certain reasonable conditions. For instance, a condition that can be imposed is that there is *some* level of loss beyond which trading strategies should not be allowed to continue; this requirement excludes versions of the "doubling strategy".[13] A model in which there are no arbitrage strategies that are admissible is said to be *free of arbitrage*. The *fundamental theorem of asset pricing* gives conditions under which a given model is arbitrage-free. The theorem can be stated as follows.

Theorem 3.2.1 *The market as specified by an objective ("real-world") probability measure $\mathbb{P}$ and a collection of asset price processes $\{Y_i\}_t$ $(i = 1, \ldots, m)$ is free of arbitrage if and only if, given any numéraire N, there is a measure $\mathbb{Q}_N$ (depending on N) which is equivalent[14] to the objective measure $\mathbb{P}$, and which is such that all relative price processes $(Y_i)_t/N_t$ are $\mathbb{Q}_N$-martingales.*

Any measure $\mathbb{Q}_N$ that has the properties mentioned in the theorem is called an *equivalent martingale measure*; "equivalent" because $\mathbb{Q}_N$ is equivalent to $\mathbb{P}$, and "martingale" because relative asset prices are martingales under $\mathbb{Q}_N$.

To see that the condition stated in the theorem indeed precludes the construction of admissible arbitrage strategies, note that, as a consequence of the relation (3.16) and Thm. 2.2.1, the assumption that all relative price processes $(Y_i)_t/N_t$ are $\mathbb{Q}_N$-martingales implies that also V_t/N_t is a $\mathbb{Q}_N$-martingale, where V_t is the value of any self-financing portfolio that can be constructed from the basic assets $(Y_i)_t$. Let V_t denote the value of such a portfolio, with $V_0 = 0$, and let N_t denote the value of a chosen numéraire. The strategy that gives rise to the portfolio value V_t would constitute an arbitrage if at some time T the conditions

$$\mathbb{P}(V_T \geq 0) = 1, \qquad \mathbb{P}(V_T > 0) > 0$$

would hold. Because the measure $\mathbb{Q}_N$ is equivalent to $\mathbb{P}$, and because the value of the numéraire N_T is certainly positive, these conditions imply

$$\mathbb{Q}_N(V_T/N_T \geq 0) = 1, \qquad \mathbb{Q}_N(V_T/N_T > 0) > 0.$$

[13] In a coin tossing game, the doubling strategy is to wager one euro and to double the stakes after each lost round; playing is stopped as soon as a win occurs. In a game with finitely many rounds, this strategy produces a large probability of a small gain, and a small probability of a large loss. In a continuous-time setting, one can play versions of this game with arbitrarily many rounds (assuming that trading can be done arbitrarily fast), even on a finite time interval. An arbitrage opportunity would arise if there would be no limit on the amount that can be bet.

[14] Recall that two probability measures P and Q are said to be *equivalent* if any event that has positive P-probability also has positive Q-probability, and vice versa. The equivalence in the theorem holds on intervals of arbitrary but finite length.

However, this state of affairs is excluded by the martingale property which implies, by Thm. 2.2.2, that

$$E^{\mathbb{Q}_N} \frac{V_T}{N_T} = \frac{V_0}{N_0} = 0$$

since $V_0 = 0$.

The converse statement, that an admissible arbitrage can be constructed when the condition of the theorem is not met, is more difficult to show. A sketch of an argument can be given as follows. Choose a numéraire. For a fixed time T, which may be taken as large as desired, consider the set of all random variables that can be constructed as final portfolio values, relative to chosen numéraire, of admissible portfolio strategies that start from portfolio value 0. Let this set be denoted by $\mathcal{H}$. Since strategies can be added and multiplied by constants, $\mathcal{H}$ is in fact a linear subspace. By the assumption of absence of arbitrage, $\mathcal{H}$ intersects the set of nonnegative random variables only in the point 0. Geometric intuition then suggests that it is possible to find a linear functional (i.e. a mapping of the space of random variables to the real line) such that its nullspace contains $\mathcal{H}$, and the set of nonzero nonnegative random variables is mapped to the positive part of the real line. The latter condition ensures that this linear functional, after proper scaling, can be written as expectation under a measure $\mathbb{Q}$ that is equivalent to the real-world measure $\mathbb{P}$. One then has the property

$$E^{\mathbb{Q}} \int_0^T \phi_t^\top \, d\frac{Y_t}{N_t} = 0$$

for all admissible strategies. By a converse of Thm. 2.2.2, it follows that the processes Y_t/N_t are martingales under $\mathbb{Q}$.[15] As emphasized, this is only a sketch of a proof; in particular, one needs to be careful when using geometric intuition in infinite-dimensional spaces. A version of the argument in the very simplest case is incorporated in Exc. 3.7.1.

The importance of the fundamental theorem of asset pricing is first of all that it gives a criterion for absence of arbitrage, which is a basic property for models of financial markets. But the theorem can also be used for pricing purposes. Suppose that, in an arbitrage-free market, an asset is introduced which is defined in terms of the existing assets; for instance it might be an option written on one of the assets in the model. Then the price of the new asset is said to be *market-consistent* if the market with this asset included is still free of arbitrage.[16] Denote the price of the new asset at time t by C_t. By the fundamental theorem of asset pricing, the price is market-consistent if and only if, for at least one of the equivalent martingale

[15] Here it is essential that ϕ_t is not subject to the self-financing constraint.

[16] The terms "fair price" and "arbitrage-free price" are also used instead of "market-consistent price".

measures $\mathbb{Q}_N$ that the given model allows, we have

$$\boxed{\frac{C_t}{N_t} = E_t^{\mathbb{Q}_N} \frac{C_T}{N_T}} \tag{3.25}$$

for any t and T with $t < T$. When the EMM is determined uniquely (this is the case of a "complete market"), then the above formula determines the price at time t uniquely in terms of the value that the asset may have at time T. For instance, in the case of a European option, the time T could be the time of maturity of the option, so that the value of the derivative at that time is given as a function of the value of the underlying asset.

The formula (3.25) is sometimes referred to as the *risk-neutral pricing formula*. To emphasize the role of the numéraire, the term *numéraire dependent pricing formula* (NDPF) will be used in this book.

3.2.2 Constructing arbitrage-free models

The formulation of the fundamental theorem of asset pricing in Thm. 3.2.1 is quite general and does not relate to a specific representation of the asset price processes Y_t. On the other hand, the theorem does not provide a concrete procedure to verify, for any given model of a financial market, whether the model is free of arbitrage. More tangible criteria for absence of arbitrage can be developed if we assume that the asset price processes are given in terms of a standard state space model (3.1). The following theorem presents a criterion for absence of arbitrage that can be verified by direct computation.

Theorem 3.2.2 *Let N be a numéraire for the model (3.1). Define $\mu_{Y/N} = \mu_{Y/N}(t,x)$ and $\sigma_{Y/N} = \sigma_{Y/N}(t,x)$ by*

$$d(Y/N) = \mu_{Y/N}\, dt + \sigma_{Y/N}\, dW. \tag{3.26}$$

The model (3.1) allows no arbitrage if and only if there exists a k-vector function $\lambda_N = \lambda_N(t,x)$ such that

$$\boxed{\mu_{Y/N} = \sigma_{Y/N}\lambda_N.} \tag{3.27}$$

Proof The condition for absence of arbitrage is that the discounted asset price process Y_t/N_t should be a martingale with respect to a measure that is equivalent to the original measure under which the model (3.1) has been formulated (the "real-world" measure). By Girsanov's theorem, in a "Markovian" version, this comes down to

the existence of a function λ_N such that when $\widetilde{W}_t$ is defned by the prescription

$$d\widetilde{W}_t = \lambda_N(t, X_t)\, dt + dW_t \tag{3.28}$$

together with the initial condition $\widetilde{W}_0 = 0$, then the discounted price process satisfies

$$d(Y/N) = \sigma_{Y/N}\, d\widetilde{W}. \tag{3.29}$$

From (3.28) and (3.26), we have

$$d(Y/N) = (\mu_{Y/N} - \sigma_{Y/N}\lambda_N)\, dt + \sigma_{Y/N}\, d\widetilde{W}$$

and so the condition to be satisfied by the function λ_N is given by (3.27). □

Note that $\mu_{Y/N}(t,x)$ is an m-vector, whereas $\sigma_{Y/N}(t,x)$ is an $m \times k$-matrix. The condition (3.27) may be expressed by the requirement that, for every $(t,x) \in \mathcal{D}$, the vector $\mu_{Y/N}(t,x)$ should belong to the subspace of $\mathbb{R}^m$ that is generated by the columns of the matrix $\sigma_{Y/N}(t,x)$. This can be verified by the techniques of linear algebra.

Suppose that (3.29) does not hold. Then, for some $(t_0, x_0) \in \mathcal{D}$, the vector $\mu_{Y/N}(t_0,x_0)$ is not in the span of the columns of $\sigma_{Y/N}(t_0,x_0)$. It follows that there exists a vector ϕ_0 such that $\phi_0^\top \sigma_{Y/N} = 0$ while $\phi_0^\top \mu_{Y/N} \neq 0$; for instance one can choose ϕ_0 such that $\phi_0^\top \mu_{Y/N} = 1$. The portfolio with composition ϕ_0 has at time t_0 and in state x_0 a positive drift and zero volatility. While that fact as such cannot be viewed as an arbitrage since the property of zero volatility holds only for an instant, some intuitive support for the condition (3.26) does emerge from it.

As an example, let us consider the standard Black-Scholes model, with the bond as a numéraire. Computation using Itô's formula shows that

$$\mu_{Y/B} = \begin{bmatrix} (\mu - r)S/B \\ 0 \end{bmatrix}, \quad \sigma_{Y/B} = \begin{bmatrix} \sigma S/B \\ 0 \end{bmatrix}.$$

The equation (3.27) in this case becomes

$$\begin{bmatrix} (\mu - r)S/B \\ 0 \end{bmatrix} = \sigma_{Y/B}\lambda_B = \begin{bmatrix} \sigma S/B \\ 0 \end{bmatrix} \lambda_B$$

where λ_B is allowed be a function of t, S, and B. If $\sigma \neq 0$, there is a unique solution which is in fact constant:

$$\lambda_B = \frac{\mu - r}{\sigma}\,. \tag{3.30}$$

If $\sigma = 0$ and $\mu \neq r$, then there is no solution and so in this case the market allows arbitrage. It is not difficult to see how to construct the arbitrage: when $\sigma = 0$, the

stock S is really a bond and when $\mu \neq r$ it carries a different interest rate than the bond B. Therefore a riskless profit can be made, starting from an initial portfolio value of zero, by borrowing money at the lower rate and lending it at the higher rate. If $\sigma = 0$ and $\mu = r$, there are many solutions, and so in this case there is no arbitrage.

It can be a bit inconvenient that the criterion (3.27) requires computation of the drift and volatility of the vector asset price process relative to the numéraire. Another criterion, which is stated directly in terms of the drift and the volatility of the asset price process itself, is given in the theorem below.

Theorem 3.2.3 *The model (3.1) admits no arbitrage if and only if there exist a k-vector valued function $\lambda = \lambda(t, x)$ and a scalar function $r = r(t, x)$ such that*

$$\boxed{\mu_Y - r\pi_Y = \sigma_Y \lambda.} \tag{3.31}$$

In order to derive this criterion from (3.27), we need a general formula for the drift and volatility of a quotient of two stochastic process in terms of the drifts and volatilities of the processes themselves (the "stochastic quotient rule"). Such a formula is given in the following lemma, which is a straightforward application of the Itô rule.

Lemma 3.2.4 *Let $\{Y_t\}$ be a vector-valued stochastic process that satisfies the stochastic differential equation*

$$dY = \mu_Y \, dt + \sigma_Y \, dW$$

and let $\{N_t\}$ be a scalar-valued process that is such that $N_t > 0$ for all t, and that satisfies

$$dN_t = \mu_N \, dt + \sigma_N \, dW. \tag{3.32}$$

Here, W_t is a standard vector Brownian motion, and μ_Y, σ_Y, μ_N, and σ_N are processes adapted to W. We then have $d(Y/N) = \mu_{Y/N} \, dt + \sigma_{Y/N} \, dW$ with

$$\mu_{Y/N} = \frac{1}{\pi_N^2}(\mu_Y \pi_N - \pi_Y \mu_N) - \sigma_{Y/N} \frac{\sigma_N^\top}{\pi_N} \tag{3.33}$$

$$\sigma_{Y/N} = \frac{1}{\pi_N^2}(\sigma_Y \pi_N - \pi_Y \sigma_N) \tag{3.34}$$

Proof From Itô's formula, we have

$$d(Y/N) = \frac{1}{\pi_N} \, dY - \frac{\pi_Y}{\pi_N^2} \, dN - \frac{1}{\pi_N^2} \, d[Y, N] + \frac{\pi_Y}{\pi_N^3} \, d[N, N]$$

$$= \frac{1}{\pi_N}(\mu_Y\,dt + \sigma_Y\,dW) - \frac{\pi_Y}{\pi_N^2}(\mu_N\,dt + \sigma_N\,dW) - \frac{1}{\pi_N^2}\sigma_Y\sigma_N^\top\,dt + \frac{\pi_Y}{\pi_N^3}\sigma_N\sigma_N^\top\,dt.$$

The formulas (3.34) and (3.33) follow from this by collecting the dt and dW terms. □

Another lemma that will be needed is formulated below. The lemma states that, if the asset prices in a particular model satisfy the equation (3.31), then the value of any self-financing portfolio formed from these assets also satisfies the same equation.

Lemma 3.2.5 *Consider the model (3.1), and suppose that the equation (3.31) holds for some functions $r(t,x)$ and $\lambda(t,x)$. Let the m-vector function $\phi(t,x)$ indicate a self-financing strategy in the assets Y, and let the associated value process be denoted by $V_t := \phi(t,X_t)^\top Y_t$. Define the functions $\mu_V = \mu_V(t,x)$ and $\sigma_V = \sigma_V(t,x)$ by*

$$dV = \mu_V dt + \sigma_V dW \tag{3.35}$$

and define the function $\pi_V = \pi_V(t,x)$ by $\pi_V = \phi^\top \pi_Y$. Under these conditions, the following relation holds:

$$\mu_V = r\pi_V + \sigma_V\lambda. \tag{3.36}$$

Proof The requirement that the strategy ϕ should be self-financing means that $dV = \phi^\top dY$. It then follows from (3.35) that $\mu_V = \phi^\top\mu_Y$ and $\sigma_V = \phi^\top\sigma_Y$. The relation (3.36) now follows upon premultiplying both sides of the equation (3.31) by $\phi^\top$. □

We can now proceed to the proof of the theorem.

Proof (of Thm. 3.2.3). Let $N_t = \nu(t,X_t)^\top Y_t$ be a numéraire; in particular, this means that the strategy $\nu(t,X_t)$ is self-financing, so that $dN = \nu^\top dY$. Define $\mu_N(t,x)$ and $\sigma_N(t,x)$ as in (3.32). We know from Thm. 3.27 that absence of arbitrage holds if and only if for each $(t,x) \in \mathcal{D}$, the m-vector $\mu_{Y/N}(t,x)$ belongs to the span of the columns of the $m \times k$-matrix $\sigma_{Y/N}(t,x)$. From the expressions in (3.34) and (3.33), we see that this happens if and only if[17]

$$\mu_Y\pi_N - \pi_Y\mu_N \in \operatorname{colsp}(\sigma_Y\pi_N - \pi_Y\sigma_N).$$

This is equivalent to the existence of a k-vector function $\lambda = \lambda(t,x)$ such that

$$\mu_Y\pi_N - \pi_Y\mu_N = (\sigma_Y\pi_N - \pi_Y\sigma_N)\lambda. \tag{3.37}$$

[17]The notation "colsp A" is used to denote the linear subspace spanned by the columns of the matrix A.

The above equation can be rewritten as

$$\mu_Y = \frac{\mu_N - \sigma_N \lambda}{\pi_N} \pi_Y + \sigma_Y \lambda. \tag{3.38}$$

Clearly, (3.31) is satisfied when (3.38) holds, that is, when there is no arbitrage. Suppose conversely that (3.31) is given; we then have to show that (3.38) holds. From Lemma 3.2.5, we have $\mu_N = r\pi_N + \sigma_N \lambda$ and so

$$r = \frac{\mu_N - \sigma_N \lambda}{\pi_N}\,.$$

It follows that (3.38) is satisfied.[18] □

The derivation as given here shows that the short rate can be defined in any arbitrage-free financial model. Indeed the short rate is a fundamental theoretical notion; it may be described as the instantaneous depreciation rate of capital, or more briefly as the "price of time". One can define an associated process called the *money market account* by the stochastic differential equation

$$dM_t = rM_t\, dt \tag{3.39}$$

together with an initial condition, for instance $M_0 = 1$. This process clearly satisfies the no-arbitrage equation $\mu_M = r\pi_M + \sigma_M \lambda$, since $\sigma_M = 0$ and $\mu_M = r\pi_M$. Given a positive initial condition, the money market account is always positive. It is often used as a numéraire.

As an example, we can apply the criterion (3.31) to the standard Black-Scholes model. A notational problem arises here, because the letter r is used both in the formulation of the BS model and in Thm. 3.2.3. Let us temporarily replace the notation $r(t,x)$ in Thm. 3.2.3 by $\rho(t,x)$. The criterion (3.31) then calls for an investigation of the solvability of the equation

$$\begin{bmatrix} \mu S \\ rB \end{bmatrix} - \begin{bmatrix} S \\ B \end{bmatrix} \rho = \begin{bmatrix} \sigma S \\ 0 \end{bmatrix} \lambda.$$

As it should be, the conditions for absence of arbitrage derived from this equation are the same as the conditions already derived above on the basis of the criterion (3.27). In particular, when $\sigma \neq 0$, we obtain unique solutions given by $\rho = r$ and

$$\lambda = \frac{\mu - r}{\sigma}\,. \tag{3.40}$$

[18] Note the role of the information that the numéraire is the value of a self-financing portfolio, i.e. it is an asset price process, rather than just any positive process.

This agrees with the expression for λ_B in (3.30), as should be the case since the bond has no volatility.

The quantity λ is often called the "price of risk"; the expression above gives its value in the BS model, as a function of the parameters of the model. This terminology has to be handled with care however; it may easily be misunderstood. A discussion is given in Section 3.6 below.

3.2.3 An alternative formulation

The no-arbitrage condition (3.31) can be rewritten in a mathematically equivalent form which allows an interpretation of its own. The rewriting is based on a fact from linear algebra, namely that the requirement that a given vector x should be in the column span of a given matrix A can be expressed not only by the condition that $x = Ay$ for some y, but also by the condition that $z^\top x = 0$ for all vectors z such that $z^\top A = 0$.

Theorem 3.2.6 *The model (3.1) admits no arbitrage if and only if there exists a scalar-valued function $r(t,x)$ such that for all $(t,x) \in \mathcal{D}$ and for all $z \in \mathbb{R}^m$ the following holds:*

$$\text{if } z^\top \sigma_Y(t,x) = 0, \text{ then } z^\top \mu_Y(t,x) \;=\; r(t,x) z^\top \pi_Y(t,x). \tag{3.41}$$

Proof If the model (3.1) is arbitrage-free, then by Thm. 3.2.3 there exist functions $\lambda(t,x)$ and $r(t,x)$ such that (3.31) holds. It follows immediately that the statement in the theorem above is true. Conversely, let $r(t,x)$ be as in the theorem above. Suppose that for a certain t and x the vector $\mu_Y(t,x) - r(t,x)\pi_Y(t,x)$ would not be in the column space of the matrix $\sigma_Y(t,x)$; then there would be an m-vector z such that $z^\top(\mu_Y(t,x) - r(t,x)\pi_Y(t,x)) = 1$ while $z^\top \sigma_Y(t,x) = 0$. This would contradict the condition of the theorem. Therefore we can conclude that for all (t,x) there must be a k-vector $\lambda(t,x)$ such that $\mu_Y(t,x) - r(t,x)\pi_Y(t,x) = \sigma_Y(t,x)\lambda(t,x)$. Absence of arbitrage follows by Thm. 3.2.3. □

The statement in Thm. 3.2.6 may be expressed as: "under all market conditions, any instantaneously riskless combination of assets earns the same instantaneous return." In particular, in a model in which a constant interest rate r is assumed, the theorem above states that arbitrage is excluded if and only if any instantaneously riskless combination of assets earns the riskless interest rate r. [19] The assets that are involved in such a combination could be the basic assets that appear in the model formulation, but they can also include derivatives written on these assets.

[19] In particular, this justifies the "alternative derivation" in the 1973 paper by Black and Scholes (see Section 1.1).

3.3 Completeness and replication

3.3.1 Completeness

Under the assumption that absence of arbitrage holds, we say that a model is *complete* if the equivalent martingale measure under which discounted price processes are martingales is determined uniquely for any given numéraire. In other words, the model (3.1) is complete if the function λ is determined uniquely by the equation (3.31). In a complete market, prices are determined uniquely. Moreover, as shown below, in a complete market it is possible to construct portfolio strategies that will produce any payoff that can be defined at a future time T as a function of variables whose value will only be known at that time, given only that the initial budget is at the appropriate level. The term "completeness" actually derives from this latter property.

The uniqueness condition leads to the following test for completeness. Recall that a matrix A is said to be *of full column rank* if the rank of A is equal to the number of columns of A (in other words, if the columns of A are independent vectors). If the linear equation $y = Ax$ has a solution, then this solution is unique if and only if the matrix A has full column rank.

Theorem 3.3.1 *Assume that the model (3.1) is arbitrage-free. The model is complete if and only if the matrix* $[\sigma_Y(t,x) \;\; \pi_Y(t,x)]$*, of size* $m \times (k+1)$*, has full column rank for all* $(t,x) \in \mathcal{D}$*.*

Proof If the condition of the theorem holds, then both r and λ in (3.31) are uniquely determined and so the model is complete. Suppose now that the condition of the theorem does not hold; then for some (t,x) either $\sigma_Y(t,x)$ has dependent columns, or there exists a k-vector η such that $\pi_Y(t,x) = \sigma_Y(t,x)\eta$. In the first case, $\lambda(t,x)$ is clearly not uniquely determined. In the second case, if (λ, r) is a solution to the equation $\mu_Y(t,x) = r\pi_Y(t,x) + \sigma_Y(t,x)\lambda$, then so is $(\lambda + c\eta, r - c)$ for any constant c; consequently, also in this case the vector λ is not uniquely determined. (Note that $\eta \neq 0$, because otherwise $\pi_Y(t,x) = 0$, which would violate Assumption 3.1.1.) In both cases it follows (via Girsanov's theorem) that the equivalent martingale measure is not unique and hence that the model is not complete. □

Since a matrix can only have full column rank if the number of rows is at least as large as the number of columns, it follows that a necessary condition for completeness of the model (3.1) is that $m \geq k + 1$, that is to say, the number of assets in the model should be at least one larger than the number of sources of uncertainty.

3.3.2 Option pricing

In a complete and arbitrage-free market, there is by definition for any given numéraire exactly one equivalent martingale measure. As a consequence, prices of contracts that can be defined within the given market are determined uniquely by the numéraire-dependent pricing formula (3.25). The formula will provide the same answer regardless which numéraire is chosen; consequently, in any specific pricing problem one can choose a numéraire that is suitable for the problem at hand, much in the same way as a suitable coordinate system may be chosen, say, to solve a specific problem in physics. Examples of the use of different numéraires are shown in Chapter 4 and Chapter 5 below.

The numéraire-dependent pricing formula (3.25) is a direct consequence of the fundamental theorem of asset pricing and holds for any contract and for any times t and T such that $T \geq t$. In applications to option pricing, the time T is typically chosen to be the time of expiration of the option. The payoff of the option may be supposed to be given as a function of the state variable at time T, where if necessary the state variables can be extended (cf. Exc. 4.5.5 for an example of state variable extension). In the generic state space model, the price of numéraire asset at time t is given as well as a function of the state variables at time t; say $N_t = \pi_N(t, X_t)$. For an option that expires at time T with value $C_T = F(X_T)$, the NDPF then takes the following specific form:

$$C_t = \pi_N(t, X_t)\, E_t^{\mathbb{Q}_N}\left[\frac{F(X_T)}{\pi_N(T, X_T)}\right]. \tag{3.42}$$

When the distribution of X_T under the measure $\mathbb{Q}_N$ is known explicitly, the expectation in the formula above can be computed as an integral which in some cases may be worked out analytically. Alternatively, computation of the option price as a function of time and of the state variables can be based on the property that the equation (3.31) must remain satisfied when C_t is included as an additional entry in the vector Y_t of asset prices. This implies that the function $\pi_C(t, x)$ satisfies the equation

$$\mu_C - r\pi_C = \sigma_C \lambda \tag{3.43a}$$

together with the boundary condition

$$\pi_C(T, x) = F(x). \tag{3.43b}$$

For given r and λ, the equation (3.43a) can be written out in full as a partial differential equation for π_C (see (3.6)). The resulting PDE is a generalized form of the *Black-Scholes equation*. Numerical methods for option pricing can be based on (3.43) as well as on (3.42), as discussed in Chapter 6 and Chapter 7 respectively.

The standard example of option pricing is the valuation of a call option within the standard Black-Scholes model. It has already been noted above (Section 3.2.2) that the BS model (with nonzero volatility) is free of arbitrage and that the price of risk is determined uniquely within the model, so that completeness holds as well. The standard numéraire with the BS model is the bond, whose price B_t at time t in the BS model is simply given by $B_t = B_0 e^{rt}$. The payoff of a call option is given by[20]

$$F(S_T) = \max(S_T - K, 0). \tag{3.44}$$

To apply the pricing formula (3.42), it suffices to know the distribution of S_T under the equivalent martingale measure $\mathbb{Q}_B$ that corresponds to taking the bond as the numéraire. An application of Itô's formula shows that

$$d\,\frac{S}{B} = (\mu - r)\,\frac{S}{B}\,dt + \sigma\,\frac{S}{B}\,dW. \tag{3.45}$$

Girsanov's theorem implies that changing to the equivalent martingale measure associated to the bond numéraire will modify the drift term in the above equation but not the volatility term. Moreover, since the process S_t/B_t must be a martingale under this measure, the drift term should vanish. In other words,

$$d\,\frac{S_t}{B_t} = \sigma\,\frac{S_t}{B_t}\,d\widetilde{W}_t \tag{3.46}$$

where $\widetilde{W}_t$ is a Brownian motion under $\mathbb{Q}_B$ (cf. 3.29). This shows that the process S_t/B_t is a geometric Brownian motion under $\mathbb{Q}_B$, just like it is under the real-world measure, be it with different parameters. From the explicit solution formula for the GBM (2.45) it follows that

$$\frac{S_T}{B_T} = \frac{S_0}{B_0}\exp(-\tfrac{1}{2}\sigma^2 T + \sigma\widetilde{W}_T). \tag{3.47}$$

Since $\widetilde{W}$ is a Brownian motion under $\mathbb{Q}_B$, the distribution of $\widetilde{W}_T$ under $\mathbb{Q}$ is normal with expectation 0 and variance T. Therefore the distribution of S_T under $\mathbb{Q}$ can be described by

$$S_T = e^{rT} S_0 \exp\bigl(-\tfrac{1}{2}\sigma^2 T + \sigma\sqrt{T}\,Z\bigr), \quad Z \sim N(0,1). \tag{3.48}$$

On the basis of (3.42), this leads to the following expression for the price of the

[20] A call option is a contract that gives the holder the right, but not the obligation, to purchase one unit of the underlying asset at the time of maturity T, by paying the strike price K which is already fixed at the initiation of the contract (time 0). If the price at time t of one unit of the underlying asset is denoted by S_t, then the value of the option contract at time T is therefore $\max(S_T - K, 0)$. This is called the "payoff".

option at time t:[21]

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \max\bigl(e^{rT} S_t \exp(-\tfrac{1}{2}\sigma^2 T + \sigma\sqrt{T}\, z) - K,\, 0\bigr) \exp(-\tfrac{1}{2} z^2)\, dz. \quad (3.49)$$

The integral can be worked out explicitly; cf. Section 4.4.2 below for details. The result is as follows:

$$C_0 = S_0 \Phi(d_1) - e^{-rT} K \Phi(d_2) \quad (3.50a)$$

where Φ is the cumulative normal distribution function and the numbers d_1 and d_2 are given by

$$d_1 = \frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \qquad d_2 = \frac{\log(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \quad (3.50b)$$

This is the celebrated *Black-Scholes formula* for the price of a call option. A collection of alternative derivations of the same formula is contained in Section 4.4.2. Since the parameter T represents time to maturity in the derivation above,[22] the price of the option at a general time $t < T$ is given by the same formula, with 0 replaced by t and T by $T - t$. The pricing function $\pi_C(t, S)$ which gives the option price as a function of the time t and the state variable S can therefore be written as

$$\pi_C(t, S) = S\Phi(d_1(t, S)) - e^{-r(T-t)} K \Phi(d_2(t, S)) \quad (3.51a)$$

where the functions $d_1(t, S)$ and $d_2(t, S)$ are given by

$$d_1(t, S) = \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2(t, S) = d_1(t, S) - \sigma\sqrt{T - t}. \quad (3.51b)$$

In the case of a call option in the Black-Scholes model, the partial differential equation (3.43) becomes (writing just π instead of π_C)

$$\frac{\partial \pi}{\partial t} + rS\frac{\partial \pi}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \pi}{\partial S^2} - r\pi = 0, \quad \pi(T, S) = \max(K - S, 0). \quad (3.52)$$

It can be verified (Exc. 3.7.3) that the function $\pi_C(t, S)$ defined in (3.51) is indeed a solution of (3.52).

3.3.3 Replication

As already mentioned, an important property of complete markets is that they allow *replication*. This is formulated in the next theorem. In the proof, again use is made

[21] Recall that if X is a random variable with density ϕ_X on an interval (a, b), then the expectation of a function of X is given by $E[f(X)] = \int_a^b f(x)\phi_X(x)\, dx$, assuming the integral exists.

[22] When current time is 0, the time *to* maturity is the same as the time *of* maturity, but in general the two are different.

of a fact from linear algebra: if a matrix A has full column rank, then the equation $y = x^\top A$, where y is given, always has a solution x.

Theorem 3.3.2 *Suppose that the model (3.1) is arbitrage-free and complete, and let r and λ be the solutions of (3.31). Let $\pi_C = \pi_C(t,x)$ be a real-valued function that satisfies the equation $\mu_C = r\pi_C + \sigma_C\lambda$, where μ_C and σ_C are defined as in (3.6). Then there exists a self-financing strategy $\phi = \phi(t,x)$ such that $\pi_C = \phi^\top \pi_Y$.*

Proof Because the matrix $[\sigma_Y(t,x) \;\; \pi_Y(t,x)]$ has full column rank for all $(t,x) \in \mathcal{D}$, there exists a vector-valued function $\phi(t,x)$ such that

$$\boxed{[\sigma_C(t,x) \;\; \pi_C(t,x)] \;=\; \phi^\top(t,x)[\sigma_Y(t,x) \;\; \pi_Y(t,x)].} \tag{3.53}$$

From the assumption that $\mu_C = r\pi_C + \sigma_C\lambda$, it follows that

$$\mu_C = [\sigma_C \;\; \pi_C] \begin{bmatrix} \lambda \\ r \end{bmatrix} = \phi^\top [\sigma_Y \;\; \pi_Y] \begin{bmatrix} \lambda \\ r \end{bmatrix} = \phi^\top \mu_Y. \tag{3.54}$$

Let $V_t = \phi_t^\top Y_t = \phi^\top(t,X_t)\pi_Y(t,X_t) = \pi_C(t,X_t)$ denote the value of the portfolio that is formed from the assets in Y by using the portfolio weights ϕ. Then

$$\begin{aligned} dV \;&=\; \mu_C\,dt + \sigma_C\,dW \;= \\ &=\; \phi^\top(\mu_Y\,dt + \sigma_Y\,dW) \;=\; \phi^\top dY \end{aligned}$$

which shows that the portfolio with weights ϕ is self-financing. □

Lemma 3.2.5 states that, under the assumption of absence of arbitrage, values of self-financing portfolios are solutions of the Black-Scholes equation. The theorem above provides a converse of this statement, under an additional assumption: in an arbitrage-free *and complete* market, any solution of the Black-Scholes equation is the value of a self-financing portfolio.

To arrive at the replication of a given payoff at some future time T, one can proceed in the following way. Suppose that C_T is the given payoff; in a state space model, we can write $C_T = f(X_T)$ where, if necessary, the state variables in the vector X_T have been extended to make it possible to represent the desired payoff in this form. Choose any numéraire $N_t = \pi_N(t,X_t)$. Define a function $\pi_C(t,x)$ by

$$\pi_C(t,x) = \pi_N(t,x)E^{\mathbb{Q}_N}\left[\frac{f(X_T)}{\pi_N(T,X_T)}\,\middle|\, X_t = x\right] \tag{3.55}$$

and note that $\pi_C(T,X_T) = f(X_T)$. The function that is defined in this way satisfies the BS equation (3.31). This is seen as follows.

Define the process C_t by $C_t = \pi_C(t, X_t)$. It follows from the definition (3.55) and the tower law of conditional expectations that the discounted process C_t/N_t is a $\mathbb{Q}$-martingale. Consequently,

$$\mu_{C/N} - \sigma_{C/N}\lambda_N = 0 \tag{3.56}$$

where $\lambda_N = \lambda_N(t, x)$ is such that the process $\widetilde{W}$ defined by (3.28) is a Brownian motion under $\mathbb{Q}$. From (3.56) we have

$$\mu_{C/N} = \sigma_{C/N}\Big(\lambda - \frac{\sigma_N^\top}{\pi_N}\Big).$$

Using the stochastic quotient rule (Lemma 3.2.4), we find from this

$$\mu_C\pi_N - \pi_C\mu_N = (\sigma_C\pi_N - \pi_C\sigma_N)\lambda$$

which can be reformulated as

$$\mu_C = \frac{\mu_N - \sigma_N\lambda}{\pi_N}\,\pi_C + \sigma_C\lambda. \tag{3.57}$$

Because by definition the numéraire portfolio is self-financing, we have $\mu_N = r\pi_N + \sigma_N\lambda$ by Lemma 3.2.5 and so the factor multiplying π_C in (3.57) is just r. This completes the argument.

Therefore, given that we start with initial value $C_0 = \pi_C(0, x_0) = N_0 E^{\mathbb{Q}_N}[C_T/N_T]$, it is possible by Thm. 3.53 to construct a portfolio strategy such that for all $t \in [0, T]$ the value of the portfolio at time t is equal to $\pi_C(t, X_t)$, and in particular the portfolio value at time T is equal to $f(X_T)$. The formula (3.53) gives the replication strategy as a solution of a matrix-vector equation which can be solved in a straightforward manner. The formula may therefore be referred to as a *replication recipe*. Some further details are given in Section 3.3.4 below.

It was shown above that solutions to the Black-Scholes equation (3.31), with given terminal condition $\pi_C(T, x) = F(x)$, can be computed as conditional expectations. The equation (3.31), as an equation for the pricing function π_C with given r and λ, can be rewritten as a partial differential equation; see (4.2) below. The fact that solutions of certain partial differential equations can be obtained from conditional expectations is the theme of the famous *Feynman-Kac theorem*.[23]

[23] Richard P. Feynman (1918–1988), American physicist; Nobel prize 1965. Mark Kac (1914–1984), Polish/American mathematician.

3.3.4 Hedging

Theorem 3.3.2 can be used to construct *hedging* strategies for liabilities that depend on future values of state variables. Assume that we have a complete and arbitrage-free market described in state space form (3.1). Suppose that current time is t and that at time $T > t$ a payment will have to be made of size $F(X_T)$, where $F : \mathbb{R} \to \mathbb{R}$ is a given function and X_T is the value of the state variable at time T. Define a function $\pi_C(t, x)$ as in (3.55). This function satisfies the Black-Scholes equation and so according to Thm. 3.3.2 there exists a self-financing portfolio that is formed from the basic assets and that has the property that at time T, whatever the realization of the state variable at that time may be, its value is exactly equal to $F(X_T)$. This confirms that the value of the contract at time t and given state x is equal to $\pi_C(t, x)$. The strategy ϕ is called a hedging strategy because it provides a perfect hedge against the liability $F(X_T)$ arising at time T; that is to say, the strategy eliminates all uncertainty in connection with this contract.

The general form of the hedging strategy is given by (3.53). The matrix $[\sigma_Y \ \pi_Y]$ must have at least as many rows as it has columns for the market to be complete. If the matrix is in fact square and invertible, then the hedging strategy is uniquely determined. In the non-square case there is some indeterminacy due to the presence of more than enough basic assets. Such redundancy may make it easier to take into account implementation considerations such as transaction costs.

A special case of the general model (3.1) is the following: $k = n-1$ (the number of sources of uncertainty is one less than the number of state variables), $\pi_Y(t, x) = x$ (the state variables are asset prices), and

$$\sigma_X(t, x) = \begin{bmatrix} \sigma_S(t, x) \\ 0 \end{bmatrix} \tag{3.58}$$

where $\sigma_S(t, x)$ is square and invertible, so that the last state variable has a zero volatility coefficient. Models of this particular form are often used in equity markets; the first $n-1$ state variables are called "stocks", the last one is a "bond". For instance, the standard Black-Scholes model is of this form. In models of this type, there is a simple expression for the hedging strategy given by (3.53). Indeed, the general expression $\sigma_C = (\partial \pi_C / \partial x)\sigma_X$ becomes, due to (3.58),

$$\sigma_C = \frac{\partial \pi_C}{\partial x_S} \sigma_S.$$

If we write $\phi^\top = [\phi_S^\top \ \ \phi_B]$, where ϕ_S has $n-1$ components and gives the positions

to be taken in stocks, then (3.53) becomes

$$\left[\begin{array}{cc} \dfrac{\partial \pi_C}{\partial x_S}\,\sigma_S & \pi_C \end{array}\right] = [\phi_S^\top \;\; \phi_B] \begin{bmatrix} \sigma_S & x_S \\ 0 & x_B \end{bmatrix} \tag{3.59}$$

In particular we have

$$\frac{\partial \pi_C}{\partial x_S}\,\sigma_S = \phi_S^\top \sigma_S$$

and hence, because σ_S is invertible,

$$\boxed{\phi_S^\top(t,x) = \frac{\partial \pi_C}{\partial x_S}(t,x).} \tag{3.60}$$

Given this, the component ϕ_B which gives the position to be taken in bonds is determined by the requirement that the portfolio should be self-financing. The strategy (3.60) is called the *delta hedge* because the partial derivatives of the contract value π_C with respect to the values of the underlying assets are known as "deltas".

In the special case of the standard Black-Scholes model, we can use a state space representation with two state variables (called S_t and B_t) and one source of randomness. For this model, the replication recipe (3.53) becomes

$$\left[\begin{array}{cc} \dfrac{\partial \pi_C}{\partial S}\,\sigma & \pi_C \end{array}\right] = [\phi_S \;\; \phi_B] \begin{bmatrix} \sigma & S \\ 0 & B \end{bmatrix}. \tag{3.61}$$

It follows that, within the BS model, replication is achieved by

$$\boxed{\phi_S(t,S) = \frac{\partial \pi_C}{\partial S}(t,S).} \tag{3.62}$$

As expected, the hedging strategy takes the form of a delta hedge. The composition of the hedge portfolio can be determined at any time by first determining the amount of stocks that should be held according to the formula above, and subsequently adjusting the amount of bonds in such a way that the portfolio is self-financing (i.e. if stocks are bought then bonds are sold, and vice versa).

In a sense there is a paradox in arbitrage-based pricing theory in perfect markets. An exact price for a contract can only be determined if the contract is replicable; but that means that the contract is in fact redundant, since under the perfect-market assumptions any market participant would be able to reproduce the same payoffs by following a suitable hedging strategy. Hence it could be said that the only contracts that can be priced are the ones for which there is no need. Following this reasoning, the enormous growth of derivatives trading since the 1970's can only be explained by

the existence of market imperfections. For instance, transactions costs are nonzero and are not the same for all market participants. However, this leads to a new paradox: while the existence of market imperfections motivates options trading, at the same time it invalidates the assumptions of frictionless markets on which most of option pricing theory is based. The resolution of this paradox is one of degree: it appears that the imperfections are large enough to motivate the existence of an options market, and at the same time small enough so that the prices that are derived under perfect-market assumptions are still useful as guidelines.

3.4 American options

Whereas a European option can only be exercised at the time of maturity, American options may be exercised at any time before or at the maturity date. The value of an American option is defined as the *minimal* premium that is needed to set up a self-financing portfolio that will *at least* replicate the payoff at any time the option may be exercised. If exercise takes place suboptimally, then the hedging portfolio will have a positive surplus. Due to the early-exercise feature, the Black-Scholes partial differential equation for the option value is replaced by a system of inequalities. This system can be derived as follows.

A hedging strategy for an American option should be such that the value of the hedging portfolio is sufficient to cover the payout obligations arising from the option, whatever the exercise strategy of the holder may be. This also means that some funds may be taken out of the portfolio in case the holder follows a non-optimal strategy. Thinking for a moment in terms of small time increments Δt, instead of the usual condition for a self-financing portfolio

$$\phi_{t+\Delta t}^{\top} Y_{t+\Delta t} = \phi_t^{\top} Y_{t+\Delta t}$$

we require now only the inequality

$$\phi_{t+\Delta t}^{\top} Y_{t+\Delta t} \leq \phi_t^{\top} Y_{t+\Delta t}. \tag{3.63}$$

Subtracting $\phi_t^{\top} Y_t$ from both sides and returning to infinitesimals, we arrive at the condition

$$dV_t \leq \phi_t^{\top} dY_t \tag{3.64}$$

for a portfolio $V_t = \phi_t^{\top} Y_t$ to be *at least self-financing*. The following theorem concerning at least self-financing strategies is a direct analog of Thm. 3.3.2.

Theorem 3.4.1 *Suppose that the model (3.1) is arbitrage-free and complete, and let r and λ be the solutions of (3.31). Let $\pi_V = \pi_V(t,x)$ be a piecewise twice*

continuously differentiable function[24] *that satisfies the inequality* $\mu_V \leq r\pi_V + \sigma_V\lambda$, *where* μ_V *and* σ_V *are defined as in (3.6). Then there exists an at least self-financing strategy* $\phi = \phi(t,x)$ *such that* $\pi_V = \phi^\top \pi_Y$.

Proof Because the matrix $[\sigma_Y(t,x) \;\; \pi_Y(t,x)]$ has full column rank for all $(t,x) \in \mathcal{D}$, there exists a vector-valued function $\phi(t,x)$ such that

$$[\sigma_V(t,x) \;\; \pi_V(t,x)] \;=\; \phi^\top(t,x)[\sigma_Y(t,x) \;\; \pi_Y(t,x)]. \tag{3.65}$$

From the assumption that $\mu_V \leq r\pi_V + \sigma_V\lambda$, it follows that

$$\mu_V \leq [\sigma_V \;\; \pi_V] \begin{bmatrix}\lambda \\ r\end{bmatrix} = \phi^\top [\sigma_Y \;\; \pi_Y] \begin{bmatrix}\lambda \\ r\end{bmatrix} = \phi^\top \mu_Y. \tag{3.66}$$

Let V_t denote the value of the portfolio that is formed from the assets in Y by using the portfolio weights ϕ. Then

$$\begin{aligned} dV \;&=\; \mu_V dt + \sigma_V dW \\ &\leq\; \phi^\top \mu_Y dt + \phi^\top \sigma_Y dW \;=\; \phi^\top dY \end{aligned}$$

which shows that the portfolio with weights ϕ is at least self-financing. □

In this proof, the fact is used that the value of the integral of a nonnegative function with respect to time is always nonnegative. Note that an analogous property does not hold for stochastic integrals. One might say that "dt is positive".

The above theorem shows that portfolios that are at least self-financing appear as solutions of the *Black-Scholes inequality*

$$\mu_V \leq r\pi_V + \sigma_V\lambda. \tag{3.67}$$

This inequality simply expresses the fact that taking money out of a portfolio decreases its growth rate.

Consider now the general case of an American option in the framework of the model (3.1). In this Markovian setting, the optimal exercise strategy of the holder of the option can be represented as a mapping which assigns to each time-state pair (t,x) a decision "exercise" or "don't exercise". The strategy can therefore be represented by the *exercise region*, that is, the set of all pairs (t,x) where the decision is to exercise. Inside the exercise region, the value of the option is clearly equal to the payoff function, say $F(t,x)$. Outside the exercise region, the value of the American option is equal to the value of a European contract that pays $F(t,x)$ at the exercise boundary. So, in the non-exercise region, the American option value must satisfy

[24] The regularity assumption ensures that the constructed portfolio strategy has continuous paths, which in turn is needed for the applicability of the stochastic calculus rules of Chapter 2.

the Black-Scholes *equation*. Also, we know that in the non-exercise region the option value cannot be less than the value of immediate exercise (otherwise the choice of the exercise region is surely not optimal). Therefore, the option value π_C satisfies the following conditions for all (t, x): either

$$\mu_C(t,x) - r(t,x)\pi_C(t,x) \leq \sigma_C(t,x)\lambda(t,x) \quad \text{and} \quad \pi_C(t,x) = F(t,x) \tag{3.68a}$$

(exercise) or

$$\mu_C(t,x) - r(t,x)\pi_C(t,x) = \sigma_C(t,x)\lambda(t,x) \quad \text{and} \quad \pi_C(t,x) \geq F(t,x) \tag{3.68b}$$

(non-exercise). Such a system of two inequalities in which always at least one has to be satisfied with equality is called a *variational inequality*. It can be shown that, under mild conditions, the variational inequality (3.68) has a unique solution that is continuously differentiable in x (the "smooth pasting condition"—see Section 4.4.1). The variational inequality can be solved explicitly in cases where a reduction to a one-dimensional problem is possible, such as in Section 4.4.1, but in general one has to take recourse to numerical methods in order to compute prices of American options. Such numerical methods are discussed in Chapter 6 and Chapter 7.

3.5 Pricing measures and numéraires

3.5.1 Change of numéraire

From the equations (3.34), (3.33), and (3.37) it is seen that the function λ that appears in (3.31) and the function λ_N that appears in (3.27) are related by

$$\lambda = \lambda_N + \frac{\sigma_N^\top}{\pi_N}\,. \tag{3.69}$$

In the particular case of the money market account, we have $\sigma_M = 0$ and so $\lambda = \lambda_M$. Consequently, we can write (cf. (3.28))

$$\boxed{dW^M = \lambda\, dt + dW^{\mathbb{P}}} \tag{3.70}$$

where $W^{\mathbb{P}}$ is a Brownian motion under the real-world measure, W^M is a Brownian motion under the risk-neutral measure, and λ is the market price of risk. Recall that λ is a vector whose length is equal to the number of entries in the vector Brownian motion $W^{\mathbb{P}}$, and that λ in general may depend on the state variables as well as on calendar time. The equation (3.69) also shows that switching from the money

market account to another numéraire can be carried out by

$$\boxed{dW^M = dW^N + \frac{\sigma_N^\top}{\pi_N}\,dt} \tag{3.71}$$

where W_t^M is a Brownian motion under $\mathbb{Q}_M$ and W_t^N is a Brownian motion under the new numéraire N. To shift from any given numéraire to any other given numéraire, one can use the formula

$$\boxed{dW^{N_1} = dW^{N_2} + \left(\frac{\sigma_{N_2}^\top}{\pi_{N_2}} - \frac{\sigma_{N_1}^\top}{\pi_{N_1}}\right) dt} \tag{3.72}$$

which follows from (3.71). In terms of the drift parameter which appears in the state equation, the change-of-numéraire formula implies

$$\boxed{\mu_X^N = \mu_X^M + \sigma_X \frac{\sigma_N^\top}{\pi_N}\,.} \tag{3.73}$$

An analogous formula holds when X is replaced by an asset price process in the same state space model. Volatility parameters are not affected by the change of measure.

The formula (3.71) holds under the usual convention that the Brownian motions involved are standard vector Brownian motions, that is to say, their components are independent. Sometimes it is convenient to write a model in terms of vector Brownian motions whose components are dependent processes. Say, for instance, that W is a vector Brownian motion with variance-covariance matrix Σ,[25] in other words: the increments $W_{t+\Delta t} - W_t$ are normally distributed with expectation 0 and variance-covariance matrix $\Delta t\,\Sigma$. The matrix Σ is nonnegative definite and therefore it can be written in the form $\Sigma = FF^\top$ where F is a (not necessarily square) matrix of loading factors. In terms of the loading factors, we can write the Brownian motion W in terms of a standard vector Brownian motion $\overline{W}$, namely as $W = F\overline{W}$. The stochastic differential equation $dX = \mu_X\,dt + \sigma_X\,dW$ can then be rewritten as $dX = \mu_X\,dt + \overline{\sigma}_X\,d\overline{W}$, where $\overline{\sigma}_X := \sigma_X F$ is a matrix of exposures to the standard Brownian motion $\overline{W}$. When this technique is applied in connection with a change of numéraire, the result is as follows:

$$dX_t = \mu_X^M\,dt + \sigma_X\,dW^M$$

[25] The use of the symbol Σ both to refer to a variance-covariance matrix and as a summation symbol should not lead to confusion.

$$
\begin{aligned}
&= \mu_X^M \, dt + \overline{\sigma}_X \, d\overline{W}^M \\
&= \mu_X^M \, dt + \overline{\sigma}_X \big(d\overline{W}^N + \frac{\overline{\sigma}_N^\top}{\pi_N} \, dt \big) \\
&= \Big(\mu_X^M + \sigma_X F \, \frac{F^\top \sigma_N^\top}{\pi_N} \Big) \, dt + \sigma_X F \, d\overline{W}^N \\
&= \Big(\mu_X^M + \sigma_X \Sigma \, \frac{\sigma_N^\top}{\pi_N} \Big) \, dt + \sigma_X \, dW^N
\end{aligned}
$$

where W^N is defined by $W^N = F\overline{W}^N$. The process W^N is a vector Brownian motion under $\mathbb{Q}_N$, with variance-covariance matrix Σ. In summary, the effect of the change of numéraire is therefore that there is no change in the volatility nor in the variance-covariance structure of the driving Brownian motion, and that the new drift is given by

$$\mu_X^N = \mu_X^M + \frac{\sigma_X \Sigma \sigma_N^\top}{\pi_N} \, . \tag{3.74}$$

The variable X might be replaced here by other variables that are formulated in the same model, such as asset price processes.

Given that the value of the numéraire N_t is given in terms of the state variables X_t by $N_t = \pi_N(t, X_t)$, we have by the Itô rule

$$\sigma_N(t,x) = \frac{\partial \pi_N}{\partial x} \, (t,x) \, \sigma_X .$$

The partial derivative on the right hand side with respect to the vector x is to be understood as the row vector with components $(\partial \pi_N / \partial x_i)(t,x)$. The function π_N can only take positive values, since N is a numéraire so that it must be positive under all circumstances. We can therefore introduce the function $\log \pi_N$. The vector σ_N / π_N which appears in the change-of-numéraire formula (3.71) can be computed as

$$\frac{\sigma_N}{\pi_N} \, (t,x) = \frac{\partial \log \pi_N}{\partial x} \, (t,x) \, \sigma_X \tag{3.75}$$

where again the partial derivative on the right hand side represents a row vector consisting of partial derivatives with respect to the components of the state variable. The formula (3.73) can consequently be written as

$$\mu_X^N = \mu_X^M + \sigma_X \sigma_X^\top \Big(\frac{\partial \log \pi_N}{\partial x} \, (t,x) \Big)^\top . \tag{3.76}$$

3.5.2 Conditions for absence of arbitrage

The no-arbitrage condition (3.27) has been derived above from the fundamental theorem of asset pricing by means of the Girsanov theorem. The FTAP states that there must be a measure $\mathbb{Q}_N$ (depending on the chosen numéraire) that is equivalent to the original ("real-world") measure $\mathbb{P}$ and that is such that all relative price

processes are martingales with respect to $\mathbb{Q}_N$. Thanks to the Girsanov theorem, the change of measure from $\mathbb{P}$ to $\mathbb{Q}_N$ can be represented by a process $\lambda_{N,t}$, which in the Markovian context of the state space model can be obtained as $\lambda_{N,t} = \lambda_N(t, X_t)$ where $\lambda_N(t, x)$ is a suitable function of time and of the state variables. The measure $\mathbb{Q}_N$ is connected to the process $\lambda_{N,t}$ by the statement that the process W_t^N defined by $W_0^N = 0$ and

$$dW_t^N = dW_t + \lambda_{N,t}\, dt$$

is a Brownian motion under $\mathbb{Q}_N$. This statement in fact fully specifies the measure $\mathbb{Q}_N$. In terms of the process W_t^N, the stochastic differential equation for the state variable in the generic model (3.1) can be written as

$$\begin{aligned} dX_t &= \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, (dW_t^N - \lambda_N(t, X_t)\, dt) \\ &= \big(\mu_X(t, X_t) - \sigma_X(t, X_t)\lambda_N(t, X_t)\big)\, dt + \sigma_X(t, X_t)\, dW_t^N \\ &= \mu_X^N(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t^N \end{aligned}$$

where $\mu_X^N(t, X_t)$ is defined by

$$\mu_X^N(t, X_t) = \mu_X(t, X_t) - \sigma_X(t, X_t)\lambda_N(t, X_t).$$

A model of the form (3.1) is said to be formulated "under $\mathbb{Q}_N$" if it is written in the form

$$\begin{aligned} dX_t &= \mu_X^N(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t^N \\ Y_t &= \pi_Y(t, X_t) \end{aligned} \tag{3.77}$$

where W_t^N is a Brownian motion under $\mathbb{Q}_N$. The relative price process Y_t/N_t may likewise be written as the solution of an SDE driven by the process W_t^N:

$$d(Y_t/N_t) = \mu_{Y/N}^N(t, X_t)\, dt + \sigma_{Y/N}(t, X_t)\, dW_t^N$$

where

$$\mu_{Y/N}^N(t, X_t) = \mu_{Y/N}(t, X_t) - \sigma_{Y/N}(t, X_t)\lambda_N(t, X_t).$$

The condition (3.31) becomes

$$\mu_{Y/N}^N(t, X_t) = 0. \tag{3.78}$$

The process Y_t satisfies the equation $dY_t = \mu_Y^N\, dt + \sigma_Y dW_t^N$ with $\mu_Y^N = \mu_Y - \sigma_Y\lambda_N$. In terms of μ_Y^N, the condition (3.31) can be written as

$$\mu_Y^N + \sigma_Y\lambda_N - r\pi_Y = \sigma_Y\lambda$$

which in view of (3.69) can also be written in the form

$$\boxed{\mu_Y^N = r\pi_Y + \sigma_Y \frac{\sigma_N^\top}{\pi_N}\,.} \tag{3.79}$$

In particular, if the money market account M_t is used as a numéraire, then we have the relationship

$$\boxed{\mu_Y^M = r\pi_Y} \tag{3.80}$$

since $\sigma_M = 0$ by definition of the money market account. In other words, in an arbitrage-free model driven by processes that are Brownian motions under the equivalent martingale measure $\mathbb{Q}_M$ that corresponds to taking the money market account as a numéraire, the relative drifts of the assets, which represent the expected returns of these assets under the measure $\mathbb{Q}_M$, are all equal to the short rate. The measure $\mathbb{Q}_M$ is sometimes called the *risk-neutral measure*. In longhand notation, the evolution of asset prices under the risk-neutral measure is given by

$$dY_t = r_t Y_t\,dt + \sigma_Y(t, X_t)\,dW_t^M$$

where W_t^M is a Brownian motion under $\mathbb{Q}_M$.

For instance, the standard Black-Scholes model is written in the following way under the measure $\mathbb{Q}_B$ that corresponds to taking the bond as a numéraire:

$$dS = rS\,dt + \sigma S\,dW \tag{3.81a}$$

$$dB = rB\,dt \tag{3.81b}$$

where W is a Brownian motion under the measure $\mathbb{Q}_B$.[26] The differential equation for B_t is deterministic, and so it keeps the same form under any change of measure. In the Black-Scholes model, the price of the stock is always positive and we can therefore take it as a numéraire. Under the corresponding measure $\mathbb{Q}_S$, the BS model is written, according to (3.79), as

$$dS = (r + \sigma^2)S\,dt + \sigma S\,dW \tag{3.82a}$$

$$dB = rB\,dt \tag{3.82b}$$

where now W stands for a Brownian motion under $\mathbb{Q}_S$.

For purposes of pricing, it is enough to have the model (3.77) together with a specification of the numéraire on which it is based. The drift terms of those state

[26] We drop the explicit reference to the numéraire in the notation, as is common in the literature.

variables that are actually prices of traded assets can easily be established on the basis of (3.79). The drift terms under $\mathbb{Q}$ of other state variables can be inferred from a *calibration* exercise; this means that some parametrization is assumed for the drift terms, and the parameters are tuned by optimizing the match of prices produced by the model to actually observed market prices of selected (liquidly traded) products. This is a general technique which can be used to determine various parameters in pricing models.

For purposes of replication, again it is enough to have the model (3.77). This is because the "replication recipe" (3.53) depends only on quantities that are invariant under change of numéraire. Indeed, the state volatility matrix σ_X is not affected by a change of numéraire, and asset volatilities depend only on the pricing function and the state volatility matrix, as shown by (3.6b). Therefore the input quantities in (3.53) can be taken directly from a given model under an arbitrary equivalent martingale measure $\mathbb{Q}$.

A disadvantage of model formulation under $\mathbb{Q}$ is that the price of risk is not determined, so that a potential warning signal is lost which could arise if, for whatever reason (for instance because the parametrized model that is used does not closely match reality) the results of the calibration process are not in line with economic intuition. The shortcuts made possible by the relation (3.79) however make the formulation under a martingale measure attractive. Moreover, it is often possible to simplify a problem formulation by a clever choice of the numéraire, as seen in various examples in Chapter 4. The price of risk, which connects models under $\mathbb{Q}$ to models under the real-world measure $\mathbb{P}$, is discussed further in Section 3.6.

3.5.3 The pricing kernel

Assume that we have an arbitrage-free model of the form (3.1), and that r and λ satisfy (3.31). Introduce a process K_t by $K_0 = 1$ and

$$dK = -K(r\,dt + \lambda^\top dW). \tag{3.83}$$

This process is called the *pricing kernel*. It has the following property: a function $\pi_C = \pi_C(t,x)$ satisfies the no-arbitrage equation (3.43a) if and only if the process C_t defined by $C_t = \pi_C(t, X_t)$ is such that the product process $K_t C_t$ is a martingale with respect to the real-world measure $\mathbb{P}$. To see this, note that

$$\begin{aligned} d(KC) &= K\,dC + C\,dK + d[K,C] = \\ &= K(\mu_C\,dt + \sigma_C\,dW) + C(-K(r dt + \lambda^\top dW)) - K\sigma_C\lambda\,dt = \\ &= K[(\mu_C - r\pi_C - \sigma_C\lambda)dt + (\sigma_C - \pi_C\lambda^\top)dW]. \end{aligned}$$

The drift term vanishes if and only if (3.43a) is satisfied.

Suppose now that we want to price a European option that matures at time T and that has a payoff function $F(x)$. From the above, it follows that the time-0 price of this derivative that is consistent with the assumed riskless rate of return r and market price of risk λ is given by

$$\pi_C(0,x) = E^{\mathbb{P}}[K_T F(X_T) \mid X_0 = x].$$

More generally, this equation may be written for a general time point $t < T$:

$$\boxed{\pi_C(t,x) = E^{\mathbb{P}}[K_T F(X_T) \mid X_t = x,\ K_t = 1].} \tag{3.84}$$

Here, the processes X_t and K_t are generated jointly by (3.4) and (3.83). It is not possible in general to express K_t as a function of X_t; the pricing kernel should rather be viewed as an additional state variable.

An advantage of the pricing kernel method is that it involves only expectations under the real-world measure $\mathbb{P}$, so that any difficulties of interpretation associated to changes of numéraire are avoided. This advantage would be called a disadvantage by those who prefer to have the freedom of choosing a suitable numéraire for a given pricing problem; see for instance the examples in Chapter 4. Actually it may even be argued that the pricing kernel method is just a particular case of the change-of-numéraire method. Indeed, it can be shown that the inverse of the pricing kernel process constitutes a self-financing portfolio whose price is always positive, and whose associated equivalent martingale measure coincides with the real-world measure.

The relationship between the pricing kernel and a given numéraire can be described on the basis of the equality

$$E^{\mathbb{Q}_N}[C_T/N_T] = E^{\mathbb{P}}[K_T C_T]$$

which must hold for any payoff C_T, given the normalization $N_0 = 1$. This relationship shows that

$$K_t = \theta_t^N / N_t \tag{3.85}$$

where θ_t^N is the Radon-Nikodym process that describes $\mathbb{Q}_N$ in terms of $\mathbb{P}$. In other words, the pricing kernel includes both a change of measure (multiply by θ_t^N) and discounting (divide by N_t). The definition (3.83) shows this as well.

3.5.4 Calibration

Practitioners often arrive at models for particular financial markets by starting from a parametrized model class specified under $\mathbb{Q}$, where a numéraire may be chosen that is convenient in relation to the products that are of interest, and then determining the unknown parameters by a procedure called "calibration". This means that the model parameters are chosen in such a way that the prices that are observed in the market for a number of selected products (typically the most liquid products in the market) are matched as closely as possible. The method is based purely on price information; historical (time series) information is not used at all. Typically the number of products that are used for calibration is larger than the number of model parameters that can be adjusted, often even much larger. Consequently, not all prices can be matched exactly and nonlinear optimization methods are used to determine a match that is best according to a chosen criterion, for instance the sum of squares of differences between prices produced by the model for the selected products and prices for these products that are observed in the market. In the search for an optimum, the selected products will have to be repriced many times with different parameter values; this is one reason for the popularity of models that allow fast pricing of liquid products.

As a simple example, consider the determination of the volatility parameter in the Black-Scholes model on the basis of the observed price of a call option with a particular strike and maturity. If we think of a model that is specified under the risk-neutral measure, then the volatility parameter is indeed the only free parameter in the BS model. If calibration is done on the basis of a single observed price, then the problem comes down to solving the equation (3.50) where now C_0 is given and σ is the unknown. To find the solution, one can make use of the fact that the derivative of the Black-Scholes price with respect to the volatility parameter can be computed exactly; indeed, calculation shows that if C_0 is given by (3.50), then

$$\frac{\partial C_0}{\partial \sigma} = S_0 \phi(d_1) \sqrt{T} \tag{3.86}$$

where $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$ is the density of the standard normal distribution, and d_1 is given by (4.21b).[27] Therefore it is possible to use *Newton's method*[28] for solving nonlinear equations. This method iteratively constructs approximations to a solution of the equation $f(x) = a$, starting from an initial guess x_0, by the rule

$$x_{k+1} = x_k - \frac{f(x_k) - a}{f'(x_k)}\,. \tag{3.87}$$

[27] The sensitivity of an option price with respect to the volatility parameter is called the *vega* of the option.

[28] Sir Isaac Newton (1643–1727), British physicist and mathematician.

In many situations, Newton's method converges quickly. The method does require that one is able to compute the value of the derivative function $f'(x)$ at any given point x. When the function f is given in analytic form it is usually not difficult to find f' as well, but in pricing applications such luxury is not always available. One then has to resort to other root-finding algorithms, which can for instance be based on the construction of an approximation to the derivative of f at a given point. Numerical algorithms must also be used when the number of prices that are used for calibration is larger than the number of free parameters in the model, so that an optimization problem has to be formulated and solved.

When the volatility parameter in the BS model is found from calibration rather than from application of an estimation technique to the observed time series of prices of the underlying asset, one speaks of the *implied volatility* (as opposed to the *historical volatility*). One reason why the implied volatility can be different from the historical volatility is that volatility is in reality not constant, and current prices reflect market's expectations regarding volatility in the future, whereas historical volatility by its nature must refer to volatility in the past. Another reason is that the BS model is quite simple and does not reflect all the risks that are perceived by the market in relation to the underlying asset. Therefore, the price based on the BS model in combination with historical volatility may produce an underestimate of the actual price. Correspondingly, the implied volatility is then higher than the historical volatility. For an example in which such a correction can be calculated explicitly, see Section 4.4.4.

3.6 The price of risk

The numéraire-dependent pricing formula (3.25) indicates that, for the purposes of derivative pricing, it is sufficient to have a model of the form (3.77), i.e. a model "under $\mathbb{Q}$". The replication recipe (3.53) shows that the same is true for replication and hedging. It therefore seems that, at least for pricing and hedging, models under $\mathbb{Q}$ are all that one needs,[29] and the real-world measure $\mathbb{P}$ can be dispensed with. In fact, this might be viewed as a happy circumstance: real-world probabilities are often difficult to determine precisely, whereas the parameters in models under $\mathbb{Q}$ can be obtained from calibration.

It should be remembered, though, that the theory of rational option pricing is built on assumptions that are at best only approximately satisfied in actual financial markets. These assumptions include the availability of costless and arbitrarily fast

[29] More precisely, a model under $\mathbb{Q}_N$ should be provided, where N is a given numéraire. However, if one such model is given, then models under equivalent martingale measures corresponding to other numéraires can be derived from the change-of-numéraire formulas in Section 3.5.1, which do not require the real-world measure.

trading, and the impossibility of market manipulation. Moreover, even if these assumptions are accepted, it may still happen that a model used for a particular situation is not correct. If a model is specified with a sufficient amount of flexibility in parameters, chances are that it can be successfully calibrated to price data, but that does not guarantee correctness of the model under changing market circumstances.

Therefore, while theoretically all risk associated with a derivative product can be hedged away in a complete market, in actual practice one still needs to be aware of risks that remain, due to market frictions, model failures, and other factors such as possible fraud. Risk management is a typical example of a domain in which real-world probabilities are important, rather than risk-neutral probabilities. In fact, one may say that while models under equivalent martingale measures are good for pricing and hedging, they are not good for anything else.[30] Since this book is about rational derivative pricing, the use of models under $\mathbb{Q}$ is pervasive, but that should not lead the reader into thinking that the real-world measure can be discarded.

The relation between the real-world measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$ can be described in terms of a Radon-Nikodym process, or equivalently, and typically more conveniently, in terms of the process λ_t that appears in Girsanov's theorem. In the context of state-space models, the process λ_t becomes a function $\lambda(t, X_t)$ of time and of the state process X_t. The term "market price of risk" that is often used for λ is motivated by the absence-of-arbitrage condition $\mu_C = r\pi_C + \sigma_C\lambda$. For instance, in the context of the Black-Scholes model, this condition can be read as stating that λ (which is a scalar in the BS model) is equal to the number of percentage points of expected return on a fixed-mix portfolio that is earned by accepting an additional percentage point of volatility. More generally, the formula describes the excess expected return[31] on an asset as the sum of products of two factors, one factor being the exposure of the asset to risk factor i as given by the i-th entry of the vector σ_C, and the other factor being the market price of risk associated to risk factor i, as given by the i-th entry of the vector λ. Decomposition of expected asset returns in this way is at the very heart of financial theory.

To connect a given model under $\mathbb{Q}$ to the real world, one needs to quantify the market prices of risk associated to the various risk factors in the model. Inserting these, one finds real-world drift terms for asset prices and other variables that may appear in the model. Alternatively, from empirical data concerning those drifts one can derive the implied market prices of risk within the model. Having a model under $\mathbb{P}$ is essential for purposes such as risk management and portfolio optimization. But even if a given model is to be used just for pricing purposes, looking at implied prices

[30] Moreover, only *expectations* under $\mathbb{Q}$ are relevant. For instance, the variance of a payoff under $\mathbb{Q}$ does not have an economic meaning, although it can be a relevant quantity in some numerical procedures (see Ch.7).

[31] The excess expected return is defined as the expected return minus the riskfree return.

of risk provides one way of testing the plausibility of proposed model parameters.

Unfortunately, empirical determination of real-world drift factors in financial models is notoriously difficult. To illustrate the problem, consider estimation of the parameter μ in the standard Black-Scholes model under $\mathbb{P}$. Assume that T years of observations are available, and that for the purpose of estimation this observation period is divided into N intervals of length $\Delta T = T/N$. According to the BS model, the log returns of the stock across these intervals are independent and normally distributed with mean $(\mu - \frac{1}{2}\sigma^2)\Delta T$ and standard deviation $\sigma\sqrt{\Delta T}$. Under the assumption that the volatility is known, taking the average of the realized log returns across the intervals of length ΔT produces an estimate of the per-period logarithmic return $(\mu - \frac{1}{2}\sigma^2)\Delta T$ that has an error bound $1.96\sigma\sqrt{\Delta T}/\sqrt{N}$ at the 95% confidence level. This implies that the error bound for the parameter μ is $\big(1.96\sigma\sqrt{\Delta T}/\sqrt{N}\,\big)/\Delta T = 1.96\sigma/\sqrt{T}$. Given that volatilities in the range of 0.1 to 0.2 are common and observation periods for stock prices are usually less than 100 years, this implies error bounds in the range of two to five or more percentage points. This is quite substantial, given that point estimates of μ are often in the range of 5 to 10 percentage points. Moreover, the assumption that the expected return is constant across periods spanning many decades is probably not warranted. We simply have to live with the fact that financial markets are too noisy (in the signal processing sense) and too unstable to allow any precise estimates of drift parameters.

The large uncertainty in estimation of drift parameters has an impact on what can be said about the market price of risk. For instance, if in the BS model one takes $\mu = 8\%$, $r = 2\%$, and $\sigma = 20\%$, then the corresponding price of stock market risk is found to be

$$\lambda = \frac{\mu - r}{\sigma} = \frac{0.08 - 0.02}{0.20} = 0.30.$$

But one may also take a lower estimate of the expected return, for instance 5%; the market price of stock market risk that would follow from this is 0.15. If one sets $\mu = 11\%$ and $\sigma = 15\%$, while keeping $r = 2\%$, then one arrives at $\lambda = 0.60$. While the differences between these estimates are substantial, still at least some indication can be derived about what is a reasonable level for the price of stock market risk.

In addition to empirical research, another view on the market price of risk can be derived from economic theory. Briefly, a standard ("neoclassical") analysis might proceed as follows. Consider a representative investor whose utility from wealth w_0 at time 0 and wealth w_T at time T is given by

$$E\big[u(w_0) + \delta_T u(w_T)\big]$$

where $u(\cdot)$ is a utility function, δ_T is a deterministic subjective discount factor,[32] and

[32]The term "subjective" here means that the number δ_T is specified as part of the agent's pref-

expectation is taken under the real-world measure. Suppose that w_0 (deterministic) and w_T (stochastic) represent the current situation of the investor, and that the investor has the opportunity to either buy or sell a small amount of a contract that generates payoff C_T at that time T and that has price C_0. The investor will neither buy nor sell if she is indifferent, in other words, if the following equality holds:

$$\Big(\frac{d}{d\alpha}E\big[u(w_0-\alpha C_0)+\delta_T u(w_T+\alpha C_T)\big]\Big)\Big|_{\alpha=0}=0. \tag{3.88}$$

The assumption the investor is representative implies that the indifference equation (3.88) is satisfied in equilibrium. In other words, (3.88) can be taken as an equilibrium equation that determines the equilibrium price at time 0 of the asset generating the stochastic payoff C_T at time T.

A more explicit form of (3.88) is obtained by differentiating and inserting $\alpha = 0$. One finds

$$-C_0u'(w_0)+\delta_T E\big[u'(w_T)C_T\big]=0$$

so that

$$C_0=E\Big[\delta_T\,\frac{u'(w_T)}{u'(w_0)}\,C_T\Big].$$

This is the pricing formula. We can write it as $C_0 = E[K_T C_T]$ if we define

$$K_T=\delta_T\,\frac{u'(w_T)}{u'(w_0)}\,. \tag{3.89}$$

In this way, the pricing kernel can be viewed as marginal utility of representative wealth at time T, normalized by marginal utility of representative wealth at time 0 and a discount factor.

Within the context of the generic state space model, the vector λ relates to the pricing kernel through (see (3.83)

$$dK_t=-K_t\big(r(t,X_t)\,dt+\lambda(t,X_t)^\top dW\big).$$

This implies (in shorthand notation)

$$d(\log\,K)=\frac{1}{K}\,dK-\frac{1}{2}\,\frac{1}{K^2}\,d[K,K]=-(r+\tfrac{1}{2}\lambda^\top\lambda)\,dt-\lambda^\top dW.$$

Given the relation of the pricing kernel to marginal utility, we can therefore think of $-\lambda_i\Delta W_i$ as the shock in the log marginal utility of the representative investor that is generated by a shock ΔW_i in the i-th driving Brownian motion. These relations are shown in the diagram of Fig. 3.1.

If the utility function is concave, as usually assumed, then marginal utility is a

erences, rather than as a market rate.

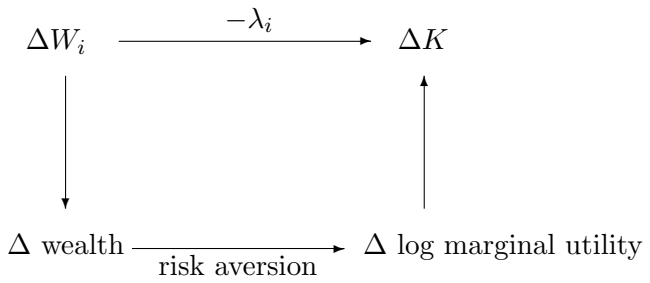

Figure 3.1: The neoclassical view of the market price of risk.

decreasing function of wealth. Therefore the following relations are suggested:

- the market price of a risk factor is *positive* when *positive* shocks correlate *positively* with the wealth of the average investor;
- the market price of a risk factor is *negative* when *positive* shocks correlate *negatively* with the wealth of the average investor.

This, for instance, provides reason to believe that the price of stock market risk should be positive. As always, correlation should not be confused with causation. For instance, higher oil prices are in general correlated with higher levels of economic activity representing good news for the representative investor; consequently, oil price risk might be priced positively, even though higher cost of energy is unfavorable to many (but not all) companies. Market prices of risk may well be state-dependent. For instance, it could be that the market price of oil price risk is positive at low levels of oil prices, but goes down as the oil price rises and becomes negative at high levels.

3.7 Exercises

1. This exercise calls for a proof of the Fundamental Theorem of Asset Pricing in the simplest possible case: two assets, two possible future states. The symbol $\mathbb{R}^2_+$ denotes the nonnegative cone in $\mathbb{R}^2$, i.e. the set of all 2-vectors with nonnegative entries, also known as the first quadrant.

a. Prove that the following statements are equivalent for a one-dimensional subspace $\mathcal{V}$ of $\mathbb{R}^2$:

(i) $\mathcal{V}$ intersects $\mathbb{R}^2_+$ only in the point 0;

(ii) there exist positive real numbers y_1 and y_2 such that

$$\mathcal{V} = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;\middle|\; y_1 x_1 + y_2 x_2 = 0 \right\}.$$

b. Consider two assets that trade at time 0 at price S_0 and B_0 respectively, and that at time 1 either have prices S_u and B_u, or S_d and B_d. Assume that all of B_0, B_u and B_d are positive. Show that exactly one of the following statements holds:

(i) there exist numbers ϕ_0 and ψ_0 such that $\phi_0 S_0 + \psi_0 B_0 = 0$, $\phi_0 S_u + \psi_0 B_u \geq 0$, $\phi_0 S_d + \psi_0 B_d \geq 0$, and at least one of $\phi_0 S_u + \psi_0 B_u$ and $\phi_0 S_d + \psi_0 B_d$ is positive;

(ii) there exist positive real numbers q_u and q_d, with $q_u + q_d = 1$, such that $q_u S_u/B_u + q_d S_d/B_d = S_0/B_0$.

2. Consider an asset whose price S_t follows a process given by

$$dS_t = \mu_S(t, S_t)\, dt + \sigma_S(t, S_t)\, dW_t.$$

Suppose that there is another traded asset whose price C_t is determined as a continuously differentiable function $\pi_C(t, S_t)$ of t and S_t. Assume that (i) the price S_t is always positive, (ii) the volatility $\sigma_S(t, S)$ is always positive, and (iii) the relative price C_t/S_t is a strictly increasing function of S_t (in other words, the function $\pi_C(t, S)/S$ is strictly increasing as a function of S for every fixed value of t).

a. Prove that the market consisting of the two assets S_t and C_t is complete and arbitrage-free.

b. Assume now that a third asset is given by the equation $dB_t = rB_t\, dt$, where r is a constant. State the conditions under which the market is still arbitrage-free.

c. Assuming the conditions of the previous part, show how the value of the asset B_t can be replicated by a self-financing portfolio consisting of the assets S_t and C_t.

3. Verify that the function defined by (3.51) solves the partial differential equation (3.52). While this can be done by brute-force differentiation, it may help to establish the following facts in order:[33]

$$\tfrac{1}{2}(d_1^2 - d_2^2) = \tfrac{1}{2}(d_1 + d_2)(d_1 - d_2) = \log(S/K) + r(T - t) \tag{3.90}$$

$$S\phi(d_1) = e^{-r(T-t)} K \phi(d_2) \tag{3.91}$$

$$\frac{\partial \pi}{\partial t}(t, S) = -re^{-r(T-t)} K \Phi(d_2) - S\phi(d_1)\, \frac{1}{2\sigma\sqrt{T - t}} \tag{3.92}$$

$$\frac{\partial \pi}{\partial S}(t, S) = \Phi(d_1) \tag{3.93}$$

$$\frac{\partial^2 \pi}{\partial S^2}(t, S) = \phi(d_1)\, \frac{1}{\sigma S\sqrt{T - t}}\,. \tag{3.94}$$

[33] The notation $\phi(x)$ is used for the derivative of the standard normal cumulative distribution function, i.e. $\phi(x)$ is the standard normal density function $(1/\sqrt{2\pi})\exp(-\frac{1}{2}x^2)$.

Be sure to verify the boundary condition as well.

4. Produce a plot of the call option value (3.50) as a function of the value of the underlying for the following parameter values: $t = 0$, $K = 100$, $T = 1$, $\sigma = 0.2$, $r = 0.04$. Also draw this plot when $T = 5$, and the other parameter values are the same as before. To which limit does the plot tend if T is increased more and more?

5. Compute the limit of the call option price (3.50) as σ tends to zero and as σ tends to infinity, while other parameters remain fixed.

6. The Black-Scholes formula (3.50) appears to give the call option price as a function of five parameters, namely the current stock price S_t, the time to maturity T, the strike K, the volatility σ, and the interest rate r. However, show that the option price relative to the stock price, C_0/S_0, can be written as a function of only two parameters, namely $m = e^{rT}S_0/K$ ("moneyness")[34] and $\sigma_{\text{tm}} = \sigma\sqrt{T}$ ("volatility to maturity"). Plot the relative call option price as a function of volatility to maturity in the range from 0 to 0.5 when moneyness is equal to 0.8, 0.9, 1, 1.1, and 1.2. Which approximation would be suggested when the moneyness is 1?

7. For fixed values of r, T, and K, define a function of two variables $f(S_0, x)$ by

$$f(S_0, x) = S_0\Phi(x) - e^{-rT}K\Phi(x - \sigma\sqrt{T}) \tag{3.95}$$

so that $f(S_0, d_1)$ with d_1 given by (3.50b) is the Black-Scholes call option value.

a. Show that the partial derivative of $f(S_0, x)$ with respect to x, when evaluated at $x = d_1$, is equal to 0.

b. Using part a., derive the relation (3.93).

c. Let the function $g(y; a)$, with parameter a, be defined for $y \in [0, 1]$ by

$$g(y; a) = \Phi\big(a + \Phi^{-1}(y)\big).$$

Prove that this function is convex if the parameter a is positive.

d. Show that the function $f(S_0, x)$ defined in part a., when considered as a function of x for fixed S_0, has a global maximum at $x = d_1$. [*Hint*: use part c.]

8. A market is given as follows:

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sigma S_t\,dW_{1,t} \\ dF_t &= \mu_1 F_t\,dt + \sigma_1 F_t\,dW_{1,t} + \sigma_2 F_t\,dW_{2,t} \\ dB_t &= rB_t\,dt. \end{aligned}$$

[34] We can also write $m = S_0/(e^{-rT}K)$, to express moneyness explicitly as the quotient of the current value of the underlying and the current value of the strike.

All three variables represent prices of traded assets; S_t is a stock index, F_t is the share price of an exchange-traded investment fund, and B_t is a bond. The parameters μ, σ, μ_1, σ_1, σ_2, and r are positive constants. The Brownian motions $W_{1,t}$ and $W_{2,t}$ are independent.

a. Show that the market as defined above is arbitrage-free and complete.

b. Show that, if the price of risk associated to the Brownian motion $W_{2,t}$ is zero, the following relation holds:[35]

$$\mu_1 - r = \frac{\sigma_1}{\sigma}(\mu - r). \tag{3.96}$$

What could be a possible economic motivation for the assumption that the price of risk associated to $W_{2,t}$ is zero?

9. In this exercise, we test the standard Black-Scholes delta hedge strategy when it is implemented in discrete time, and relate its behavior to the final asset value and to realized volatility. We consider a call option with strike $K = 100$ maturing at time $T = 1$; current time is 0, current value of the underlying asset is $S_0 = 100$, the drift parameter is $\mu = 0.12$, the volatility parameter is $\sigma = 0.20$, and the interest rate is $r = 0.04$; the time step is $\Delta t = 0.01$.

Write a program to generate scenarios for the joint evolution of the following variables: stock price S; bond price B; value of the replicating portfolio V; stock holdings in the replicating portfolio ϕ_S; bond holdings in the replicating portfolio ϕ_B. Let the initial value of the replicating portfolio be equal to the price of the option according to the Black-Scholes formula (3.50). At each time step, update ϕ_S according to the delta strategy (note that the delta corresponding to the call option price is given in (3.93)) and then update ϕ_B in such a way that the replicating portfolio is self-financing. To update the values of S and B, you may use the exact formulas or use Euler approximations.

a. The hedge error is defined as

$$\text{HE} = V_T - \max(S_T - K, 0)$$

where V_T is the value of the replicating portfolio at the time of maturity. Use the program you have written to generate 1000 scenarios, and plot a histogram of the hedge errors that you obtain from these scenarios. Also compute the expectation and the standard deviation of the hedge error. Compare the standard deviation of the hedge error to the standard deviation of the option payoff itself. Is the hedge

[35] The formula (3.96) is reminiscent of the relation that is derived in the Capital Asset Pricing Model (CAPM) between the excess expected return of a given investment portfolio and the excess expected return on a broad market portfolio.

effective? [*Hint*: If the standard deviation of the hedge error that you find is larger than 1% of the value of the underlying at time 0, you may want to have another look at your implementation of the delta hedge.]

b. Plot a scatter diagram of the hedge error versus the final asset value S_T. Which phenomenon do you observe?

c. The *realized volatility* corresponding to a particular price path $(S_{t_0}, S_{t_1}, \dots, S_{t_N})$, where $t_i = i\Delta t$ and $N = T/\Delta t$, is defined by[36]

$$\sigma_{\text{real}} = \left(\frac{1}{T} \sum_{i=1}^{N} \left(\frac{S_{t_{i+1}} - S_{t_i}}{S_{t_i}} \right)^2 \right)^{\frac{1}{2}}.$$

Extend your program so that it also computes realized volatility along each of the trajectories. Again generate 1000 scenarios and plot a scatter diagram of the hedge error versus the realized volatility. Which phenomenon do you observe?

10. In this exercise we compare two replication strategies for a call option in the BS model, using the same parameters as in Exc. 9. One is the standard delta hedge, and the other is the so called stop-loss hedge. The latter strategy holds one unit of the stock as long as the option is in the money (i.e. $S_t > e^{-r(T-t)}K$), and sells the stock as soon as the option goes out of the money. The idea behind this strategy is that the hedger (i.e. the party that has written (sold) the option) will need to deliver one unit of the stock if the option ends in the money, while no delivery needs to take place in the opposite case. Of course, the trading strategy in stocks needs to be accompanied by an appropriate trading strategy in bonds in order to form a self-financing hedge portfolio. To assess the quality of the two strategies, we can use a numéraire, as in (3.16).

a. Take the bond as a numéraire. Show that the relative asset price $\hat{S}_t := S_t/B_t$ satisfies the SDE

$$d\hat{S}_t = (\mu - r)\hat{S}_t\, dt + \sigma \hat{S}_t\, dW_t.$$

Define $\hat{K} = e^{-rT}K$. Verify that, in terms of the variable $\hat{S}_t$, the delta hedge is given by

$$\phi^{\Delta}_{1,t} = \Phi(d_{1,t}), \qquad d_{1,t} = \frac{\log(\hat{S}_t/\hat{K}) + \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

and the stop-loss hedge by

$$\phi^{\text{SL}}_t = \begin{cases} 1 & \text{if } \hat{S}_t > \hat{K} \\ 0 & \text{if } \hat{S}_t \leq \hat{K}. \end{cases}$$

[36] An alternative definition that is also sometimes used replaces the relative return $(S_{t_{i+1}} - S_{t_i})/S_{t_i}$ by the log return $\log(S_{t_{i+1}}/S_{t_i})$. The difference between the two quantities is small when $S_{t_{i+1}}$ is close to S_{t_i}.

In both cases, ϕ_t indicates the amount of units of the stock that are in the hedge portfolio at time t.

b. Write a script that generates scenarios of the stock price in discrete time steps Δt. Use the exact formula for the geometric Brownian motion, rather than the Euler approximation. Compute the sum of trading gains/losses corresponding to the stop-loss hedge in 1000 scenarios (see the summation term in (1.13)), using step size $\Delta t = 0.1$, and also the outcomes in the same scenarios of the call option payoff (relative to the numéraire at time T) $\hat{C}_T = \max(\hat{S}_T - \hat{K}, 0)$. Draw a scatter plot, and compute the standard deviation of the difference between the two variables in the scenarios that you generated. Repeat these steps for the delta hedge. Which of the two strategies leads to the smallest standard deviation?

c. Repeat the experiment of part b. using $\Delta t = 10^{-k}$ for $k = 2, 3, 4$. Which trends do you observe?

d. Explain how the Black-Scholes price of the option is reflected in the scatter diagrams.

11. The dynamic hedging strategy as developed in Section 3.3.4 is model-dependent. Alternatively, one can hedge a given product by replicating its payoff as well as possible by a linear combination of other products which are already liquidly traded in the market. This is called *static hedging*.

a. A *butterfly option*, written on an underlying with price S_t at time t, is a contract that has a payoff of the following form:

$$C_T = \begin{cases} 0 & \text{if } S_T \leq K_1 \text{ or } S_T \geq K_3 \\ L(S_T - K_1)/(K_2 - K_1) & \text{if } K_1 \leq S_T \leq K_2 \\ L(K_3 - S_T))/(K_3 - K_2) & \text{if } K_2 \leq S_T \leq K_3 \end{cases} \tag{3.97}$$

where K_1, K_2, K_3 and L are parameters such that $K_1 < K_2 < K_3$ and $L > 0$. Show that the payoff of this contract can be perfectly replicated by a linear combination of three call options.

In cases where a given product cannot be perfectly replicated by static hedging, one may still look look for a linear combination of available assets that replicates the target product as closely as possible. As a measure of hedge quality, one can use for instance

$$\text{HQ} = 1 - \frac{\text{std}(C_T - H_T)}{\text{std}(C_T)} \tag{3.98}$$

where C_T is the payoff of the target product, H_T is the value at time T of the hedge portfolio, and the standard deviation is taken under the real-world probability measure.

b. Prove that, if H_T results from an linear combination of available assets which is optimal in the sense of the criterion (3.98), then

$$\mathrm{HQ} = 1 - \sqrt{1 - \rho^2} \tag{3.99}$$

where ρ is the correlation coefficient of C_T and H_T. What value does the correlation coefficient ρ need to have in order to achieve 50% hedge quality? Which value is needed to get 80% hedge quality?

12. Let S_t denote the price of an asset and suppose that S_t follows a geometric Brownian motion, so $dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$. Let C_t denote the price of some derivative and assume that the interest rate is zero, so that C_t is given by $C_t = \pi_C(t, S_t)$ where π_C is a function of two variables t and S that satisfies the partial differential equation

$$\frac{\partial \pi_C}{\partial t}(t, S) + \tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 \pi_C}{\partial S^2}(t, S) = 0. \tag{3.100}$$

Consider now the portfolio that is composed as follows: long one unit of the derivative, short $\frac{\partial \pi_C}{\partial S}$ units of the asset. So the value of the portfolio is given by $V_t = f(t, S_t)$ with

$$f(t, S) \;=\; \pi_C(t, S) - S\,\frac{\partial \pi_C}{\partial S}(t, S). \tag{3.101}$$

a. Prove that

$$dV_t \;=\; -S_t^2 \frac{\partial^2 \pi_C}{\partial S^2}(t, S_t)(\mu\, dt + \sigma\, dW_t). \tag{3.102}$$

[*Hint*: You may want to use, among other things, the equation that you get by differentiating (3.100) with respect to S.]

b. Is the portfolio V_t self-financing, i. e. does it satisfy $dV_t = dC_t - \frac{\partial \pi_C}{\partial S}(t, S_t)\, dS_t$?

13. Prove the following generalization of the replication theorem Thm 3.3.2. Assume the same conditions as in the theorem, except for the completeness, so that the equation $\mu_Y - r\pi_Y = \sigma_Y \lambda$ may have non-unique solutions. Show that a portfolio value function $\pi_C = \pi_C(t, x)$ can be replicated if and only if the equation $\mu_C - r\pi_C = \sigma_C \lambda$ is satisfied for *all* pairs (r, λ) such that $\mu_Y - r\pi_Y = \sigma_Y \lambda$.

14. For a European investor, a savings account with an American bank is not the same as a savings account with a European bank, because the value in euros of the American account is affected by the euro/dollar exchange rate, whereas this is not the case for the European account. Also, the American interest rate may be different from the European rate. The situation may be modeled as follows:

$$dB_t^E = r_E B_t^E\, dt$$

$$
\begin{aligned}
dB_t^A &= r_A B_t^A \, dt \\
dZ_t &= \mu Z_t \, dt + \sigma Z_t \, dW_t \\
S_t &= Z_t B_t^A
\end{aligned}
$$

where B_t^E and B_t^A are the values of the European and the American accounts in their respective currencies, r_E and r_A denote the European and the American rate respectively, Z_t is the exchange rate from dollars to euros, μ and σ are constants, and S_t is the value of the American account in euros.

a. Show that the model is complete (with assets B^E and S), and determine the risk-free rate of return and the market price of risk.

b. Suppose that B^E is taken as a numéraire. Give the exchange rate dynamics under the corresponding equivalent martingale measure.

15. **a**. Verify the validity of the expression (3.86) for the vega of a call option in the Black-Scholes model.

b. Implement Newton's method (3.87) to find the implied volatility, according to the Black-Scholes model, of a call option that has time to maturity $T = 1$, strike $K = 100$, and market price $C_0 = 8.00$, given that the current price of the stock is $S_0 = 100$ and the interest rate is $r = 0.04$. As an initial guess for the implied volatility, take any positive number that you like. As a stopping criterion for the iteration (3.87), accept the value x_k as the true value when the difference between x_k and the previous estimate x_{k-1} is less than 10^{-5}. How many steps does the method take to converge? Verify the validity of the answer by applying the Black-Scholes formula in which you take σ equal to the value that you have found from the iteration.

c. Repeat part b., but this time stop only when the difference between successive iterates in Newton's method is less than 10^{-10} instead of 10^{-5}. How many steps does it take now for the algorithm to converge?

This page intentionally left blank

Chapter 4

Analytical option pricing

This chapter gives an outline, with examples, of the basic methods of computation of prices for contingent claims. The emphasis here is on cases in which an analytical solution may be found. Numerical methods are discussed in later chapters.

4.1 Three ways of pricing

4.1.1 The Black-Scholes partial differential equation

Let a model for a financial market be given in state space form (3.1). If the model is free of arbitrage, there exist functions $r(t,x)$ and $\lambda(t,x)$ such that the Black-Scholes equation (3.31) is satisfied. Lemma 3.2.5 then states that, for any self-financing strategy $\phi = \phi(t,x)$, the pricing function $\pi_C = \phi^\top \pi_Y$ satisfies the equation $\mu_C = r\pi_C + \sigma_C\lambda$ where μ_C and σ_C are related to π_C as in (3.6). Writing this equation explicitly, we obtain

$$\frac{\partial \pi_C}{\partial t} + \frac{\partial \pi_C}{\partial x}\mu_X + \tfrac{1}{2}\operatorname{tr}\frac{\partial^2 \pi_C}{\partial x^2}\sigma_X\sigma_X^\top = r\pi_C + \frac{\partial \pi_C}{\partial x}\sigma_X\lambda \tag{4.1}$$

or, in slightly rewritten form,

$$-\frac{\partial \pi_C}{\partial t} = \tfrac{1}{2}\operatorname{tr}\frac{\partial^2 \pi_C}{\partial x^2}\sigma_X\sigma_X^\top + \frac{\partial \pi_C}{\partial x}(\mu_X - \sigma_X\lambda) - r\pi_C. \tag{4.2}$$

In the theory of partial differential equations, the above equation is classified as being of backward parabolic type. Under mild conditions, when a function $F(x)$ is given, there exist solutions of this equation for $t \le T$ which are such that $\pi_C(T,x) = F(x)$. In a complete market, this means that there is a self-financing portfolio that generates a given state-dependent payoff at time $t = T$; in other words, the portfolio replicates a European derivative with payoff function $F(X_T)$. The value of this portfolio at any time prior to T is the arbitrage-free price of the derivative at time t, expressed as a function of the current time and the current state. More

generally, solutions can be defined by conditions along other boundaries as well. This is illustrated in the example below.

Example 4.1.1 Consider the pricing of a perpetual double-barrier option in the standard Black-Scholes model. The model is

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t \tag{4.3a}$$

$$dB_t = rB_t\,dt \tag{4.3b}$$

where both S and B are tradable assets. The contract that we consider here pays 1 euro when S_t reaches a given lower level L and expires worthless when S_t reaches a given upper level U. It is assumed that $L < S_0 < U$; the contract remains alive as long as neither the level L nor the level U has been reached. The Black-Scholes equation corresponding to the model (4.3) is

$$-\frac{\partial\pi}{\partial t}(t,S) \;=\; \tfrac{1}{2}\sigma^2S^2\frac{\partial^2\pi}{\partial S^2}(t,S) + rS\frac{\partial\pi}{\partial S}(t,S) - r\pi(t,S). \tag{4.4}$$

To obtain the pricing function of the contract, add the boundary conditions

$$\pi(t,L) = 1, \quad \pi(t,U) = 0 \qquad (0 < t < \infty). \tag{4.5}$$

Calendar time plays no role in the definition of the contract and so its pricing function should not depend on t. Therefore, we are looking for a function $\pi = \pi(S)$ that solves the ordinary differential equation

$$\tfrac{1}{2}\sigma^2S^2\frac{d^2\pi}{dS^2} + rS\frac{d\pi}{dS} - r\pi \;=\; 0 \tag{4.6}$$

on the interval $[L,U]$, with the boundary conditions

$$\pi(L) = 1, \quad \pi(U) = 0. \tag{4.7}$$

The equation (4.6) is linear and of second order, so it is expected that all solutions to this equation can be written as a linear combination of two independent particular solutions. One solution is obvious: the function $\pi(S) = S$ must satisfy (4.6) because it is clearly the pricing function of a self-financing portfolio, and it is easy to verify that it is indeed a solution. We need a second solution to form a linear combination with the first one that satisfies the boundary conditions (4.7). Let us try a solution of the form $\pi(S) = S^\alpha$. Inserting this in the Black-Scholes PDE leads to

$$0 \;=\; \tfrac{1}{2}\sigma^2S^2\cdot\alpha(\alpha-1)S^{\alpha-2} + rS\cdot\alpha S^{\alpha-1} - rS^\alpha \;=\; (r + \tfrac{1}{2}\sigma^2\alpha)(\alpha-1)S^\alpha$$

so that $\pi(S) = S^\alpha$ is a solution when $\alpha = 1$ (as expected) and when $\alpha = -2r/\sigma^2$.

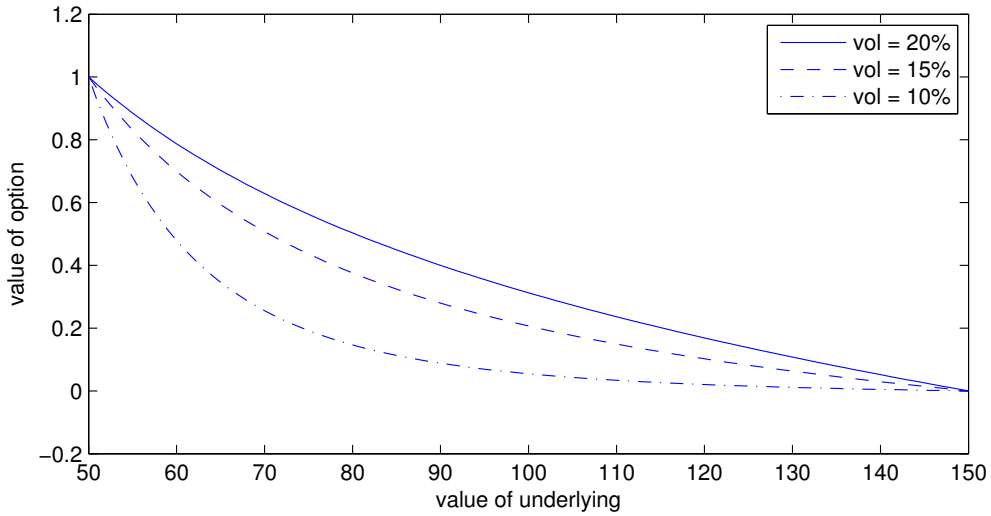

Figure 4.1: Barrier option value as a function of the value of the underlying. The parameter values are as follows: lower barrier $L = 50$, upper barrier $U = 150$, interest rate $r = 2\%$. Three different values of the volatility σ are used, namely 0.2, 0.15 and 0.1.

Abbreviate $2r/\sigma^2$ as γ. Any function of the form $\pi(S) = c_1 S + c_2 S^{-\gamma}$ where c_1 and c_2 are constants is a solution of (4.6). The constants c_1 and c_2 must be chosen to satisfy the boundary conditions; that is, we must have $c_1 L + c_2 L^{-\gamma} = 1$ and $c_1 U + c_2 U^{-\gamma} = 0$. This leads to

$$c_1 = -\frac{L^\gamma}{U^{\gamma+1} - L^{\gamma+1}}, \qquad c_2 = \frac{L^\gamma U^{\gamma+1}}{U^{\gamma+1} - L^{\gamma+1}}.$$

Therefore the price of the option is

$$\pi(t, S) = \frac{L^\gamma}{U^{\gamma+1} - L^{\gamma+1}} \left[\left(\frac{U}{S} \right)^{\gamma+1} - 1 \right] S \tag{4.8}$$

for all $t \geq 0$ and $S \in [L, U]$.

The value of the option is justified from the point of view of the seller, because the amount that is received is enough to set up a self-financing portfolio which will cover the liabilities generated by the contract, irrespective of whether the lower barrier or the higher barrier will be hit first. The price is also justified from the buyer's perspective, because the contract allows the buyer to go short in the replicating portfolio, which will generate an amount at time 0 that is exactly equal to the option price as computed above. As soon as one of the barriers is hit, the position will be unwound without any remaining liabilities; it doesn't matter that it is not known at what time this will occur. Of course, all of this depends on the standard assumptions of frictionless trading.

4.1.2 The equivalent martingale measure

The numéraire-dependent pricing formula is given in (3.25). To compute prices of contingent claims on the basis of this formula, it is easiest to work with a model that is formulated under the equivalent martingale measure associated to the chosen numéraire; cf. (3.77). Suppose that we have such a model, and suppose that we want to price a contract that has a single payoff at time T, defined by $C_T = f(X_T)$ where $f(x)$ is a given function. To determine the expected value of C_T/N_T under $\mathbb{Q}$ given information at time t, we solve the equation (3.77) with a given value of X_t at time t. In some cases it is possible to obtain an analytical solution. If not, an approximation for $E_t^{\mathbb{Q}}[F(X_T)/N_T]$ can be obtained by simulating a large number of scenarios of (3.77) taking $\widetilde{W}_t$ as a standard Wiener process, and computing the average of the obtained values of $f(X_T)/N_T$. This is the *Monte Carlo method* which is discussed more extensively in Chapter 7. A case in which an analytical solution can be obtained is shown in the following example.

Example 4.1.2 Consider the pricing of a digital option in the standard Black-Scholes model. The model is given by (4.3) as before. A *digital option* with strike K is given by the payoff function

$$F(S) = 1_{S>K} \tag{4.9}$$

which means that one unit will be paid when the price of the underlying at maturity exceeds K, and otherwise nothing will be paid. Under the equivalent martingale measure that corresponds to taking B as a numéraire, the standard Black-Scholes model is written as (3.82) where the driving process W_t is a Brownian motion under $\mathbb{Q}_B$. The price of the digital option at time 0 can now be computed from the numéraire-dependent pricing formula (3.25) in the following way:

$$\begin{aligned} C_0 &= E\Big[\frac{1_{S_T>K}}{e^{rT}}\Big] = e^{-rT}\mathbb{Q}(S_T > K) = e^{-rT}\mathbb{Q}(\log S_T > \log K) = \\ &= e^{-rT}P(\log S_0 + (r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,Z > \log K) = \\ &= e^{-rT}P\left(Z > \frac{-\log(S_0/K) - (r - \tfrac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) = e^{-rT}\Phi(d_2) \end{aligned} \tag{4.10}$$

where Z is a standard normal variable, P is the probability associated with Z (not to be confused with the "objective" measure $\mathbb{P}$), $\Phi(\cdot)$ is the cumulative normal distribution function, and the constant d_2 is defined as in (3.50b). At time t, the pricing formula is the same except that T is replaced by $T - t$ and S_0 by S_t.

4.1.3 The pricing kernel method

The pricing kernel method was discussed in Section 3.5.3. The method requires an explicit specification of the market prices of risk associated to the driving Brownian motions that appear in a given model. However, as will be shown in the example below, it may happen that the price of risk can be absorbed into model parameters. The computation can then be carried out as if the price of risk is zero, but within a model in which certain parameters have been replaced by risk-adjusted versions.

Example 4.1.3 Consider the valuation of bonds in the Vasicek model. The model is given by the state equation

$$dX_t = a(b - X_t)\,dt + \sigma\,dW_t \tag{4.11}$$

together with the following specifications of the short rate and the market price of risk:

$$r(t,x) = x, \quad \lambda(t,x) = \lambda \text{ (constant)}. \tag{4.12}$$

The product that we consider is a default-free zero-coupon bond; it pays one unit of currency at a given time T. The value of this contract at time 0, according to the Vasicek model, is given by

$$\pi_T(0,x) = E[K_T \cdot 1 \mid X_0 = x_0] \tag{4.13}$$

where the pricing kernel process is determined by (writing $r_t := r(t, X_t)$, so that $r_t = X_t$)

$$dK_t = -K_t(r_t\,dt + \lambda\,dW_t), \quad K_0 = 1. \tag{4.14}$$

The Black-Scholes partial differential equation associated to the Vasicek model is

$$-\frac{\partial \pi_T}{\partial t} = \tfrac{1}{2}\sigma^2 \frac{\partial^2 \pi_T}{\partial x^2} + [a(b-x) - \lambda\sigma]\frac{\partial \pi_T}{\partial x} - x\pi_T. \tag{4.15}$$

This shows that the Vasicek bond pricing model, which appears at first to have four parameters, actually has only three (namely a, σ, and $ab - \lambda\sigma$). For instance, the model with parameters (a, b, σ, λ) leads to the same bond prices as the model with parameters $(a, b - \lambda\sigma/a, \sigma, 0)$. We can therefore simplify the calculations by taking $\lambda = 0$, keeping in mind that the interpretation of the parameter b is then that it represents a "risk-adjusted" average level of the short rate, rather than the actual average level. Now, from (4.11) and (4.14) we obtain

$$d\begin{bmatrix} \log K_t \\ X_t \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & -a \end{bmatrix} + \begin{bmatrix} 0 \\ ab \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dW_t. \tag{4.16}$$

This is a linear stochastic differential equation (see Section 2.6). After some manipulation, one finds[1]

$$E[\log K_T \mid X_0 = r_0] = -\frac{1 - e^{-aT}}{a}(r_0 - b) - bT \tag{4.17}$$

and

$$\mathrm{var}[\log K_T \mid X_0 = x_0] = \frac{\sigma^2}{a^2}\left(T - 2\,\frac{1 - e^{-aT}}{a} + \frac{1 - e^{-2aT}}{2a}\right). \tag{4.18}$$

Using the standard rule for the expectation of a lognormally distributed variable, we find

$$\pi_T(0, r_0) = \exp\left(-\left[\left(b - \frac{\sigma^2}{2a^2}\right)T + \left(r_0 - b + \frac{\sigma^2}{a^2}\right)\frac{1 - e^{-aT}}{a} - \frac{\sigma^2}{2a^2}\,\frac{1 - e^{-2aT}}{2a}\right]\right). \tag{4.19}$$

Current time is taken to be 0 in the above formula, but the choice of the initial time in the Vasicek model is arbitrary, so that the formula above also holds for $\pi_T(t, x_t)$ except that in the right hand side x_0 must be replaced by x_t and T must be replaced by $T - t$. One way to check the correctness of the calculations is to verify that the function $\pi_T(t, x)$ satisfies the Vasicek-Black-Scholes equation (4.15) and the boundary condition $\pi_T(T, x) = 1$. The equation (4.19) describes the term structure as a linear combination of a constant and two functions of time to maturity; it therefore usually does not provide a very good fit to the actually observed term structure.

4.2 Five derivations of the Black-Scholes formula

The pricing formula (3.50) for a call option appeared in 1973 in a paper by Fischer Black and Myron Scholes that was published in the *Journal of Political Economy*. The developments leading up to this paper have been discussed in Section 1.1. Let the price of the underlying asset at time t be denoted by S_t, and take the current time to be $t = 0$ so that the current value of the underlying asset is S_0. Consider a call option with time of maturity T and strike K; the payoff of this option at time T is then given by $C_T = \max(S_T - K, 0)$. Suppose that the price of the underlying asset follows the stochastic differential equation

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t \tag{4.20}$$

where μ and σ are constants, and assume there is a constant interest rate r that applies to all maturities. On the basis of these assumptions, Black and Scholes

[1] Detailed calculations are shown in Section 4.4.3.

argued that the fair price to be paid for the option at time $t = 0$ is given by

$$C_0 = S_0\Phi(d_1) - e^{-rT}K\Phi(d_2) \tag{4.21a}$$

where Φ is the cumulative normal distribution function and the numbers d_1 and d_2 are given by

$$d_1 = \frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \qquad d_2 = \frac{\log(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\,. \tag{4.21b}$$

The publication of this formula is considered to mark the birth of modern mathematical finance. In honor of the historical formula above, five derivations of it are presented below.

In their Nobel prize winning paper, Black and Scholes derive the partial differential equation (4.24) below from the premise that, in the absence of arbitrage, an instantaneously riskless combination of assets must earn the riskless return (cf. Thm. 3.2.6). They then solve the PDE by applying a logarithmic substitution to transform the equation to one that can be found in textbooks. The equation they arrive at is the so called *heat equation* that was first studied by Fourier.[2] The first derivation below follows this line of reasoning and shows how Fourier might have solved the Black-Scholes equation. The second derivation uses the pricing kernel method, and the last three are all based on the numéraire-dependent pricing formula, with different choices of the numéraire.

The following integral identity will be used in most of the derivations to be presented below:

$$\frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{1}{2}x^2+bx}\,dx = e^{\frac{1}{2}b^2}\Phi(b-a) \tag{4.22}$$

where a and b are constants. A special case is

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}x^2+bx}\,dx = e^{\frac{1}{2}b^2}. \tag{4.23}$$

The identity is obtained as follows:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{1}{2}x^2+bx}\,dx &= \frac{e^{\frac{1}{2}b^2}}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{1}{2}(x-b)^2}\,dx \quad \overset{y=x-b}{=} \\ &= \frac{e^{\frac{1}{2}b^2}}{\sqrt{2\pi}}\int_{a-b}^\infty e^{-\frac{1}{2}y^2}\,dy = e^{\frac{1}{2}b^2}(1-\Phi(a-b)) = \\ &= e^{\frac{1}{2}b^2}\Phi(b-a).\end{aligned}$$

[2] Joseph Fourier (1768–1830), French mathematician and physicist.

4.2.1 Solving the Black-Scholes equation

Start from the partial differential equation

$$-\frac{\partial \pi}{\partial t}(t,S) \;=\; \tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 \pi}{\partial S^2}(t,S) + rS\frac{\partial \pi}{\partial S}(t,S) - r\pi(t,S). \tag{4.24}$$

We want to solve this together with the boundary condition

$$\pi(T,S) = \max(S-K,0). \tag{4.25}$$

The first step is to simplify the equation somewhat by a change of the independent variables. We introduce new variables $x = \log S$ and $\tau = T - t$ and define a new unknown function by

$$F(\tau,x) = \pi(T-\tau,\, e^x). \tag{4.26}$$

In terms of the new variables, the partial differential equation (4.24) becomes (in shorthand notation)

$$\frac{\partial F}{\partial \tau} = (r - \tfrac{1}{2}\sigma^2)\frac{\partial F}{\partial x} + \tfrac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial x^2} - rF \tag{4.27}$$

whereas the boundary condition (4.25) is transformed to

$$F(0,x) = \max(e^x - K, 0). \tag{4.28}$$

Our task is to find a solution of the PDE (4.27) that satisfies the boundary condition (4.28). The equation (4.27) by itself, without the boundary condition, should have many solutions since we should be able to accommodate many possible payoffs that may take place at time T, not just the one represented by (4.25) or equivalently (4.28). To start with, note that the PDE (4.27) is linear: if $F_1(\tau,x)$ and $F_2(\tau,x)$ are solutions, then so is $aF_1(\tau,x) + bF_2(\tau,x)$ for any constants a and b. To find particular solutions, let us try functions of the form

$$F(\tau,x) = g(\tau)h(x).$$

Inserting this trial solution into (4.27), we find

$$g'(\tau)h(x) = (r - \tfrac{1}{2}\sigma^2)g(\tau)h'(x) + \tfrac{1}{2}\sigma^2 g(\tau)h''(x) - rg(\tau)h(x)$$

or in other words

$$\frac{g'(\tau)}{g(\tau)} = \frac{(r - \tfrac{1}{2}\sigma^2)h'(x) + \tfrac{1}{2}\sigma^2 h''(x) - rh(x)}{h(x)}. \tag{4.29}$$

On the left hand side we have a function of τ, on the right hand side a function of x; they can only be equal if both are constant, say equal to c. In this way we obtain ordinary differential equations for the two functions $g(\tau)$ and $h(x)$. Both are linear equations with constant coefficients. The differential equation

$$\frac{g'(\tau)}{g(\tau)} = c$$

has $g(\tau) = e^{c\tau}$ as a solution. Inserting $h(x) = e^{\lambda x}$, we find that the right hand side of (4.29) is equal to c if the constant λ is such that

$$(r - \tfrac{1}{2}\sigma^2)\lambda + \tfrac{1}{2}\sigma^2\lambda^2 - r = c.$$

Therefore, the function

$$F(\tau, x) = \exp\Big(\big((r - \tfrac{1}{2}\sigma^2)\lambda + \tfrac{1}{2}\sigma^2\lambda^2 - r\big)\tau + \lambda x\Big) \tag{4.30}$$

is, for any constant λ, a solution of the partial differential equation (4.27). By the linearity property of (4.27), any linear combination of solutions of this type is again a solution. This gives us arbitrarily many degrees of freedom; it is not immediately clear however how to use these to satisfy the condition (4.28), which should hold for all x and which therefore represents an infinite number of constraints. To proceed, first note that the function in (4.30) may be written as

$$F(\tau, x) = e^{-r\tau} e^{\frac{1}{2}\sigma^2\lambda^2\tau} e^{(x + (r - \frac{1}{2}\sigma^2)\tau)\lambda}. \tag{4.31}$$

Using (4.23), we can write

$$e^{\frac{1}{2}\sigma^2\lambda^2\tau} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2 + \sigma\lambda\sqrt{\tau}z}\, dz$$

so that

$$F(\tau, x) = \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} e^{\lambda(x + (r - \frac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}z)}\, dz. \tag{4.32}$$

Due to the linearity of (4.27), we can conclude that a solution of (4.27) is given by any expression of the form

$$G(\tau, x) = \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \phi\big(x + (r - \tfrac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}z\big)\, dz \tag{4.33}$$

where the function ϕ is defined by

$$\phi(y) = \sum_{i=1}^{N} c_i e^{\lambda_i y} \tag{4.34}$$

and the c_i's, the λ_i's, and the number N can be chosen arbitrarily. The jump from the finite to the infinite can now be made by dropping (4.34) as a constraint and looking at (4.33), where ϕ is "any" function, as a representation of solutions of the partial differential equation (4.27).

Having parametrized the solutions of the PDE (4.27) in terms of an arbitrary function, we can now hope to be able to satisfy the boundary condition (4.28). We have

$$G(0,x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \phi(x)\, dz = \phi(x).$$

Therefore, the constraint (4.28) is satisfied by taking

$$\phi(x) = \max(e^x - K, 0)$$

and the corresponding solution is

$$G(\tau,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \max\Big(\exp\big(x + (r - \tfrac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}z\big) - K, 0\Big)\, dz. \quad (4.35)$$

This may already be viewed as an explicit expression. An evaluation in terms of the cumulative normal distribution function can be carried out as follows. Noting that

$$\exp\big(x + (r - \tfrac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}z\big) - K \geq 0 \;\Leftrightarrow\; z \geq -d$$

where

$$d := \frac{x - \log K + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$

we can write, using (4.22),

$$\begin{aligned} G(\tau,x) &= \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-d}^{\infty} e^{-\frac{1}{2}z^2} \big(\exp\big(x + (r - \tfrac{1}{2}\sigma^2)\tau + \sigma\sqrt{\tau}z\big) - K\big)\, dz = \\ &= \frac{e^{x - \frac{1}{2}\sigma^2\tau}}{\sqrt{2\pi}} \int_{-d}^{\infty} e^{-\frac{1}{2}z^2 + \sigma\sqrt{\tau}z}\, dz - \frac{e^{-r\tau}K}{\sqrt{2\pi}} \int_{-d}^{\infty} e^{-\frac{1}{2}z^2}\, dz = \\ &= e^x \Phi(d + \sigma\sqrt{\tau}) - e^{-r\tau} K \Phi(d). \end{aligned} \quad (4.36)$$

The rigor of the reasoning we have applied, in particular dropping the constraint (4.34), may be subject to debate. However, it is possible to show by direct calculation (see Exc. 3.3) that the function above is indeed a solution of the partial differential equation (4.27) with the boundary condition (4.28). An expression for the option price at time 0 in terms of the initial stock price is obtained by setting $\tau = T$ and $x = \log S_0$. We arrive at the Black-Scholes formula (4.21).

4.2.2 The pricing kernel method

The pricing kernel for the Black-Scholes model is given by

$$dK = -rK\,dt - \frac{\mu - r}{\sigma}\,K\,dW, \qquad K_0 = 1.$$

This is a geometric Brownian motion and so its solution can be written down explicitly:

$$K_t = \exp\Big[\Big(-r - \tfrac{1}{2}\Big(\frac{\mu - r}{\sigma}\Big)^2\Big)t - \frac{\mu - r}{\sigma}\,W_t\Big].$$

In terms of the pricing kernel, the price of the call option at time 0 is given by

$$C_0 = E[K_T C_T] = E[K_T \max(S_T - K, 0)].$$

Since $S_t = S_0 \exp\big((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\big)$, we find

$$\begin{aligned} C_0 = E\big[&\exp\big((-r - \tfrac{1}{2}(\mu - r)^2/\sigma^2)T - ((\mu - r)/\sigma)W_T\big) \cdot \\ &\cdot \max\big(S_0 \exp\big(\big(\mu - \tfrac{1}{2}\sigma^2\big)T + \sigma W_T\big) - K, 0\big)\big] = \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} &e^{-\frac{1}{2}z^2} \exp\big((-r - \tfrac{1}{2}(\mu - r)^2/\sigma^2)T - ((\mu - r)/\sigma)\sqrt{T}z\big) \cdot \\ &\cdot \max\big(S_0 \exp\big(\big(\mu - \tfrac{1}{2}\sigma^2\big)T + \sigma\sqrt{T}z\big) - K, 0\big)\,dz. \end{aligned}$$

The first two factors in the integrand above together form an exponential function with exponent

$$-\tfrac{1}{2}z^2 + (-r - \tfrac{1}{2}(\mu - r)^2/\sigma^2)T - ((\mu - r)/\sigma)\sqrt{T}z = -rT - \tfrac{1}{2}\big(z + \frac{\mu - r}{\sigma}\sqrt{T}\big)^2.$$

This suggests a change of variable $y = z + \frac{\mu - r}{\sigma}\sqrt{T}$. The exponent in the third factor of the integrand is then transformed to

$$(\mu - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\big(y - \frac{\mu - r}{\sigma}\sqrt{T}\big) = (r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}y.$$

We obtain

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \max\Big(S_0 \exp\big(\big(r - \tfrac{1}{2}\sigma^2\big)T + \sigma\sqrt{T}y\big) - K, 0\Big)\,dy.$$

This is the same as the right hand side of (4.35) with $x = \log S_0$ and $\tau = T$. We therefore again arrive at the Black-Scholes formula.

4.2.3 Taking the bond as a numéraire

The price of the call option can be obtained from the formula

$$\frac{C_0}{B_0} = E^{\mathbb{Q}_B} \frac{\max(S_T - K, 0)}{B_T} \tag{4.37}$$

where $\mathbb{Q}_B$ is the equivalent martingale measure that corresponds to taking the bond (which in the Black-Scholes model is the same as the money market account) as a numéraire. Since the expression above involves an expectation under $\mathbb{Q}_B$, it is convenient to work with the specification of the Black-Scholes model under the same measure. This specification is given in (3.81). In particular, the evolution of the stock price is described by

$$dS = rS\,dt + \sigma S\,dW$$

where W is a Brownian motion under the risk-neutral measure $\mathbb{Q}_B$. Since S is a traded asset, the fact that the drift term is of the form rS also follows from the rule (3.80). From the standard solution formula for geometric Brownian motion, we find

$$S_T = S_0 \exp\big((r - \tfrac{1}{2}\sigma^2)T + \sigma W_T\big). \tag{4.38}$$

Under $\mathbb{Q}_B$, the stochastic variable W_T follows a normal distribution with expectation 0 and variance T. Consequently, the pricing formula (4.37) can be written more explicitly as

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \max\big(S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}z} - K,\, 0\big)\,dz \tag{4.39}$$

which is the same as the right hand side of (4.35) with $x = \log S_0$ and $\tau = T$. Once more, we arrive at the Black-Scholes formula.

4.2.4 Taking the stock as a numéraire

In the Black-Scholes model, the stock price is always positive so that the stock can be selected as a numéraire. Therefore, the price of the call option can be obtained from the formula

$$\frac{C_0}{S_0} = E^{\mathbb{Q}_S} \frac{\max(S_T - K, 0)}{S_T} = E^{\mathbb{Q}_S} \max(1 - S_T^{-1}K, 0) \tag{4.40}$$

where $\mathbb{Q}_S$ is the equivalent martingale measure that corresponds to taking the stock as a numéraire. In this case it is convenient to work with a specification of the model under $\mathbb{Q}_S$. This specification is given in (3.82). This specification can be obtained from the formula (3.79). Alternatively (and somewhat more laboriously),

starting from the original specification of the Black-Scholes model under the real-world measure $\mathbb{P}$ (driven by a process W that is a Brownian motion under $\mathbb{P}$), the Girsanov process λ_S that is such that the process $\widetilde{W}_t$ defined by $\widetilde{W}_0 = 0$ and

$$d\widetilde{W} = \lambda_S \, dt + dW$$

is a Brownian motion under $\mathbb{Q}_S$ can be determined by noting that the process B_t/S_t must be a martingale under $\mathbb{Q}_S$, and

$$\begin{aligned} d\,\frac{B}{S} &= \frac{1}{S}\,dB - \frac{B}{S^2}\,dS + \frac{B}{S^3}\,d[S,S] = \\ &= (r - \mu + \sigma^2)\,\frac{B}{S}\,dt - \sigma\,\frac{B}{S}\,dW \\ &= -\sigma\,\frac{B}{S}\Big(\frac{\mu - r - \sigma^2}{\sigma}\,dt + dW\Big) \end{aligned}$$

so that we find

$$\lambda_S = \frac{\mu - r - \sigma^2}{\sigma}\,.$$

Inserting this in the formulation of the Black-Scholes model under $\mathbb{P}$, we also arrive at (3.82).

In particular the evolution of the stock price under $\mathbb{Q}_S$ is given by

$$dS = (r + \sigma^2)S\,dt + \sigma S\,dW \tag{4.41}$$

where W denotes a Brownian motion under $\mathbb{Q}_S$. From the standard formula for geometric Brownian motion, we have

$$S_T = S_0 \exp\big((r + \tfrac{1}{2}\sigma^2)T + \sigma W_T\big). \tag{4.42}$$

Under $\mathbb{Q}_S$, the stochastic variable W_T follows a normal distribution with expectation 0 and variance T. Consequently, the pricing formula (4.37) can be written more explicitly as

$$C_0 = \frac{S_0}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \max\big(1 - S_0^{-1} e^{-(r+\frac{1}{2}\sigma^2)T - \sigma\sqrt{T}z} K, 0\big)\,dz. \tag{4.43}$$

A change of variable $y = z + \sigma\sqrt{T}$ would lead us to (4.39). Alternatively, we can evaluate the integral directly; the computation is essentially the same as the one that was applied to evaluate the right hand side of (4.35). Note that

$$1 - S_0^{-1} e^{-(r+\frac{1}{2}\sigma^2)T - \sigma\sqrt{T}\,z} K \geq 0 \;\Leftrightarrow\; z \geq -d$$

where d is defined by

$$d = \frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

Therefore, we obtain

$$\begin{aligned} C_0 &= \frac{S_0}{\sqrt{2\pi}} \int_{-d}^{\infty} e^{-\frac{1}{2}z^2}\, dz - \frac{K}{\sqrt{2\pi}} e^{-(r+\frac{1}{2}\sigma^2)T} \int_{-d}^{\infty} e^{-\frac{1}{2}z^2 - \sigma\sqrt{T}z}\, dz \\ &= S_0\Phi(d_1) - e^{-rT} K \Phi(d_2) \end{aligned}$$

where d_1 and d_2 are defined as in (4.21b); in particular, $d_1 = d$.

4.2.5 Splitting the payoff

The payoff of the call option can be split into two parts as follows:

$$\max(S_T - K, 0) = 1_{\{S_T - K \geq 0\}}\,(S_T - K) = 1_{\{S_T - K \geq 0\}}\, S_T - 1_{\{S_T - K \geq 0\}}\, K.$$

Define

$$C_T^s = 1_{\{S_T - K \geq 0\}}\, S_T \tag{4.44}$$

$$C_T^b = 1_{\{S_T - K \geq 0\}}\, K \tag{4.45}$$

and let C_0^s and C_0^b denotes the prices at time 0 of the contracts that have the above payoffs at time T. Taking the stock as a numéraire to determine C_0^s, we find

$$\frac{C_0^s}{S_0} = E^{\mathbb{Q}_S} \frac{C_T^s}{S_T} = E^{\mathbb{Q}_S}\, 1_{\{S_T - K \geq 0\}} = \mathbb{Q}_S(S_T \geq K)$$

where $\mathbb{Q}_S(S_T \geq K)$ denotes the probability of the event $S_T \geq K$ under the measure $\mathbb{Q}_S$. To determine the price of C_0^b, take the bond as a numéraire; we find

$$\frac{C_0^b}{B_0} = E^{\mathbb{Q}_B} \frac{C_T^b}{B_T} = \frac{K}{B_T}\, E^{\mathbb{Q}_B}\, 1_{\{S_T - K \geq 0\}} = \frac{K}{B_T}\, \mathbb{Q}_B(S_T \geq K).$$

Since we must have $C_0 = C_0^s - C_0^b$, we obtain

$$C_0 = S_0\, \mathbb{Q}_S(S_T \geq K) - e^{-rT} K\, \mathbb{Q}_B(S_T \geq K). \tag{4.46}$$

It remains to determine the two probabilities. From (4.38) and (4.42), it follows that

$$\mathbb{Q}_B(S_T \geq K) = P\big(S_0 \exp\big((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z\big) \geq K\big)$$

and

$$\mathbb{Q}_S(S_T \geq K) = P\big(S_0 \exp\big((r + \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z\big) \geq K\big)$$

where Z is a standard normal variable and P denotes the probability associated with Z (not to be confused with the "objective" measure $\mathbb{P}$). We find

$$\mathbb{Q}_B(S_T \geq K) = \Phi(d_2), \qquad \mathbb{Q}_S(S_T \geq K) = \Phi(d_1)$$

and again the Black-Scholes formula appears.

4.2.6 Comments

The pricing kernel method can alternatively be viewed as an application of the numéraire-dependent pricing formula, taking as a numéraire the portfolio whose value is given by

$$V_t = K_t^{-1} = \exp\Big[\Big(r + \tfrac{1}{2}\Big(\frac{\mu - r}{\sigma}\Big)^2\Big)t + \frac{\mu - r}{\sigma} W_t\Big].$$

Since this value process is such that $K_t V_t$ is a $\mathbb{P}$-martingale (clearly, because $K_t V_t$ is constant), it is the value of a self-financing portfolio (see Section 3.5.3). In fact, it can be verified that V_t is the value at time t of a fixed-mix portfolio that starts at $V_0 = 1$ and that invests a fraction $(\mu - r)/\sigma^2$ in stocks. Among all self-financing portfolio strategies, this one has a special position in that it optimizes the expected growth rate.[3] In conclusion, all of the presented methods except the first one are applications of the numéraire-dependent pricing formula.

The option defined by the payoff (4.44) is called an *asset-or-nothing* option. The payoff (4.45) can only take two values, one of which is zero; a contract of this type is called a *digital option*.

The fifth method is the only one that does not require the integral identity (4.22). On the other hand, the other methods are more general: they lead to an integral expression

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} f\big(e^{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}z} S_0\big)\, dz \tag{4.47}$$

where f can be any payoff function that determines the contract value $C_T = f(S_T)$ at the time of maturity T. When f is a piecewise linear function, such as in the case of call and put options, the above integral can be evaluated in terms of the cumulative normal distribution function. In other cases, the integral may be evaluated numerically.

A situation in which the payoff function f is not piecewise linear arises in the pricing of *compound options*, such as the following. A "call-on-a-call" is a contract that gives the holder the right to buy at time T_1, for a price K_1 that is already determined at the initiation of the contract, a call option with strike K_2 that expires

[3] This is not a special property of the Black-Scholes model. It is true in general that the inverse of the pricing kernel is the value of the growth-optimal portfolio.

at time $T_2 > T_1$. The value of the call option at time T_1 is given, as a function of S_{T_1}, by the Black-Scholes formula. Write this function as $f_1(S_{T_1})$. Then the call-on-a-call contract may be viewed as a contract that matures at time T_1 with a payoff function defined by $f(S_{T_1}) = \max(f_1(S_{T_1}) - K_1, 0)$. The value of this contract can be obtained from the integral expression (4.47). Compound options are discussed more extensively in Section 4.3.5.

4.3 Variations

4.3.1 Multiple payoffs

A contract may specify payments at more than one date. For instance, a standard mortgage contract is paid back in monthly instalments during a long period of time. The valuation of the stream of mortgage payments is a relatively simple matter if it is assumed that payments will continue during the full running time of the contract, and the size of the payments is fixed in advance, but there are also contracts in which payments may vary in time and then the valuation can be quite a bit harder. In principle though, the pricing formula for a series of payoffs C_{T_i} at times $T_1, \ldots, T_n$ is a straightforward extension of the numéraire-dependent pricing formula (3.25):

$$C_0 = N_0 \sum_{i=1}^{n} E^{\mathbb{Q}} \frac{C_{T_i}}{N_{T_i}} . \tag{4.48}$$

In the special case of constant interest rates, we can take the money market account $M_t = e^{rt}$ as the numéraire; then

$$C_0 = \sum_{i=1}^{n} e^{-rT_i} E^{\mathbb{Q}}[C_{T_i}]. \tag{4.49}$$

This shows the NDPF as a generalized *net present value formula*.

4.3.2 Random time of expiry

In a standard European option, the time of expiry is fixed as part of the contract. However, there are also situations in which the payoff occurs at a time which cannot be predicted in advance. For instance, in the perpetual double barrier option that was discussed in Example 4.1.1, the contract expires when the price of the underlying assets first reaches one of the two boundaries. In this situation, the source of randomness in the time of expiry is the same as the one that generates the randomness in the payoff, so that there is in fact no new uncertainty; the valuation problem can be solved by adjusting the boundary conditions for the Black-Scholes partial differential equation, whereas the PDE itself remains the same. As another exam-

ple, consider the case of a *reverse mortgage* contract. In a usual mortgage contract, the debt at the time of initiation is high, and the amount of the debt is gradually decreased in time through regular repayments. In a reverse mortgage contract, the debt is zero or small in the beginning but is increasing in time; the client (usually at an advanced age) essentially takes out successive loans, to support a higher level of spending than would otherwise be possible. The debt will be settled from the proceeds of selling the home at the time at which the owner moves out, typically by admission to long-term care or to eternity. It is not known in advance when this will happen. The source of randomness in the time of expiry is in this case of a non-financial nature; in absence of sufficient data to establish an associated price of risk or more directly a probability distribution under a risk-neutral measure, the situation needs to be viewed as constituting an *incomplete* market.

In a *credit default swap*, the buyer pays regular amounts to the seller during a fixed period, in return for the seller's commitment to pay a compensation amount to the buyer when, within this period, a credit event of a certain type (specified in the contract) occurs. Such a credit event could for instance be a default of a specific company, or the occurrence of a certain number of defaults within a designated group of companies. When indeed a credit event takes place during the contract period, the series of contributions from the buyer to the seller is terminated. In the case of such contracts, the time at which payoff takes place is again random, but there may be enough trading in related instruments to establish a risk-neutral probability distribution. Consider, in general, a situation in which a contract provides a payoff $C_\tau = F(X_\tau)$, where τ is a random time, X_τ is the value of the state variable, and F is a payoff function. Suppose that a description is available for the joint distribution of the random time τ and the state process X_t under a pricing measure $\mathbb{Q}$ corresponding to a chosen numéraire N_t. The option with random time of expiry could be approximated as an option with multiple payoffs with payoff times $\Delta t, 2\Delta t, \ldots$, and with payoffs that are zero except at time $k\Delta t$, where k is such that $(k-1)\Delta t < \tau \leq k\Delta t$. According to (4.48), the time-0 value of the option is then given by

$$C_0 = N_0 \sum_{k=1}^{n} E^{\mathbb{Q}}\left[\mathbb{1}_{(k-1)\Delta t<\tau\leq k\Delta t}\,\frac{F(X_{t_k})}{N_{t_k}}\right]$$

where $n = T/\Delta t$ and T is the length of the contract period (supposed to be a multiple of Δt). Letting the time step Δt tend to 0, one can write

$$\frac{C_0}{N_0} = \lim_{\Delta t\downarrow 0} E^{\mathbb{Q}} \sum_{k=1}^{n}\left[\mathbb{1}_{(k-1)\Delta t<\tau\leq k\Delta t}\,\frac{F(X_{t_k})}{N_{t_k}}\right] = E^{\mathbb{Q}}\left[\mathbb{1}_{\tau\leq T}\,\frac{C_\tau}{N_\tau}\right] \tag{4.50}$$

which represents the option value in a form that is reminiscent of the NDPF. If it assumed that the random variable τ has a density $f_\tau(t)$ under $\mathbb{Q}$ (meaning that the

probability of the event $t \leq \tau < t + \Delta t$ under $\mathbb{Q}$ is approximately equal to $f_\tau(t)\Delta t$, for any $0 \leq t \leq T$), then one can also write

$$\frac{C_0}{N_0} = \int_0^T E^{\mathbb{Q}}\left[\frac{C_t}{N_t} \,\middle|\, \tau = t \right] f_\tau(t)\, dt.$$

When there is no dependence (under $\mathbb{Q}$) between the relative payoff and the time of expiry, then the conditioning in the expectation within the integral can be dropped, and we get a weighted average of regular option prices.

4.3.3 Path-dependent options

Sometimes the payoff of an option is defined not just in terms of the value of the underlying at the time of expiry, but in terms of the values of the underlying at several time points in the period up to expiry. The option is then said to *path-dependent*. For instance, an *Asian call option* is, by definition, a contract that pays the amount

$$C_T = \max\Big(\frac{1}{n}\sum_{i=1}^{n} S_{T_i} - K, 0\Big).$$

at time T, with $T_1 < T_2 < T_n = T$. A straightforward way to price such an option is to apply the Monte Carlo method with Euler discretization. One just needs to make sure that the sample points T_i are part of the time grid, and that the relevant cumulative sum is kept in computer memory and updated during time stepping when appropriate. Since the notion of "state variable", from a programming perspective, is just "variable that needs to be updated during the time stepping loop", this means that in fact the cumulative sum is given the status of an additional state variable. In other words, by suitably extending the set of state variables, the option payoff at the time of expiry becomes a function of the extended state at expiry only; in other words, the path dependence is removed.

In the case of the Asian call, the state variables that are added undergo discrete changes at deterministic times, so their evolution cannot be described in terms of stochastic differential equations driven by Brownian motion. It is not difficult to extend the modeling framework so that the Asian call or similar products can be included. The idea of state extension can however also be illustrated within the Brownian SDE framework, if we consider a "continuous Asian call option" with payoff

$$C_T = \max\Big(\frac{1}{T}\int_0^T S_t\, dt - K, 0\Big).$$

In this case, the payoff in fact depends on the entire path of the underlying up to the time of maturity, rather than on its values at a finite set of sample points. Such an option cannot be realized in practice, but it could be viewed as an approximation

of an Asian option with a high density of sample points. The state dependence can be accommodated by introducing a new state variable A_t which satisfies

$$dA_t = S_t\,dt, \quad A_0 = 0. \tag{4.51}$$

The option payoff is then given by

$$C_T = \max\left(\tfrac{1}{T}A_T - K, 0\right).$$

When also an SDE is given for S_t and the state vector X_t is defined as the vector with components S_t and A_t, then the resulting model appears as a standard state space model with the option payoff given as a function of the state variable at time T. There is no explicit expression for the density of A_T, so the option price at time 0 must be evaluated by Monte Carlo. If the *geometric* average is used instead, then the option price can be computed analytically.

4.3.4 Costs and dividends

In the theory we assume that assets are self-financing, but, in the real world, stocks often generate dividends, and commodities typically bring storage costs. To fit assets that generate costs and dividends into the theory, a strategy that may be applied is to specify the way in which the dividends are used (for instance they could be placed in a savings account), or, in the case of costs, how these are financed (for instance, the required funds may be taken from a savings account). In this way, the given asset becomes part of a self-financing portfolio. Assuming a complete market, the distribution of the asset under a suitable pricing measure can then be derived.

To illustrate, suppose that S_t is the price at time t of a dividend-paying stock, and assume for convenience that dividend is paid continuously at a fixed rate, as a percentage of the stock price. Suppose that the stock price follows the usual Black-Scholes model

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sigma S_t\,dW_t \\ dB_t &= rB_t\,dt. \end{aligned}$$

The assumption that the stock pays dividends implies that S_t is *not* the price of a self-financing portfolio. Consequently, one cannot argue that the quotient S_t/B_t must be a martingale under the risk-neutral measure, and it would not be correct that the drift term in the SDE for S_t under the risk-neutral measure is given by rS_t. Instead, one can do the following.

Since it is assumed that the stock pays a fixed-percentage dividend continuously, the dividend received from one unit of the stock during the interval from t to $t + \Delta t$

is $qS_t\Delta t$ where q is a constant. Assume that the dividends arising from one unit of the stock are placed into a savings account, and let D_t denote the value of this account. The value in the account changes from time t to time $t+\Delta t$ due to interest that is received as well as to dividends that flow into the account. Up to first order in Δt, one can write

$$D_{t+\Delta t} = D_t + rD_t\Delta t + qS_t\Delta t.$$

This leads to the continuous-time equation

$$dD_t = (rD_t + qS_t)\,dt.$$

Contrary to the asset S_t, the portfolio $V_t := S_t + D_t$ is self-financing. Therefore, under the risk-neutral measure $\mathbb{Q}_B$, one has

$$dV_t = rV_t\,dt + \sigma S_t\,dW_t^B.$$

From the relation $dV_t = dS_t + dD_t$ it follows that $dS_t = dV_t - dD_t$. Therefore,

$$\begin{aligned} dS_t &= rV_t\,dt + \sigma S_t\,dW_t^B - (rD_t + qS_t)\,dt \\ &= r(S_t + D_t)\,dt + \sigma S_t\,dW_t^B - (rD_t + qS_t)\,dt \\ &= (r-q)S_t\,dt + \sigma S_t\,dW_t^B. \end{aligned}$$

Another destination for the dividends could be chosen. For instance, they might be re-invested into the stock. Let V_t be the value at time t of a portfolio which is completely invested in the stock with price S_t, with reinvestment of dividends. The change in value of this portfolio between time t and time $t+\Delta t$ is due to the change in price of the stock and to dividend received. Note that V_t/S_t is the number of units of the stock. The dividend received in the period from t to $t+\Delta t$ is (to first order) equal to qV_t. We have for small Δt:

$$V_{t+\Delta t} = V_t + \frac{V_t}{S_t}\big(S_{t+\Delta t} - S_t\big) + qV_t\Delta t.$$

The corresponding continuous-time equation is:

$$dV_t = \frac{V_t}{S_t}\big(dS_t + qS_t\,dt\big) = (\mu + q)V_t\,dt + \sigma V_t\,dW_t.$$

The portfolio V_t is self-financing, so under $\mathbb{Q}_B$:

$$dV_t = rV_t\,dt + \sigma V_t\,dW_t^B.$$

From $dV_t = (V_t/S_t)(dS_t + qS_t\,dt)$, we have $dS_t = (S_t/V_t)(dV_t - qV_t\,dt)$. Therefore

$$dS_t = (r-q)S_t\,dt + \sigma S_t\,dW_t^B$$

which is the same as the result that was obtained before, as it should be.

After obtaining the SDE under the risk-neutral measure for the dividend-paying asset, pricing formulas for options can be derived. For instance, the formula for a standard European call option written on S_t becomes

$$C_0 = e^{-qT}S_0\Phi(d_1) - e^{-rT}K\Phi(d_2)$$
$$d_1 = \frac{\log(S_0/K) + (r - q + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \qquad d_2 = d_1 - \sigma\sqrt{T}.$$

This is similar to the Black-Scholes formula, with redefined parameters. When $q = 0$, the standard formula is recovered.

4.3.5 Compound options

A *compound option* is an "option on an option". For instance, a standard put option gives the holder the right to buy the underlying asset at a given time T for a predetermined price K. Instead, one could also think of a contract that gives the holder the right to buy, at time T_1, for a predetermined price K_1, a put option on a given underlying asset that matures at time T_2 and has strike K_2. This is a "put on a put". Likewise, one could have a put on a call, a call on a put, and so on. Besides such fanciful products, there are many contracts that can be modeled as compound options. Here are some examples:

- options that give the holder the right to exercise at one or more time points before expiry (Bermudan options)
- forward start options (for instance a call option that will start at time T_1 with a strike determined at that time)
- Asian options with a finite number of sample points.

There is a universal principle by which all compund options can be priced. The principle states that one should work *backwards* from the time of expiry, using the rule

$$\text{payoff of option starting at } T_{i-1} = \text{value of option starting at } T_i.$$

The application of this principle is illustrated in the following examples.

Example 4.3.1 Consider a call option that will start at time $T_1 > 0$ and that will expire at time $T_2 > T_1$, with a strike given by the value of the underlying at time T_1.

This is a particular case of a forward start option. Let the value of the underlying at time t be given by S_t, and assume that the Black-Scholes model holds. To determine the value of the option at time 0, first determine its value at time T_1. We have

$$C_{T_1} = S_{T_1}\Phi(d_1) - e^{-r(T_2-T_1)}S_{T_1}\Phi(d_2)$$

with (since $K = S_{T_1}$)

$$d_1 = \frac{r + \frac{1}{2}\sigma^2}{\sigma}\sqrt{T_2 - T_1}\,, \quad d_2 = \frac{r - \frac{1}{2}\sigma^2}{\sigma}\sqrt{T_2 - T_1}\,.$$

Therefore,

$$C_{T_1} = S_{T_1}\big[\Phi(d_1) - e^{-r(T_2-T_1)}\Phi(d_2)\big]$$

where d_1 and d_2 are *deterministic*. So, for $t < T_1$,

$$C_t = S_t\big[\Phi(d_1) - e^{-r(T_2-T_1)}\Phi(d_2)\big]$$

Another way of writing this is

$$C_t = (C_0/S_0)S_t, \qquad 0 \le t \le T_1$$

where C_0 is the value at time 0 of a call option with strike $K = S_0$ and maturity $T_2 - T_1$. Therefore, the value of the forward start option up to time T_1 is just a constant multiple of the price of the underlying. For $t \ge T_1$, the option value is that of a call option with strike S_{T_1} and maturity $T_2 - t$.

Example 4.3.2 As another example of a product that can be modeled as a compound option, consider a Bermudan option with two exercise dates. Specifically, suppose we have a put option with strike K that can be exercised at times T_1 and at time $T_2 > T_1$. At time T_1, the value of the option is

$$C_{T_1} = \max\big(\max(K - S_{T_1}, 0), C^{\text{cv}}_{T_1}\big) = \max\big(K - S_{T_1}, C^{\text{cv}}_{T_1}\big)$$

where $C^{\text{cv}}_{T_1}$ ("cv" for "continuation value") is the value at T_1 of a put option with strike K maturing at time T_2. The time-0 value of the Bermudan option is the same as the time-0 value of a contract that matures at T_1 with payoff C_{T_1}. This contract is a standard European option, be it with a somewhat complicated payoff function. Within the Black-Scholes model, the payoff at time T_1 can be given in analytic form. To compute the value at time 0, a numerical technique has to be used. The process can be repeated when there are more than two exercise dates.

4.4 Further worked examples

4.4.1 The perpetual American put

The perpetual American put is a claim that can be exercised at any time to produce a payoff $\max(K - S_t, 0)$, where K is the strike and S_t is the value of the underlying at the time of exercise. Assume that we are in the standard framework of the Black-Scholes model. The holder of the option is likely to follow an exercise strategy which maximizes the option's value. One can expect that for high values of S_t, exercise is not optimal; in particular if $S_t > K$ then exercise would clearly be useless. On the other hand, for low values of S_t it may be attractive to exercise the option. There is no reason why the exercise strategy should depend on calendar time, and so the optimal strategy is expected to be of the following form: do not exercise as long as $S_t > c$ for a certain constant $c < K$, and exercise as soon as the stock price S_t reaches the level c. The value of the American option then becomes equal to the value of the perpetual European option which pays $K - c$ as soon as the stock price S_t takes the value c. The constant c can be chosen by the holder and is therefore determined implicitly by the requirement that it should maximize the option's value.

To find the exercise boundary given by the constant c, we evaluate the value of the corresponding European option. This problem is similar to the one that was discussed in Example 4.1.1. The general solution to the time-homogeneous Black-Scholes equation is

$$\pi(S) = aS + bS^{-\gamma}, \quad \gamma = 2r/\sigma^2, \quad c \leq S < \infty$$

where this time a and b have to be chosen to meet the boundary conditions

$$\pi(c) = K - c, \quad \pi(S) \text{ bounded as } S \to \infty.$$

The second condition implies that $a = 0$. Therefore, the value of the European option as a function of S is always of the form $bS^{-\gamma}$, where the constant b depends on the unknown c.

If we consider the collection of curves of the form $bS^{-\gamma}$ for values of b that increase from 0, then we see that for low values of b the curve crosses the line $K - S$ twice until for some critical value of b it just touches this line; for higher values of b the two curves do not cross. Since there must be at least one value of S (namely $S = c$) such that $K - S = bS^{-\gamma}$, the high values of b do not qualify. Among all the curves that remain, the one that just touches the line $K - S$ is the one that produces the highest option values (for *all* values of S). Therefore, the values of b and c are determined by the two conditions

$$K - c = bc^{-\gamma} \tag{4.52a}$$

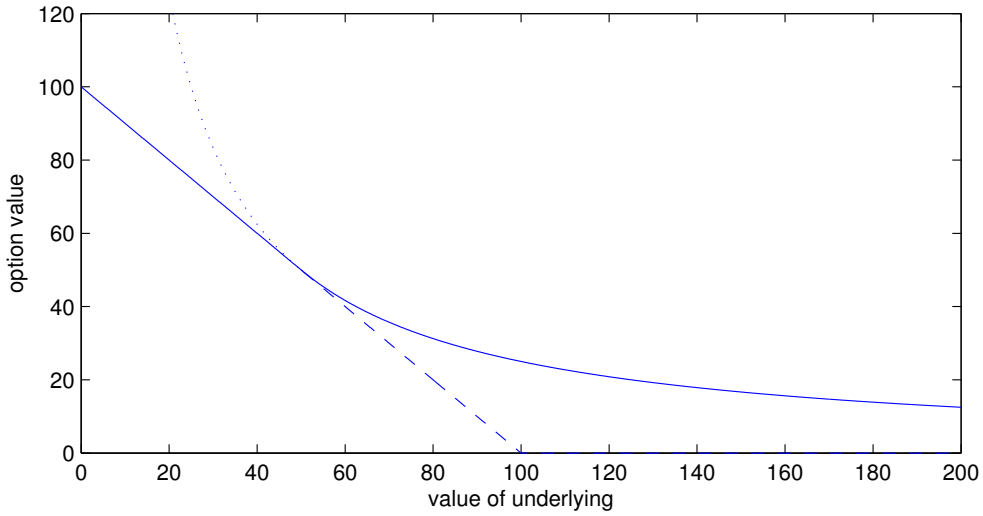

Figure 4.2: The drawn curve represents the value of a perpetual American put option, as a function of the value of the underlying. The parameter values are: strike $K = 100$; interest rate $r = 2\%$; volatility $\sigma = 0.2$. The dashed curve (partly covered by the drawn curve) indicates the value of immediate exercise. The dotted curve (also partly covered) represents a function of the form $bS^{-\gamma}$, with $\gamma = 2r/\sigma^2$.

$$-1 = -b\gamma c^{-\gamma-1}. \tag{4.52b}$$

This produces the following result:

$$c = \frac{r}{r + \frac{1}{2}\sigma^2} K. \tag{4.53}$$

So for $S \geq c$, the value of the American put option is given by

$$\pi(S) = \frac{\frac{1}{2}\sigma^2 K}{r + \frac{1}{2}\sigma^2} \left[\left(1 + \frac{\sigma^2}{2r}\right) \frac{S}{K} \right]^{-2r/\sigma^2} \qquad \text{for } S \geq \frac{r}{r + \frac{1}{2}\sigma^2} K \tag{4.54a}$$

whereas for $S \leq c$ the option is exercised immediately and so its value is

$$\pi(S) = K - S \qquad \text{for } S \leq \frac{r}{r + \frac{1}{2}\sigma^2} K. \tag{4.54b}$$

Figure 4.2 shows a graph of the value of a perpetual American put option, together with a graph of the value of immediate exercise. The exercise region is determined in such a way that the payoff function, as a function of the state variable, is not only continuous but also continuously differentiable; there is no "kink" at the exercise boundary. This is called the *smooth pasting condition*. This condition can be shown to hold also in case of non-perpetual contracts. The smooth pasting condition serves as an extra boundary condition that may be used to determine the solution in cases where other boundary conditions do not suffice to define the solution uniquely.

4.4.2 A defaultable perpetuity

A *perpetuity* is a contract that entitles the holder to a series of fixed annual payments that will go on forever.[4] Such a contract can only be credibly sold by an institution that is expected to exist eternally.[5] Perhaps more realistically, one can imagine a trust fund which is, at some point in time, equipped with a certain amount of capital, and which then pays a fixed amount each year to eligible beneficiaries, until the fund runs out of money.[6] The product that is defined in this way could be called a *defaultable perpetuity*. Suppose that the fund holds a fixed-mix portfolio, and assume that the Black-Scholes model applies with $r > 0$. For simplicity, consider a continuous payment stream of c units of currency per year, instead of payments at discrete times. The evolution of the fund's capital is then given by the SDE

$$dS_t = (rS_t - c)\,dt + \sigma S_t\,dW_t \tag{4.55}$$

where W_t is a Brownian motion under the risk-neutral measure $\mathbb{Q}$. Let us determine the value of this contract to the holder. As in the case of the barrier option discussed above, the value should be determined completely by S_t (which in this case is the current size of the capital in the trust fund) and should not depend on time. Therefore, we again arrive at an ODE, which now is given by (taking into account that the contract generates a dividend of size $c\Delta t$ in an interval of length Δt)

$$c + (rS - c)\,\frac{d\pi_C}{dS}(S) + \tfrac{1}{2}\sigma^2 S^2\,\frac{d^2\pi_C}{dS^2}(S) = r\pi_C(S). \tag{4.56}$$

Since this is a second-order ODE, two boundary conditions should be provided to determine a unique solution. One boundary condition is fairly obvious:

$$\pi_C(0) = 0.$$

To arrive at a second boundary condition, note that, when the fund's capital is large relative to the annual payment, the fund will not easily run out of money, so that the value of the stream of payments should be close to the value of a *guaranteed* annuity, which is

$$\int_0^\infty e^{-rt} c\,dt = \frac{c}{r}\,.$$

[4] Clearly, the holder may not be able to enjoy all of those payments herself. However, she might view the bond as an investment and at some point sell it, or she might bequest it.

[5] For instance, the government of the United Kingdom has issued such instruments (in 1751, and again in 1927) under the name of "consol bonds". These bonds carried a redemption provision, however. The last of the outstanding consol bonds were redeemed in 2015.

[6] The term *endowment fund* is also often used for funds that generate an ongoing income stream on the basis of an initially given capital.

Therefore, as a second boundary condition it is reasonable to impose

$$\lim_{S\to\infty} \pi_C(S) = \frac{c}{r}\,. \tag{4.57}$$

To solve the equation (4.56), it is useful to differentiate both sides with respect to S. To shorten the notation, write $\pi'_C(S)$ instead of $d\pi_C(s)/dS$, and likewise for the second derivative. Applying differentiation to (4.56) leads to

$$\big((r+\sigma^2)S - c\big)\pi''_C(S) + \tfrac{1}{2}\sigma^2 S^2 \pi'''_C(S) = 0 \tag{4.58}$$

(note that differentiation generates a term $r\pi'_C(S)$ both on the left and on the right, so that a cancellation occurs). If we can find functions $g(S)$ and $h(S)$ such that

$$h(S)g(S) = \tfrac{1}{2}\sigma^2 S^2, \qquad h(s)g'(S) = (r+\sigma^2)S - c \tag{4.59}$$

then solutions of (4.58) can be found by solving $\pi''_C(S) = a_1/g(S)$ where a_1 is an arbitrary constant; indeed, we then have $g'(S)\pi''_C(S) + g(S)\pi'''_C(S) = 0$, and multiplication by $h(S)$ shows that (4.58)is satisfied. To find a solution to (4.59), write

$$\frac{d}{dS}\log g(S) = \frac{g'(S)}{g(S)} = \frac{(r+\sigma^2)S - c}{\frac{1}{2}\sigma^2 S^2} = 2\left(\frac{r}{\sigma^2}+1\right)\frac{1}{S} - \frac{2c}{\sigma^2 S^2}$$

A solution of this is given by

$$\log g(S) = 2\Big(\frac{r}{\sigma^2}+1\Big)\log S + \frac{2c}{\sigma^2 S} \quad\Rightarrow\quad g(S) = S^{2r/\sigma^2+2}e^{2c/(\sigma^2 S)}.$$

Consequently, (4.58) is solved by

$$\pi_C(S) = a_1 f(S) + a_2 S + a_3 \tag{4.60}$$

where a_1, a_2, and a_3 are constants and the function $f(x)$ is a second integral of $1/g(x)$; specifically, we can take

$$f(x) = \int_0^x \int_0^u z^{-2r/\sigma^2-2}\, e^{-2c/(\sigma^2 z)}\, dz\, du. \tag{4.61}$$

The three constants are to be determined by the two boundary conditions and by the requirement that the solution should satisfy (4.56) rather than just the differentiated form (4.58). The boundary condition $\pi_C(0) = 0$ leads to the condition $a_3 = 0$. Because (4.58) is obtained from (4.56) by differentiation, any solution of (4.58) is such that

$$c + (rS - c)\pi'_C(S) + \tfrac{1}{2}\sigma^2 S^2 \pi''_C(S) - r\pi_C(S) = \kappa \tag{4.62}$$

where κ is a constant. A given solution of (4.58) is also a solution of (4.56) if and only if this constant is 0. To verify this, it is sufficient to evaluate the left hand side of (4.62) at any point, for instance $S = 0$; the condition to be fulfilled is therefore

$$c - c\pi_C'(0) - r\pi_C(0) = 0.$$

Since $\pi_C(0) = 0$ and $\pi_C'(0) = a_2$, the condition that we find is $a_2 = 1$. It remains to determine the constant a_1 on the basis of the second boundary condition. This takes a bit more work.

The double integral that appears in the function $f(x)$ can be simplified by means of integration by parts. If we define (writing $a = 2r/\sigma^2$, $b = 2c/\sigma^2$)

$$g(u) = \int_0^u z^{-a-2} e^{-b/z}\, dz$$

then

$$\begin{aligned} f(x) &= \int_0^x g(u)\, du = ug(u)\Big|_0^x - \int_0^x ug'(u)\, du = xg(x) - \int_0^x u \cdot u^{-a-2} e^{-b/u}\, du \\ &= x\int_0^x z^{-a-2} e^{-b/z}\, dz - \int_0^x z^{-a-1} e^{-b/z}\, dz. \end{aligned}$$

This is a representation in terms of single integrals. The integrals can be rewritten in terms of a standard special function by the substitution $y = b/z$, as follows:

$$\begin{aligned} \int_0^x z^{-a-2} e^{-b/z}\, dz &= \int_{b/x}^\infty b^{-a-2} y^{a+2} e^{-y} b y^{-2}\, dy = b^{-a-1} \int_{b/x}^\infty y^a e^{-y}\, dy \\ &= b^{-a-1}\Gamma(a+1, b/x) \end{aligned}$$

where Γ denotes the upper incomplete gamma function[7]

$$\Gamma(a, w) = \int_w^\infty t^{a-1} e^{-t}\, dt.$$

Likewise,

$$\int_0^x z^{-a-1} e^{-b/z}\, dz = b^{-a}\Gamma(a, b/x)$$

The expression (4.60) can therefore be rewritten, using $a_2 = 1$ and $a_3 = 0$, as

$$\pi_C(S) = a_4\big(S\Gamma(a+1, b/S) - b\Gamma(a, b/S)\big) + S$$

[7] When the upper incomplete gamma function is divided by the value $\Gamma(a) = \Gamma(a, 0)$ of the complete gamma function at a (a normalization which is in some sources, but not here, included in the definition), then the result is the decumulative distribution function of the Gamma distribution. In other words, when X is a variable that follows a Gamma distribution with shape parameter a and rate parameter 1, then the probability that X exceeds a given value $w \geq 0$ is equal to $\Gamma(a, w)/\Gamma(a)$.

$$= S\big(1 + a_4\Gamma(a+1, b/S)\big) - a_4 b\Gamma(a, b/S)$$

where the constant a_4 is still to be determined. It is seen from the expression above that the only choice for a_4 that has a chance of satisfying the boundary condition at infinity is $a_4 = -1/\Gamma(a+1)$; any other choice will lead to a limit value of $\pi_C(S)$ that is either ∞ or $-\infty$. To see that setting $a_4 = -1/\Gamma(a+1)$ indeed satisfies (4.57), note that

$$\lim_{S\to\infty} S\left(1 - \frac{\Gamma(a+1, b/S)}{\Gamma(a+1)}\right) = \lim_{x\downarrow 0} b\,\frac{1 - \Gamma(a+1, x)/\Gamma(a+1)}{x} = -\lim_{x\downarrow 0}\frac{bx^a e^{-x}}{\Gamma(a+1)} = 0$$

by L'Hôpital's rule,[8] and

$$\lim_{S\to\infty} b\,\frac{\Gamma(a, b/S)}{\Gamma(a+1)} = b\,\frac{\Gamma(a)}{\Gamma(a+1)} = \frac{b}{a} = \frac{c}{r}$$

by the standard fact (proved by integration by parts) that $\Gamma(a+1) = a\Gamma(a)$. In summary, the solution is as follows:

$$\pi_C(S) = S\left(1 - \frac{\Gamma\big(a+1, b/S)\big)}{\Gamma(a+1)}\right) + \frac{c}{r}\,\frac{\Gamma(a, b/S))}{\Gamma(a)} \tag{4.63}$$

where $a = 2r/\sigma^2$, $b = 2c/\sigma^2$.

A plot of $\pi_C(S)$ as a function of S is shown in Fig. 4.3. Also shown are the two upper bounds that must hold for the value of the defaultable perpetuity: the value cannot be higher than the value of the available capital, nor can it be higher than the value of a guaranteed perpetuity. For low values of capital (relative to the annual payout), the value of the perpetuity is approximately equal to the value of available capital. In this situation, it is likely (under the real-world measure, and even more so under the risk-neutral measure) that capital will run out at some point, i.e., in the language of actuarial science, ruin occurs. Therefore there is a high probability that all capital will be spent. On the other hand, if the initial capital is high, there are many scenarios in which ruin never occurs. This means that there is a substantial probability that not all capital will be spent, so that the value of the perpetuity is less than the value of available capital. The "law of conservation of value" does not apply in this case, because part of the initial capital is pushed towards infinity.

As is seen from the graph, the waste of capital is not negligible even in cases where the initial capital is in the range of 50 to 90 percent of what would be needed for a guaranteed perpetuity (as probably often happens in practice), especially when a high volatility is chosen. In a Black-Scholes world with fixed interest rates, trustees of a fund that aims to pay a fixed income stream will probably not maintain the fixed-

[8] Guillaume François Antoine, marquis de L'Hôpital (1661–1704), French mathematician.

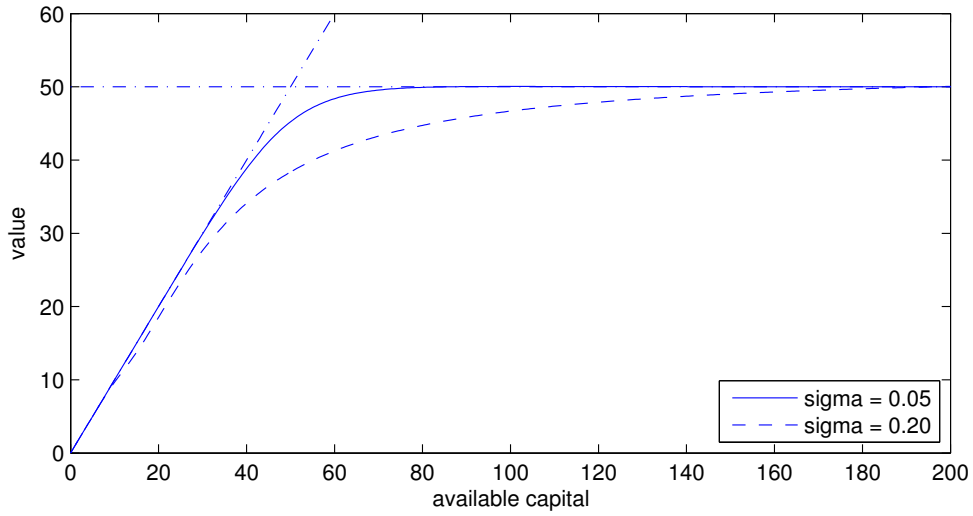

Figure 4.3: Value of a defaultable perpetuity, as a function of available capital. The annual payment rate is taken to be 1. A Black-Scholes model is assumed for the evolution of capital. Outcomes are shown for two values of the volatility, namely $\sigma = 0.05$ (drawn line) and $\sigma = 0.2$ (dashed line). Also shown are the two upper bounds that are given by the available capital and by the value of a guaranteed perpetuity (dash-dotted lines).

mix policy when the value of capital grows above the amount that is needed to pay a guaranteed perpetuity; in such a case, the downward risk cannot be compensated by the upward potential, because further capital gains, if they occur, will never be used. A fully efficient[9] policy could be designed as the replication policy that corresponds to the value function $\pi_C(S)$. This policy would gradually reduce the volatility of the investment portfolio when capital approaches the value of a guaranteed perpetuity.

4.4.3 The Vasicek model

Contracts that pay a fixed amount at a given time T are known as *default-free zero-coupon bonds*. The price at time t of a zero-coupon bond that matures at time T (written as $P_t(T)$) determines the *discount factor* that must be applied at time t to a payment that will be received at time T. This discount factor is usually expressed in terms of an *interest rate* (discretely compounded r_T^{d} or continuously compounded r_T^{c}) through

$$P_0(T) = d_T = \frac{1}{(1 + r_T^{\mathrm{d}})^T}\,, \qquad P_0(T) = d_T = \exp(-r_T^{\mathrm{c}} T).$$

In this way the prices of zero-coupon bonds for different maturities determine what is called the *term structure of interest rates*. The term structure reflects the "time

[9] "Efficient" is meant here in the sense that the initial capital is fully used, i.e. no funds are shifted towards infinity.

value of money"; it will be discussed more extensively in Chapter 5.

The Vasicek model is a model for the short rate process r_t. It was already used to illustrate the pricing kernel method in Example 4.1.3. The Vasicek bond pricing formula will be re-derived here in somewhat more detail and using the numéraire-dependent pricing formula rather than the pricing kernel. As a numéraire, we take the money market account, and we specify the model under the corresponding equivalent martingale measure $\mathbb{Q}_M$ in the form

$$dr_t = a(b - r_t)\,dt + \sigma\,dW_t \tag{4.64}$$

where W_t is a Brownian motion under $\mathbb{Q}$. Let the price at time $t \leq T$ of the contract that pays 1 at time T be denoted by $P_t(T)$. Apply the NDPF:

$$\frac{P_0(T)}{M_0} = E^{\mathbb{Q}_M}\frac{P_T(T)}{M_T} = E^{\mathbb{Q}_M}\frac{1}{M_T}.$$

We can set $M_0 = 1$.

For convenience of notation, drop the superscript $\mathbb{Q}_M$ now. The problem comes down to: determine $E[1/M_T]$ when M_T is given by

$$\begin{aligned} dM_t &= r_t M_t\,dt \\ dr_t &= a(b - r_t)\,dt + \sigma\,dW_t. \end{aligned}$$

with $M_0 = 1$. Write $m_t := \log M_t$ and note that $dm_t = r_t\,dt$. The problem therefore is: determine $E[\exp(-m_T)]$ when m_T is given by

$$\begin{aligned} dm_t &= r_t\,dt \\ dr_t &= a(b - r_t)\,dt + \sigma\,dW_t \end{aligned}$$

with $m_0 = 0$. The equations can be written in vector form as

$$d\begin{bmatrix} r_t \\ m_t \end{bmatrix} = \begin{bmatrix} -a & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} r_t \\ m_t \end{bmatrix}dt + \begin{bmatrix} ab \\ 0 \end{bmatrix}dt + \begin{bmatrix} \sigma \\ 0 \end{bmatrix}dW_t. \tag{4.65}$$

These are linear equations. Therefore the distribution of m_T is *normal*, and the distribution of $\exp(-m_T)$ is *lognormal*. We have

$$E[\exp(-m_T)] = \exp\bigl(-E[m_T] + \tfrac{1}{2}\operatorname{var}(m_T)\bigr).$$

So we must determine $E[m_T]$ and $\operatorname{var}(m_T)$. To solve the vector differential equation

(4.65), one can for instance use the "left eigenvector method". Note that

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} -a & 0 \\ 1 & 0 \end{bmatrix} = -a \cdot \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & a \end{bmatrix} \begin{bmatrix} -a & 0 \\ 1 & 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 & a \end{bmatrix}.$$

Therefore we can write *scalar* SDEs for r_t (as already known) and for $r_t + am_t$.

$$\begin{aligned} dr_t &= -ar_t\,dt + ab\,dt + \sigma\,dW_t \\ &\Rightarrow \quad r_T = e^{-aT}r_0 + \textstyle\int_0^T e^{-a(T-t)}ab\,dt + \int_0^T e^{-a(T-t)}\sigma\,dW_t \end{aligned}$$

$$\begin{aligned} d(r_t + am_t) &= ab\,dt + \sigma\,dW_t \\ &\Rightarrow \quad r_T + am_T = r_0 + am_0 + abT + \textstyle\int_0^T \sigma\,dW_t. \end{aligned}$$

We find (note $m_0 = 0$)

$$m_T = \frac{1 - e^{-aT}}{a}(r_0 - b) + bT + \frac{\sigma}{a}\int_0^T \left(1 - e^{a(T-t)}\right) dW_t.$$

Using the property "you can't beat the system" (Thm. 2.2.2) and the rule (2.49) for the variance of a stochastic integral with deterministic integrand, one finds

$$\begin{aligned} E[m_T] &= \left[b + \frac{1 - e^{-aT}}{aT}(r_0 - b) \right] T \\ \operatorname{var}(m_T) &= \frac{\sigma^2}{a^2}\int_0^T \left(1 - e^{a(T-t)}\right)^2 dt \\ &= \frac{\sigma^2}{a^2}\left[1 - 2\,\frac{1 - e^{-aT}}{aT} + \frac{1 - e^{-2aT}}{2aT} \right] T. \end{aligned}$$

Consequently, the price at time 0 of the zero-coupon bond that pays 1 at time T is given in the Vasicek model by

$$P_0(T) = \exp\left(-\left[b + \frac{1 - e^{-aT}}{aT}(r_0 - b) - \frac{\sigma^2}{2a^2}\left(1 - 2\,\frac{1 - e^{-aT}}{aT} + \frac{1 - e^{-2aT}}{2aT} \right) \right] T \right) \tag{4.66}$$

where it should be recalled that the parameter b is the one that appears in the model under $\mathbb{Q}$; consequently, this parameter must be interpreted as a risk-adjusted mean reversion level of the short rate. As a soundness check, one can verify that the formula above reduces to $P_0(T) = \exp(-r_0 T)$ if $\sigma = 0$ and $r_0 = b$. Since T represents time to maturity in the derivation above, the discount factor $P_t(T)$ from a general time $t \le T$ to T under the Vasicek model is given by the same formula, with r_0 replaced by r_t and T by $T - t$.

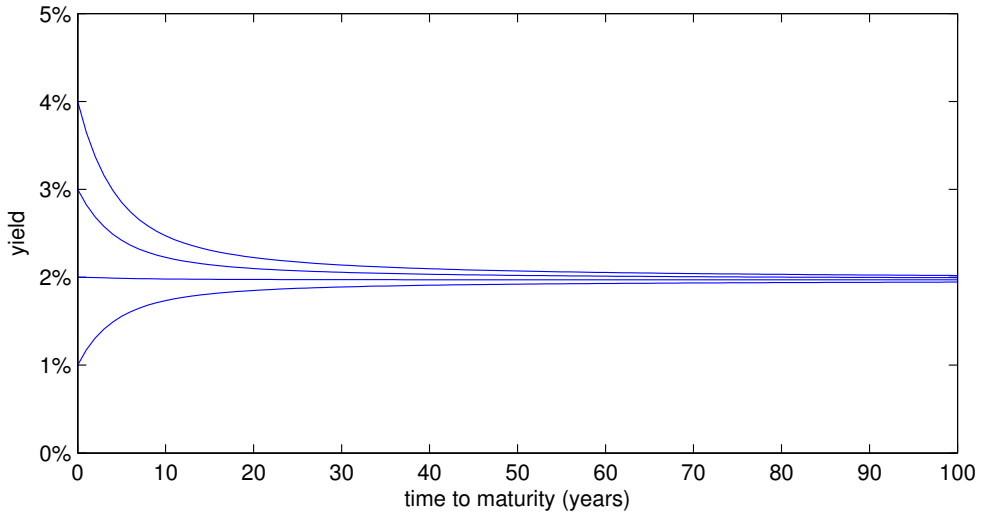

Figure 4.4: Yield curves according to the Vasicek model with parameter values $a = 0.4$, $b = 0.02$, $\sigma = 0.01$. The yield curves are shown for several values of the current short rate.

The term structure of interest rates is usually not expressed directly in terms of bond prices, but rather in terms of the corresponding interest rates. From (4.66), it is seen that the continuously compounded interest rate for maturity T according to the Vasicek model is given by

$$R_0(T) = b + \frac{1 - e^{-aT}}{aT}(r_0 - b) - \frac{\sigma^2}{2a^2}\left(1 - 2\,\frac{1 - e^{-aT}}{aT} + \frac{1 - e^{-2aT}}{2aT}\right). \qquad (4.67)$$

A plot of the function of T that is defined in this way (the "yield curve") is shown in Fig. 4.4. In fact, several curves are shown, corresponding to different values of the short rate r_0.

4.4.4 Put option in Black-Scholes-Vasicek model

The next example concerns a situation with two sources of risk. The solution method that is shown here (as usual, it is not the only possible method) uses a technique similar to that of Section 4.2.5: split the payoff into two parts that are handled by different numéraires. In this way, the example also serves to illustrate the change-of-numéraire technique in a situation with multiple driving Brownian motions.

The Black-Scholes-Vasicek model, written under the risk-neutral measure $\mathbb{Q}_M$, takes the following form:

$$\begin{aligned} dS_t &= r_t S_t\,dt + \sigma_S S_t\,dW_{S,t} \\ dM_t &= r_t M_t\,dt \\ dr_t &= a(b - r_t)\,dt + \sigma_r\,dW_{r,t}. \end{aligned}$$

The Brownian motions $W_{S,t}$ and $W_{r,t}$ are assumed to be independent (for simplicity). Our purpose in this section will be to derive a pricing formula for a put option with S_t as the underlying. The put pricing problem is relevant for an institution that seeks nominal downward protection for its investments in the face of variable interest rates. The institution might buy put options, but it could also choose to create a put option synthetically by following a *replication* strategy.

We can use the splitting method:

$$\max(K - S_T, 0) = \mathbb{1}_{\{S_T \leq K\}} K - \mathbb{1}_{\{S_T \leq K\}} S_T.$$

When the money market account is used as a numéraire for the first part and the stock as a numéraire for the second part, then the put value C_0 at time 0 is written as

$$C_0 = M_0 E^{\mathbb{Q}_M} \left[\frac{\mathbb{1}_{\{S_T \leq K\}} K}{M_T} \right] - S_0 E^{\mathbb{Q}_S} \left[\mathbb{1}_{\{S_T \leq K\}} \right].$$

Unfortunately, calculation of the expectation on the left requires the joint distribution of S_T and M_T. This complication can be avoided by using, instead of the money market account, another numéraire: the zero-coupon bond that matures at time T. The corresponding EMM is called the *forward measure*. It is denoted by $\mathbb{Q}_T$.

The numéraire-dependent pricing formula for a general contract C under the forward measure is as follows:

$$\frac{C_t}{P_t(T)} = E_t^{\mathbb{Q}_T} \left[\frac{C_T}{P_T(T)} \right] = E_t^{\mathbb{Q}_T} [C_T]$$

because $P_T(T) = 1$. This leads to the pricing formula

$$C_t = P_t(T) E_t^{\mathbb{Q}_T} [C_T].$$

This expression is of the form "first take expectation, then apply the discount factor that corresponds to the time of payment". It should be noted that expectation is taken here under $\mathbb{Q}_T$.

Returning now to the put valuation problem, the expression for the put value that is obtained from the splitting method with the T-bond and the stock as numéraires is

$$C_0 = P_0(T) K \mathbb{Q}_T(S_T \leq K) - S_0 \mathbb{Q}_S(S_T \leq K). \tag{4.68}$$

To calculate the probability of the event $S_T \leq K$ under $\mathbb{Q}_T$, we rewrite the model under this measure. For that purpose we need σ_T, the volatility of the bond with maturity T. We already computed the price of the T-bond in the Vasicek model.

The pricing function is of the form

$$\pi_T(t, r_t) = \exp\Big(-\frac{1 - e^{-a(T-t)}}{a}\, r_t + f_T(t)\Big)$$

where $f_T(t)$ depends on t only. Let σ_T denote the corresponding volatility. Note that σ_T is a 1×2 vector, since there are two Brownian motions in the model. We have

$$\sigma_T = \begin{bmatrix} \frac{\partial \pi_T}{\partial S} & \frac{\partial \pi_T}{\partial r} \end{bmatrix} \begin{bmatrix} \sigma_S S & 0 \\ 0 & \sigma_r \end{bmatrix} =: \begin{bmatrix} \sigma_{T,S} & \sigma_{T,r} \end{bmatrix}.$$

The quantities that we need can be readily computed:

$$\frac{\partial \pi_T}{\partial S} = 0, \qquad \frac{1}{\pi_T}\frac{\partial \pi_T}{\partial r} = \frac{\partial(\log\, \pi_T)}{\partial r} = -\frac{1 - e^{-a(T-t)}}{a}\,.$$

Therefore

$$\frac{\sigma_T^\top}{\pi_T} = \frac{1}{\pi_T}\begin{bmatrix} \sigma_{T,S} \\ \sigma_{T,r} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\sigma_r}{a}\Big(1 - e^{-a(T-t)}\Big) \end{bmatrix}.$$

With this, we can rewrite the model under the forward measure $\mathbb{Q}_T$; use the formula $\mu_X^T = \mu_X + \sigma_X \sigma_T^\top / \pi_T$. In the case of the BSV model:

$$\mu_S^T = r_t S_t + \begin{bmatrix} \sigma S_t & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{\sigma_r}{a}\Big(1 - e^{-a(T-t)}\Big) \end{bmatrix} = r_t S_t$$

$$\mu_r^T = a(b - r_t) + \begin{bmatrix} 0 & \sigma_r \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{\sigma_r}{a}\big(1 - e^{-a(T-t)}\big) \end{bmatrix} = a(b - r_t) - \frac{\sigma_r^2}{a}\Big(1 - e^{-a(T-t)}\Big).$$

The BSV model under $\mathbb{Q}_T$ is:

$$\begin{aligned} dS_t &= r_t S_t\, dt + \sigma_S S_t\, dW_{S,t}^T \\ dM_t &= r_t M_t\, dt \\ dr_t &= a\big(b(t) - r_t\big)\, dt + \sigma_r\, dW_{r,t}^T \end{aligned}$$

where

$$b(t) = b - \frac{\sigma_r^2}{a^2}\Big(1 - e^{-a(T-t)}\Big).$$

We can write

$$d\begin{bmatrix} \log\, S_t \\ r_t \end{bmatrix} = \left(\begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix}\begin{bmatrix} \log\, S_t \\ r_t \end{bmatrix} + \begin{bmatrix} \frac{1}{2}\sigma_S^2 \\ ab(t) \end{bmatrix}\right) dt + \begin{bmatrix} \sigma_S & 0 \\ 0 & \sigma_r \end{bmatrix} d\begin{bmatrix} W_{S,t}^T \\ W_{r,t}^T \end{bmatrix}.$$

This shows that the vector consisting of $\log S_t$ and r_t satisfies a linear SDE, as

discussed in Section 2.6.3. It follows that $\log S_t$ and r_t are, at any time t, jointly normally distributed. Expectations and variances can be computed as indicated in Section 2.6.3, and subsequently the quantity $\mathbb{Q}_T(S_T \leq K)$, which will be needed in the expression (4.68) for the put value, can be obtained.

Before entering into the computation of the expectation and variance of $\log S_T$, let us determine the form of the BSV model under $\mathbb{Q}_S$, which will be needed as well. Using the standard formulas for change of numéraire, we find

$$\mu_S^S = \mu_S^M + [\sigma_S \quad 0]\,\frac{\sigma_S^\top}{\pi_S} = (r_t + \sigma_S^2)S_t$$
$$\mu_r^S = \mu_r^M + [0 \quad \sigma_r]\,\frac{\sigma_S^\top}{\pi_S} = a(b - r_t).$$

The BSV model under $\mathbb{Q}_S$ is

$$\begin{aligned} dS_t &= (r_t + \sigma_S^2)S_t\,dt + \sigma_S S_t\,dW_{S,t}^S \\ dM_t &= r_t M_t\,dt \\ dr_t &= a(b - r_t)\,dt + \sigma_r\,dW_{r,t}^S. \end{aligned}$$

The joint SDE for $\log S_t$ and r_t is

$$d\begin{bmatrix}\log S_t \\ r_t\end{bmatrix} = \left(\begin{bmatrix}0 & 1 \\ 0 & -a\end{bmatrix}\begin{bmatrix}\log S_t \\ r_t\end{bmatrix} + \begin{bmatrix}\frac{1}{2}\sigma_S^2 \\ ab\end{bmatrix}\right)dt + \begin{bmatrix}\sigma_S & 0 \\ 0 & \sigma_r\end{bmatrix} d\begin{bmatrix}W_{S,t}^T \\ W_{r,t}^T\end{bmatrix}$$

so that, also in this case, we find a linear vector SDE. Therefore, expectation and variance of $\log S_T$ can again be computed from the formulas of Section (2.6.3).

Now, we proceed to the calculation of expectation and variance of $\log S_T$ both under $\mathbb{Q}_S$ and under $\mathbb{Q}_T$. Consider first $\mathbb{Q}_S$. The differential equation for the expectations of $\log S_T$ and r_t is given by (2.81), and the joint variance-covariance matrix of $\log S_t$ and r_t is described by the differential equation (2.84). Write $m_1(t) = E^{\mathbb{Q}_S}[\log S_t]$, $m_2(t) = E^{\mathbb{Q}_S} r_t$. These quantities are subject to the following differential equations:

$$\frac{dm_1}{dt}(t) = m_2(t) + \tfrac{1}{2}\sigma_S^2, \qquad \frac{dm_2}{dt}(t) = -am_2(t) + ab.$$

These equations can be solved successively, starting with m_2:

$$m_2(t) = e^{-at}m_2(0) + (1 - e^{-at})b = e^{-at}(m_2(0) - b) + b$$
$$m_1(t) = m_1(0) + \frac{1 - e^{-at}}{a}\,(m_2(0) - b) + (b + \tfrac{1}{2}\sigma_S^2)t. \tag{4.69}$$

As a variation of the method used in Section 4.4.3 to determine variances, we can

use the matrix differential equation (2.84). The equation becomes in the present case (time arguments are suppressed to alleviate the notation)

$$\frac{d}{dt}\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix}\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} + \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & -a \end{bmatrix} + \begin{bmatrix} \sigma_S^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix}$$

where it should be noted that $h_{12}(t) = h_{21}(t)$ because of the symmetry of the variance-covariance matrix. Writing this in coordinates, one obtains:

$$\begin{aligned}
\frac{dh_{11}}{dt}(t) &= h_{21}(t) + h_{12}(t) + \sigma_S^2 = 2h_{12}(t) + \sigma_S^2 \\
\frac{dh_{12}}{dt}(t) &= h_{22}(t) - ah_{12}(t) \\
\frac{dh_{22}}{dt}(t) &= -2ah_{22}(t) + \sigma_r^2.
\end{aligned}$$

Since $\log S_0$ and r_0 are given, we have $H(0) = 0$, so that the initial conditions for all variables are 0. Again the equations can be solved consecutively.

$$\begin{aligned}
h_{22}(t) &= \frac{1 - e^{-2at}}{2a}\,\sigma_r^2 = \frac{\sigma_r^2}{2a} - \frac{\sigma_r^2}{2a}\,e^{-2at} \\
h_{12}(t) &= \frac{\sigma_r^2}{2a}\,\frac{1 - e^{-at}}{a} - \frac{\sigma_r^2}{2a}\int_0^t e^{-a(t-s)}e^{-2as}\,ds \\
&= \frac{\sigma_r^2}{2a^2}\,(1 - e^{-at}) - \frac{\sigma_r^2}{2a}\,e^{-at}\int_0^t e^{-as}\,ds \\
&= \frac{\sigma_r^2}{2a^2}\,(1 - e^{-at}) - \frac{\sigma_r^2}{2a}\,e^{-at}\,\frac{1 - e^{-at}}{a}\,ds \\
&= \frac{\sigma_r^2}{2a^2}\big(1 - 2e^{-at} + e^{-2at}\big) \\
h_{11}(t) &= \frac{\sigma_r^2}{a^2}\left(t - 2\,\frac{1 - e^{-at}}{a} + \frac{1 - e^{-2at}}{2a}\right) + \sigma_S^2 t \\
&= \Big(\frac{\sigma_r^2}{a^2} + \sigma_S^2\Big)t - \frac{\sigma_r^2}{a^3}\,\big(\tfrac{3}{2} - 2e^{-at} + \tfrac{1}{2}e^{-2at}\big).
\end{aligned} \tag{4.70}$$

The differential equations for the expectations of $\log S_t$ and r_t under Q_T are slightly more involved, due to the presence of a more complicated forcing term:

$$\begin{aligned}
\frac{dm_1}{dt}(t) &= m_2(t) - \tfrac{1}{2}\sigma_S^2 \\
\frac{dm_2}{dt}(t) &= -am_2(t) + a\Big(b - \frac{\sigma_r^2}{a^2}\Big(1 - e^{-a(T-t)}\Big)\Big)
\end{aligned}$$

where $m_1(t)$ and $m_2(t)$ now refer to expectations under $\mathbb{Q}_T$. The equations can still

be solved consecutively:

$$\begin{aligned} m_2(t) &= e^{-at}m_2(0) + a\Big(b - \frac{\sigma_r^2}{a^2}\Big)\frac{1-e^{-at}}{a} + \frac{\sigma_r^2}{a}\int_0^t e^{-a(t-s)}e^{-a(T-s)}\,ds \\ &= e^{-at}m_2(0) + \Big(b - \frac{\sigma_r^2}{a^2}\Big)(1-e^{-at}) + \frac{\sigma_r^2}{a}e^{-a(T+t)}\,\frac{e^{2at}-1}{2a} \\ &= b - \frac{\sigma_r^2}{a^2} + \Big(m_2(0) - b + \frac{\sigma_r^2}{a^2}\Big)e^{-at} + \frac{\sigma_r^2}{2a^2}\,e^{-aT}(e^{at}-e^{-at}) \\ m_1(t) &= m_1(0) + \Big(b - \tfrac{1}{2}\sigma_S^2 - \frac{\sigma_r^2}{a^2}\Big)t + \Big(m_2(0) - b + \frac{\sigma_r^2}{a^2}\Big)\frac{1-e^{-at}}{a} \\ &\qquad + \frac{\sigma_r^2}{2a^2}\,e^{-aT}\left(\frac{e^{at}-1}{a} - \frac{1-e^{-at}}{a}\right). \end{aligned} \tag{4.71}$$

The variance of $\log S_t$ under $\mathbb{Q}_T$ is the same as under $\mathbb{Q}_S$. We therefore find from (4.70)

$$\mathrm{var}^{\mathbb{Q}_S}(S_T) = \mathrm{var}^{\mathbb{Q}_T}(S_T) = \sigma^2 T$$

where σ^2 is defined by

$$\begin{aligned} \sigma^2 &= \sigma_S^2 + \frac{\sigma_r^2}{a^2}\left(1 - \frac{\frac{3}{2} - 2e^{-aT} + \frac{1}{2}e^{-2aT}}{aT}\right) \\ &= \sigma_S^2 + \frac{\sigma_r^2}{a^2}\left(1 - 2\,\frac{1-e^{-aT}}{aT} + \frac{1-e^{-2aT}}{2aT}\right). \end{aligned} \tag{4.72}$$

Recall the expression in (4.67) for the interest rate $R_0(T)$ that corresponds to maturity T according to the Vasicek model. From (4.69) and (4.71), the following expressions are obtained:

$$\begin{aligned} E^{\mathbb{Q}_S}[\log S_T] &= \log S_0 + (b + \tfrac{1}{2}\sigma_S^2)T + (r_0 - b)\,\frac{1-e^{-aT}}{a} \\ &= \log S_0 + \big(R_0(T) + \tfrac{1}{2}\sigma_1^2\big)T \end{aligned} \tag{4.73}$$

$$\begin{aligned} E^{\mathbb{Q}_T}[\log S_T] &= \log S_0 + \Big(b - \tfrac{1}{2}\sigma_S^2 - \frac{\sigma_r^2}{a^2}\Big)T + \Big(m_2(0) - b + \frac{\sigma_r^2}{a^2}\Big)\,\frac{1-e^{-aT}}{a} \\ &\qquad + \frac{\sigma_r^2}{2a^3}(1 - 2e^{-aT} + e^{-2aT}) \\ &= \log S_0 + \big(R_0(T) - \tfrac{1}{2}\sigma^2\big)T \end{aligned} \tag{4.74}$$

Consequently, if the parameter r is set equal $R_0(T)$, and σ is defined by (4.72), then the time-0 value of the put option in the Black-Scholes-Vasicek model can be written in the same way as the put option value in the BS model, namely

$$C_0 = -S_0\Phi(-d_1) + e^{-rT}K\Phi(-d_2) \tag{4.75a}$$

where the parameters d_1 and d_2 are defined by the familiar formulas

$$d_1 = \frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \quad d_2 = \frac{\log(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\,. \tag{4.75b}$$

In the BSV model, we find in this way that the interest rate with respect to the time of maturity T of the put option should be taken, whereas there is no such indication in the BS model since that model assumes that the interest rate is the same for all maturities. Moreover, the formula uses a volatility that is higher than the volatility of the underlying asset. This is natural, since the variability of the interest rate introduces an additional uncertainty into the model. The BSV put pricing formula therefore provides a partial explanation for the fact that the implied volatility (i.e. the volatility that is backed out from observed option prices) is often higher than the historical volatility that is obtained from time series analysis of the price of the underlying asset. The correction term in (4.72) is horizon-dependent and becomes more important for longer maturities, but for typical parameter values it is small relative to the Black-Scholes volatility σ_S^2. The numerical impact of the volatility correction relating to interest rate variability is therefore usually small; it is much more important to use the correct value of the interest rate.

4.5 Exercises

1. Re-derive the price of a digital option in the Black-Scholes model (see (4.10) using the numéraire-dependent pricing formula with S as the numéraire, rather than B.

2. Obtain the price of a digital option in the Black-Scholes model by means of the pricing kernel method. The computations may be simplified by noting that the solution should not depend on the parameter μ in the Black-Scholes model (since this parameter doesn't appear in the corresponding Black-Scholes equation), so that μ can be assigned any convenient value.

3. The following market was considered in Exc. 3.8:

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sigma S_t\,dW_{1,t} \\ dF_t &= \mu_1 F_t\,dt + \sigma_1 F_t\,dW_{1,t} + \sigma_2 F_t\,dW_{2,t} \\ dB_t &= r B_t\,dt. \end{aligned}$$

All three variables represent prices of traded assets; S_t is a stock index, F_t is the share price of an exchange-traded investment fund, and B_t is a bond. The parameters μ, σ, μ_1, σ_1, σ_2, and r are positive constants. The Brownian motions $W_{1,t}$ and $W_{2,t}$ are independent. It has been shown in Exc. 3.8 that the above market is arbitrage-free

and complete. Investors who buy shares in the fund may be concerned that the return of the fund may lag behind the stock market index. Therefore they may be interested in a contract that will pay to the holder at time T the difference between the stock index value and the fund share value, if that difference is positive. In other words, the payoff of the contract is $\max(S_T, F_T)$. In order to determine the value of this contract, it is convenient to work with S_t as a numéraire. The associated equivalent martingale measure is denoted by $\mathbb{Q}_S$.

a. Rewrite the model in a form in which the driving processes are Brownian motions under $\mathbb{Q}_S$.

b. Determine the value at time 0 of the contract with payoff $\max(S_T, F_T)$. As a special case, determine the value of the contract when the fund share value and the index value are initially equal, i.e. $F_0 = S_0$.

4. Re-derive formula (4.66) for the price of a bond in the Vasicek model by solving the partial differential equation (4.15) with the boundary condition $\pi_T(T, x) = 1$, as follows. Try a solution of the form $\pi_T(t, x) = \exp(f(t) + g(t)x)$ for suitable functions f and g. Use the fact that the resulting equation must hold for all x, and solve the resulting ordinary differential equations for f and g with the appropriate boundary conditions.

5. Prove that the Black-Scholes model extended with (4.51) is complete and free of arbitrage. Write down the pricing equation for options whose payoff is determined by A_T.

6. A *continuously sampled geometric Asian option* is a contract whose payoff depends on the geometric average of the underlying asset during a certain period. The continuous geometric average is defined by

$$A_T = S_0 \exp\left(\frac{1}{T}\int_0^T \log\frac{S_t}{S_0}\,dt\right). \tag{4.76}$$

a. Show that the valuation of geometric Asian options becomes possible in the Black-Scholes model after addition of one extra state variable with suitable dynamics. Prove that the resulting model is complete and free of arbitrage.

b. Write down the pricing equation for geometric Asian options that follows from the model constructed in part a.

c. Give a formula for the value of a digital geometric Asian option.

d. Give a formula for the value of a geometric Asian call option, and compare the result to the Black-Scholes formula for the value of a European call option.

7. In the Black-Scholes model, a general formula for the price at time 0 of a contract

that pays $f(S_T)$ at time T, where f is a given payoff function, is given by (4.47):

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}z^2) f\big(S_0 \exp((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}z)\big)\, dz. \tag{4.77}$$

In cases where the integral cannot be evaluated analytically, we may still evaluate it numerically by making use of the approximation formula[10]

$$C_0 \approx \frac{e^{-rT}}{\sqrt{2\pi}} \sum_{k=1}^{N} \exp(-\tfrac{1}{2}z_k^2) f\big(S_0 \exp((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}z_k)\big)\, \Delta z \tag{4.78}$$

where the z_k's form a uniform grid with step size Δz. To make the approximation accurate, the grid should be sufficiently wide and the step size should be sufficiently small.

In the calculations below, assume the following parameter values: $r = 0.04$, $\sigma = 0.2$, $S_0 = 100$.

a. Consider first a standard call option with time of maturity $T = 1$ and strike $K = 100$. Plot the integrand (i.e. the function that appears under the integral sign) in (4.77) as a function of z. Determine a suitable grid width and a suitable step size, and compute the integral by the approximation (4.78). Compare to the exact value as obtained from the Black-Scholes formula. Which grid width and step size do you need to get an approximation error of less than one cent?

Now consider the pricing of a compound option. Specifically, consider a contract that gives the holder the right, but not the obligation, to buy at time T_1 at the price K_1 a put option on the underlying asset S which expires at time $T_2 > T_1$ and which has strike K_2 (a "call on a put"). Note that the value of the contract at time T_1 is $\max(P - K_1, 0)$ where P is the value of a put with strike K_2 whose underlying asset value is S_{T_1} and which has $T_2 - T_1$ units of time to maturity. Take $T_1 = 1$, $T_2 = 2$, $K_1 = 5$, $K_2 = 100$.

b. Write down the value of the compound option in the form (4.77). Plot the integrand to determine a suitable grid width. Compute approximations of the value of the option by the formula (4.78) using several different step sizes, and determine the value of the option up to an accuracy of one cent. (You may judge the convergence "by eye".)

8. Consider the pricing of an option with payoff $f(S_T)$ within the Black-Scholes model. Instead of using a numerical integration method based on discretization, as

[10]The formula (4.78) is a simple example of a *numerical integration scheme*. The numerical evaluation of single integrals is also known as *quadrature*, because it comes down to computing the area of a certain region in a plane. Numerical techniques for multiple integrals go by the name of *cubature* methods.

in (4.78), one might also use a method based on function approximation. Introduce a function $F(z)$ by

$$F(z) = f(S_0 \exp\big((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\, z). \tag{4.79}$$

The option price is then given by $C_0 = e^{-rT}E[F(Z)]$ with $Z \sim N(0,1)$. If one has an approximation of the form

$$F(z) \approx \sum_{i=1}^{n} c_i \phi_i(z) \tag{4.80}$$

where the functions ϕ_i are such that $E[\phi_i(Z)]$ is known analytically, then an approximate value of the option price is given by

$$C_0 \approx e^{-rT} \sum_{i=1}^{n} c_i E[\phi_i(Z)]. \tag{4.81}$$

a. Consider the same situation as in Exc. 7.a. Use the following basis functions: $\phi_1(z) = 1$, $\phi_2(z) = \sin(az)$, $\phi_3(z) = \cos(az)$, $\phi_4(z) = \sin(2az)$, $\phi_5(z) = \cos(2az)$, where a is a constant that is still to be selected. Find an approximation to the function F defined in (4.79) in terms of these basis functions by taking a grid $\{z_1, \ldots, z_m\}$ and regressing the vector of function values $[F(z_1) \cdots F(z_m)]$ on the corresponding vectors $[\phi_i(z_1) \cdots \phi_i(z_m)]$ $(i = 1, \ldots, 5)$ formed from the basis functions. A fine grid is not needed for this purpose; take for instance a grid from -4 to 4 with steps of size 1. Plot the function F as well as its approximation on a wider grid than you used in the regression, to see the quality of the approximation. Experiment to find a value of a that generates a good approximation. Then compute an approximate option value from the formula (4.81), using the known values of $E[\phi_i(Z)]$ (see (2.99)). Compare to the exact value as found from the Black-Scholes formula.

b. Rewrite the function $F(z)$ in (4.79) as $F(z; S_0)$ to indicate the parameter S_0 explicitly. Verify that, for $\eta > 0$, we have

$$F(z; \eta S_0) = F\big(z + (\log \eta)/\big(\sigma\sqrt{T}\,\big), S_0\big).$$

Consequently, (4.80) implies

$$F(z; \eta S_0) \approx \sum_{i=1}^{n} c_i \phi_i\big(z + (\log \eta)/\big(\sigma\sqrt{T}\,\big)\big).$$

Use the standard formulas

$$\sin(x+y) = \sin x\, \cos y + \cos x\, \sin y, \quad \cos(x+y) = \cos x\, \cos y - \sin x\, \sin y$$

to show that the approximation coefficients found in part a. can be used to find approximate option values not only for S_0 as given, but in fact for all possible initial values of the underlying asset. Plot a graph of the approximate option values as a function of the initial value of the underlying, and compare to the exact solution.[11]

9. Write down the concrete form of the variational inequality (3.68) for the case of the perpetual American put in the standard Black-Scholes model, and verify that the function (4.54) is the unique continuously differentiable solution. Provide an interpretation of the result that you get by applying the Black-Scholes differential operator to function (4.54) in the exercise region. [*Hint*: consider the delta hedging position taken by the writer of the option when the price of the underlying is in the exercise region.]

10. Is the second derivative of the function (4.54) a continuous function? Explain the answer on the basis of what you found in the previous exercise.

11. Assume the standard Black-Scholes model, and consider the perpetual American option with payoff given by

$$F(S) = \begin{cases} 1 \text{ if } S \geq U \text{ or } S \leq L \\ 0 \text{ if } L < S < U \end{cases}$$

where L and U are given constants with $L < U$. Determine the value of the contract by direct reasoning (cf. Example 4.1.1). Does the smooth pasting principle hold? Is this option a "true American"?

12. A *straddle* is a contract with payoff function $F(S) = |S - K|$. Determine the value of a perpetual American straddle under the standard Black-Scholes assumptions.

13. The payoff function of a straddle, $F(S) = |S - K|$, is the sum of the payoffs of a put and a call, both with strike K. Does it follow that the value of an American straddle is the sum of the value of the American call and the American put with the same strike price?

[11] A similar extension from a single value of the underlying to all values of the underlying is possible when exponentials are used as basis functions, or all polynomials up to a certain degree. The key property is that the space of basis functions should be closed under the shift operators defined by $(T_h\phi)(x) = \phi(x + h)$, or equivalently, that the space of linear combinations of basis functions should be closed under differentiation.

Chapter 5

The term structure of interest rates

5.1 Term structure products

One of the principal roles of financial markets is that they allow cashflows to be shifted over time. Lending agreements are presumably the oldest financial contracts, and today the markets related to interest rates are still the largest among all financial markets. Economic activity is greatly facilitated by the ability to borrow.

Interest rates depend on time in two different ways. First of all, there is a dependence on the *time to maturity*. Short-term interest rates are typically different from (usually lower than) long-term interest rates. This fact is expressed by saying that interest rates have a *term structure*. The dependence of interest rates on the time to maturity can be expressed in several ways, as discussed in the following section. In addition to the dependence on the time to maturity, there is also a dependence on *calendar time*. This is analogous to the dependence of, for instance, stock prices on calendar time. While in the case of a stock price we have just a single quantity that depends on time, in the case of interest rates there are many quantities involved which do not move completely in step but which are also not completely independent. If today the three-month rate is higher than it was yesterday, then presumably the six-month rate is also up, but perhaps not to the same extent, and it may well happen that the twenty-year rate has actually gone down. Modeling the evolution of the term structure of interest rates is therefore more complicated than modeling the evolution of, say, a stock price.

The simplest term structure product is the *zero-coupon bond*. This is just an agreement in which party A pays a given amount V_1 to party B at time 0 (now), and party B pays to party A a given amount V_2 at time T (the time of maturity). When party A is an individual consumer and party B is a bank, the zero-coupon bond is also known as a *deposit*. The unit of time that is used is usually a year, so

that T is expressed in years. The corresponding continuously compounded annual interest rate R_T is given implicitly by the equation

$$V_2 = e^{R_T T} V_1$$

or explicitly by $R_T = (1/T)\log(V_2/V_1)$. The corresponding discretely compounded rate R_T^d is obtained from

$$V_2 = (1 + R_T^d)^T V_1.$$

Given the initial capital V_1 and the maturity T, it is equivalent to specify either the final amount V_2 or (for instance) the discretely compounded annual interest rate, since one can be computed from the other and vice versa. To express final capital V_2 in terms of a discretely compounded annual rate is an example of a *quoting convention*. Zero-coupon bonds are also known as *discount bonds* because the value at current time 0 of a zero-coupon bond with face value 1 that matures at time T gives exactly the discount factor that must be applied to a cashflow that will occur in T years from now.

When bonds are issued by companies or by governments as a means of financing their activities, they are usually *coupon-paying bonds* rather than zero-coupon bonds. For instance the Dutch government might issue a 5-year bond with 4% coupons paying annually. For a face value of, say, 100 euros, such a bond will pay 4 euros one year after its initiation, another 4 euros after two, three, and four years, and 104 euros at the time of maturity after five years. The face value of the bond is used to compute these cashflows. The actual market value of a coupon-paying bond depends on the current term structure. A coupon-bond can be viewed as a portfolio of zero-coupon bonds; for instance, the 5-year bond with 4% annual coupon and face value 100 can be considered as a portfolio consisting of a one-year zero-coupon bond with face value 4 euro, a two-year zero-coupon bond with face value also 4 euro, and so on, and finally a 5-year zero-coupon bond with face value 104 euro. Therefore, if the prices of zero-coupon bonds of all maturities are known, then the prices of coupon-paying bonds are also known.

When party A sells a zero-coupon bond to party B, this is effectively the same as that party B provides a loan to party A which will be paid back with interest by party A at the time of maturity of the bond. A variation of this theme is the *forward rate agreement*. An FRA also involves two cashflows, one from party B to party A and one in the other direction; the difference is however that both cashflows take place in the future, say at times T_1 and T_2. The contract implies a loan from party B to party A which will be effectuated at time T_1 rather than at the time at which the contract is agreed. The amount that will be paid back at time T_2 is already specified at the time, say 0, at which the contract is agreed. In other words, the interest rate that will be paid by party A is fixed at time 0. Such an agreement

could be useful for instance when an institution already knows that it will need a loan at a given later time T_1, and the institution prefers to have the interest rate set already, rather than to wait and see at which rates loans will be available at time T_1. In a forward rate agreement, there is no cashflow at the time at which the contract is entered; the only cashflows are V_1 at time T_1 and V_2 at time T_2 in the opposite direction. The value of the contract at time 0 should therefore be zero. This means that the equality $P(T_1)V_1 = P(T_2)V_2$ should hold, where $P(T_1)$ is the discount factor (at time 0) for maturity T_1 and $P(T_2)$ is the discount factor (at time 0) for maturity T_2. The *forward rate*, i.e. the interest rate that is implied by the ratio V_2/V_1 and the time difference $T_2 - T_1$, is therefore determined by the *current* term structure.

Many interest rate products are designed to provide protection against high interest rates. Suppose for instance that a company has a loan with variable interest rate, which may be renewed say every year on the basis of the current one-year rate. The company might be interested in a contract that effectively limits its future interest payments to a given maximum. Suppose that the desired maximum is 5% and that the principal of the loan is one million euro. To achieve the required protection for the interest payment that is to be made say on January 1 of year $t + 1$, the company should receive on January 1 of year $t + 1$ the amount by which the interest rate payment that is to be made at January 1 of year $t + 1$ (namely the one-year discretely compounded interest rate as it is derived from the term structure at January 1 of year t) minus 50 000 euro, if the result of the subtraction is positive. The contract is therefore similar to a call option, except that the underlying quantity is now a future interest rate rather than a future stock price. Such a contract which provides protection for a single interest payment is called a *caplet*. This name derives from the term *interest rate cap* that is used for a series of such contracts, providing protection during a period in which several interest payments are to be made. In addition to interest rate caps, there exist also interest rate *floors* which provide protection in the opposite direction. Caps become more valuable when interest rates go up, and floors become more valuable when interest rates go down; depending on the period that is specified in a cap or floor contract, there will be sensitivity in particular to certain parts of the term structure. These contracts may therefore in general be of interest to parties who would would like to modify their exposure to particular changes of the term structure. Further refinements may be achieved by making use of options on caps and floors; these are known as *captions* and *floortions* respectively. These are instruments that provide the holder for instance with the right, but not the obligation, to enter into a cap contract with specified features (coverage period, cap rate, etc.) at a specified time in the future.

There are also term structure products that effectively replace a loan with variable interest rate by a loan with fixed interest rate for a certain period. These

instruments go by the name of *interest rate swaps* or just simply *swaps*. Swaps are very popular instruments, since they make it possible to exchange short-term funding effectively by long-term funding, or vice versa. In a swap contract, party A (called the *payer*) pays a fixed interest rate on a given principal amount to party B, while party B (called the *receiver*) pays a floating rate on the same amount to party A. Only interest payments are made between the two parties; the principal amount is only used as a reference and is not actually exchanged, so that one speaks of a *notional* principal. The dates at which payments are made are known as *tenor dates*. The floating rate payment that is to be made at tenor date T_i is usually determined by the rate that holds at time T_{i-1} for bonds that will mature at time T_i. In other words, the *fixing date* is T_{i-1}, while the *payment date* is T_i. In this way, the floating rate payments replicate the costs of roll-over short-term funding. The fixed rate that is paid by party A is called the *swap rate*. In a standard swap contract, this rate is chosen in such a way that the value of the contract at the time of initiation is zero; the swap rate that achieves this is called the *par swap rate*. At first sight it may seem that, while the current value of the fixed-rate payments (as a function of the swap rate) is easy to obtain from the current term structure, the same may not hold for the floating-rate payments, since these refer to future interest rates which are not known at the time at which the contract is agreed. However, the floating-rate payments can be generated by a strategy of which the cost can be determined on the basis of the term structure at the time of initiation of the contract. This strategy works as follows.

Suppose that the tenor dates are $T_0 = 0, T_1, \ldots$; the last tenor date is T_n. Let a notional principal be given, say one million euro. Suppose that at time T_0 an initial capital of one million euro is available. The capital can be invested at time T_0 in zero-coupon bonds that will mature at time T_1. At time T_1, the investment generates an interest payment whose size is equal to the interest rate that holds at time T_0 for the period until T_1, times the principal of one million euro. This interest rate payment exactly covers the first floating interest payment in the swap contract. At time T_1, the principal amount of one million euro is available for reinvestment and can be used to buy bonds that will mature at time T_2. The interest rate payment that is received at time T_2 is then equal to the rate for maturity T_2 *as it held at time* T_1, times the principal of one million euro. Again this is exactly equal to the floating rate payment that is to be made as a result of the swap contract. This goes on until the final tenor date T_n. This reasoning shows that it is possible in principle[1] to generate the floating interest rate payments by borrowing the principal at time T_0 and returning it at time T_n. The cost of this strategy is the difference between

[1] Implementation of the strategy as described may not be so easy due to the limited availability of default-free zero-coupon bonds with a given maturity date. However, the strategy serves only as part of a valuation argument and does not need to be actually carried out.

the value of one million euro now and the value now of one million euro to be paid at time T_n. By this argument, the value of the "floating leg" of a swap contract is $(1-P(T_n))V$ where $P(T_n)$ is the discount factor for maturity T_n and V denotes the notional principal.

The value of the "fixed leg" is equal to $(\sum_{i=1}^{n} P(T_i))r^{\text{sw}}V$ where r^{sw} is the swap rate. From the valuation of the floating leg as described above, it follows that the par swap rate is given by

$$r^{\text{sw}} = \frac{1 - P(T_N)}{\sum_{i=1}^{N} P(T_i)}\,. \tag{5.1}$$

In particular, the par swap rate is completely determined by the term structure at the time of initiation of the contract; no model for the evolution of the term structure is required. In fact the relation between swap rates and default-free discount factors is often used in reverse, as a way of determining discount factors from swap rates that are observed in the market, rather than vice versa. The reason for this is that *credit risk* (the risk of default of the counterparty) plays a role especially in the pricing of long-term bonds, whereas this type of risk is less of an issue in swap contracts since the principal is not exchanged. The mutual relation between swap rates and discount factors is worked out in a special case below in Section 5.2.4.

Options on swap contracts, known as *swaptions*, are very popular as well. A swaption gives the holder the right, but not the obligation, to enter into a specific swap contract at a specified time in the future. In a swaption contract, there is an *option maturity date* which is the time at which the holder has to decide whether or not to exercise the right to enter into the swap, and the *swap maturity date* which is the last date for interest payments as a result of the swap contract if the option is indeed exercised. The tenor dates for payments between the option maturity date and the swap maturity date are usually standardized (say, floating-rate payments every three months and fixed-rate payments every six months). For instance, a "5-by-10" swaption has an option maturity date of five years, and if the option is exercised will lead to swap payments during ten years. The specifications of the swap contract include furthermore the fixed rate that will be paid. The swap rate that is agreed in a swaption contract is similar to the strike in a call or a put option on a stock; it is a contract parameter that influences the value of the option. In contrast to swaps, the value of swaptions cannot be derived from the current term structure alone.

Credit risk plays an important role in the valuation of some term structure products. Bonds that have the same time to maturity and the same coupon rate can still have very different values in the market, depending on the credit status of the issuer. Naturally, a discount will be applied to the value of a bond that is issued by a company if investors believe that there is a nonnegligible probability that the company will go into default before the time of maturity, so that the bond

will redeemed only partly or not at all. This means that an issuer of low credit status has to pay a higher effective interest rate than the rate paid by an issuer that is considered completely safe. The difference between these interest rates (which may depend on maturity) is called the *credit spread*. Credit issues may be avoided or mitigated in interest rate contracts for instance by the posting of collateral. There are also instruments available that specifically target credit risk, such as *credit default swaps*. Modeling of credit risk is a subject of its own which will not be covered in this chapter.

5.2 Term structure descriptions

The term structure of interest rates can be represented in several different ways. In this section we discuss the four most common representations: the discount curve, the yield curve, the forward curve, and the swap curve. We will assume continuous compounding throughout, since this is convenient for theoretical purposes; in practice however, discrete compounding is often used. There are discretely compounded analogs of all of the formulas presented below.

The simplest product that can be used to determine the interest rate for a given maturity T is the default-free zero-coupon bond that is certain to pay a given amount at time T and that produces no other cashflows. In practice, bonds are usually coupon-paying; as already noted above, such contracts can be viewed as portfolios of zero-coupon bonds. The term "bond" when used below without further qualification refers to a default-free zero-coupon bond.

In this section, we consider different representations of the term structure at a given moment of time. Without loss of generality we can let this moment be 0. Note that, with this convention, the parameter T denotes both time *of* maturity and time *to* maturity. Upon conversion to a general time t, the parameter T must be replaced by $T - t$ in all instances where T denotes time *to* maturity. To simplify notation, the subscript 0 denoting current time will usually be suppressed in this section.

5.2.1 The discount curve

The value of a default-free, zero-coupon bond maturing at time T, relative to the value of its principal, is a dimensionless number called the (riskless) *discount factor* for maturity T. We will use the notation $P(T)$ for this quantity. The curve that is obtained by looking at the values of the discount factor for different maturities T is called the *discount curve*. This is one of the possible representations of the term structure. In a sense it is the most basic representation; the discount factor is defined independently of any compounding conventions.

Under the assumption that cash can be stored costlessly, an arbitrage opportunity can be constructed if the discount factor for maturity T_2 exceeds the discount factor for maturity T_1, with $T_1 < T_2$. This is done as follows. At time 0, buy a bond that pays 1 at time T_1, and sell a bond that pays 1 at time T_2. If $P(T_2) > P(T_1)$, this produces a net gain at time 0. At time T_1, store the amount 1 that is received, and use it at time T_2 to fulfill the commitment from the sold bond; no obligations remain. It follows that, under the assumption of costless storage of cash, the discount curve must be nonincreasing. Other requirements that follow from absence of arbitrage (even without the assumption of costless storage of cash) are that $P(T) > 0$ for all $T \geq 0$, and $P(0) = 1$. Empirically it is found that the discount curve is always quite smooth, so that for theoretical purposes one may reasonably assume that the curve is differentiable.

5.2.2 The yield curve

The (continuously compounded) *yield* for a given maturity $T > 0$, denoted by $R(T)$, is defined as the continuously compounded constant interest rate that is implied by the discount factor $P(T)$. That is to say, the number $R(T)$ is the solution of the equation

$$P(T) = e^{-R(T)T}$$

which is given in explicit form by

$$R(T) = -\frac{1}{T} \log P(T). \tag{5.2}$$

The definition above cannot be applied as such at $T = 0$, since $\log P(0) = 0$. The value of $R(0)$ is defined by taking the limit as the time of maturity T tends to 0:

$$R(0) = -\lim_{T \downarrow 0} \frac{\log P(T)}{T} = -\frac{P'(T)}{P(T)}\bigg|_{T=0} = -P'(0). \tag{5.3}$$

The quantity defined above is called the *short rate*; it is the instantaneous rate of growth of riskless capital. The short rate is an important theoretical notion. In practice, the rates on loans of very short maturities (overnight loans) are affected by various factors that term structure models usually do not aim to cover, so that the three-month rate is often considered to provide a better proxy for the short rate than the overnight rate. The *yield curve* is the curve that is obtained by plotting $R(T)$ as a function of T. This is the most common representation of the term structure.

It was noted above that, under the assumption that cash can be stored at no cost, the discount curve must be nonincreasing. This implies in particular that $P(T) \leq P(0) = 1$ for all $T \geq 0$, and hence that the yields for al maturities are

nonnegative. However, a nonnegative yield curve does not always correspond to a nonincreasing discount curve. Differentiating the relation $R(T)T = -\log P(T)$, we find

$$R'(T)T + R(T) = -\frac{P'(T)}{P(T)}$$

so that the requirement that the discount curve should be nonincreasing is satisfied if and only if the corresponding yield curve satisfies

$$-R'(T) \leq \frac{1}{T}\, R(T) \tag{5.4}$$

for all $T > 0$. This means that the yield curve cannot decrease too steeply.

5.2.3 The forward curve

The idea of a forward rate was already introduced above in connection with a discussion of forward rate agreements. Assume that party A will pay to party B the amount V_1 at time T_1 and that party B will pay to party A the amount V_2 at time $T_2 > T_1$. Suppose moreover that the relative sizes of the volumes V_1 and V_2 have been chosen to make the current value of the contract equal to zero, that is,

$$P(T_1)V_1 = P(T_2)V_2. \tag{5.5}$$

The *forward rate*, denoted by $F(T_1, T_2)$, is the rate that is implied by the quotient V_2/V_1, which is equal to $P(T_1)/P(T_2)$ according to the formula above, and the time difference $T_2 - T_1$. Under our convention of continuous compounding, this means that the forward rate is determined from the following equation:

$$e^{-R(T_1)T_1} e^{R(T_2)T_2} = e^{F(T_1,T_2)(T_2-T_1)}.$$

The explicit form is

$$F(T_1, T_2) = \frac{R(T_2)T_2 - R(T_1)T_1}{T_2 - T_1}\,. \tag{5.6}$$

An alternative way of writing the above relations is

$$e^{R(T_1)T_1} e^{F(T_1,T_2)(T_2-T_1)} = e^{R(T_2)T_2}.$$

This shows that, if the yield curve is increasing, the forward rate can be looked at as a "catch-up rate" which speeds up the growth of capital so that growth during time T_1 at the (low) rate $R(T_1)$ followed by growth during time $T_2 - T_1$ at the rate $F(T_1, T_2)$ leads to the same final capital as growth at the (higher) rate $R(T_2)$ during time T_2. Especially at longer maturities, a moderate increase of the yield curve therefore leads to high values of forward rates. Conversely, if the yield curve

is decreasing, the forward rate can be viewed as a slow-down rate. When the yields for maturities T_1 and T_2 are the same, say $R(T_1) = R(T_2) = R$, then the forward rate corresponding to maturities T_1 and T_2 is equal to R as well.

If for a given time of maturity T we take $T_1 = T$ and $T_2 = T + \Delta T$ and let ΔT tend to 0, then both the numerator and the denominator in the quotient at the right hand side of (5.6) tend to 0. The quotient itself tends to a finite limit which is called the *instantaneous forward rate* for maturity T:

$$F(T) = \lim_{\Delta T \to 0} \frac{R(T+\Delta T)(T+\Delta T) - R(T)T}{\Delta T} = \frac{d}{dT}(R(T)T) = R'(T)T + R(T). \tag{5.7}$$

Since $R(T)T = -\log P(T)$, we can also write

$$F(T) = -\frac{d}{dT}\log P(T) = -\frac{P'(T)}{P(T)}\,. \tag{5.8}$$

Conversely (using $P(0) = 1$), the discount rates can be expressed in terms of the forward rates by

$$P(T) = \exp\Big(-\int_0^T F(t)\,dt\Big) \tag{5.9}$$

The curve formed by the instantaneous forward rates is called the *forward curve*. The instantaneous forward rate for maturity T is also simply just called the forward rate for maturity T.

Note that $F(0) = R(0)$, so that the left endpoint of the forward curve is given by the short rate, just as it is the case for the yield curve. Since $F(T) = \frac{d}{dT}(R(T)T)$ and $(R(T)T)|_{T=0} = 0$, the yields can be expressed in terms of the forward rates by

$$R(T) = \frac{1}{T}\int_0^T F(t)\,dt. \tag{5.10}$$

In other words, the yields for different maturities can be viewed as cumulative averages of the forward rates.

It was argued above that, under the assumption of costless storage of cash, there is a certain constraint on the yield curve that must be satisfied to prevent a simple arbitrage opportunity. This is the condition that the yield curve should satisfy to make sure that the corresponding discount curve is nonincreasing. It is seen from (5.8) (or from (5.9)) that the corresponding condition in terms of the forward curve is just that the forward rates for all maturities should be nonnegative. This is a simpler condition than the requirement (5.4) in terms of the yield curve.

5.2.4 The swap curve

The fourth representation of the term structure that we will discuss is constructed from swap contracts. As discussed in Section 5.1, the (non-annualized) par swap rate in a swap contract is given implicitly by the equation

$$1 - P(T_N) = r^{\text{sw}} \sum_{i=1}^{n} P(T_i). \tag{5.11}$$

Rewriting this equation in the form

$$1 = r^{\text{sw}} \sum_{i=1}^{n} P(T_i) + P(T_n)$$

leads to another interpretation: for a coupon-paying bond with face value 1 that matures at time T_n and that provides coupon payments with a coupon rate r^{sw} at dates $T_1, \ldots, T_n$, the left hand side represents its face value, while the right hand side represents the market value of the bond. That is to say, the par swap rate is equal to the *par coupon rate*, which is defined as the coupon rate that needs to be paid so that the market value of the bond is equal to its face value. The par coupon rate is also known as the *par yield*.

The value of the floating leg in a swap contract only depends on the final time of maturity of the contract T_N, but the value of the fixed leg depends on the choice of the tenor dates. In keeping with the spirit of continuous compounding, we may choose to work with the limit situation in which the fixed payments are made continuously. The value of the fixed leg is then equal to $\int_0^T P(t) r V \, dt$ when the continuously paid rate is equal to r. The *continuously-paid par swap rate for maturity* T, denoted by $S(T)$, is then defined as the value of the continuously paid fixed rate that makes the present value of the swap contract with final maturity T equal to 0:

$$S(T) = \frac{1 - P(T)}{\int_0^T P(t) \, dt}. \tag{5.12}$$

Using (5.8) and the fact that $P(0) = 1$, we can write

$$1 - P(T) = -\int_0^T P'(t) \, dt = \int_0^T P(t) F(t) \, dt$$

so that the definition of the swap rate may also be written as

$$S(T) = \frac{\int_0^T P(t) F(t) \, dt}{\int_0^T P(t) \, dt}. \tag{5.13}$$

This may be compared to (5.10). Whereas the yields for different maturities can be

considered as *unweighted* cumulative averages of the forward rates, the swap rates are *weighted* cumulative averages. The weights are provided by the discount factors, which are higher for short maturities. Therefore the swap rates are influenced more by the short end of the forward curve than the yields, the difference being noticeable especially for longer maturities. When the forward curve is increasing as is usually the case, the swap curve lies below the yield curve; since the yield curve is increasing when the forward curve is increasing, this may also be described by saying that the swap rates lag behind the yields, so that for instance the 20-year swap rate might be equal to the 15-year yield.

The relation between swap rates and forward rates can also be written in a different way. By computing the derivatives with respect to T of both sides of the relation

$$S(T) \int_0^T P(t)\,dt = 1 - P(T)$$

(another form of (5.12)), we find

$$S'(T) \int_0^T P(t)\,dt + S(T)P(T) = -P'(T) \tag{5.14}$$

which implies, by (5.8), that

$$F(T) = S(T) + \frac{\int_0^T P(t)\,dt}{P(T)}\, S'(T). \tag{5.15}$$

This may be compared to (5.7). It follows from (5.7) that the forward curve lies above the yield curve at maturities where the yield curve is increasing, and lies below the yield curve at points where the yield curve is decreasing. The relation (5.15) implies that the forward curve has the same property with respect to the swap curve.

5.2.5 Summary and examples

A summary of the most important relations between the various representations of the term structure of interest rates is provided in Table 5.1. Each representation describes the term structure completely. All formulas have analogs for the case of discretely compounded interest. The discrete formulas do not look as nice as their continuous counterparts, but in practice they are important because discrete compounding is used in many types of contracts. The relations between yields, forward rates, and swap rates are illustrated in a few specific examples in Figs. 5.1–5.3. The examples also illustrate several different parametrizations that are popular in the literature. The parametrizations can be described either in terms of yield curves, forward curves, or swap curves; the formulas are simplest when the forward

	in terms of discount factors	in terms of yields	in terms of forward rates
$P(T)$		$\exp(-TR(T))$	$\exp\big(-\int_0^T F(t)\,dt\big)$
$R(T)$	$-\dfrac{1}{T}\log P(T)$		$\dfrac{1}{T}\displaystyle\int_0^T F(t)\,dt$
$F(T)$	$-P'(T)/P(T)$	$(TR(T))'$	
$S(T)$	$\dfrac{1-P(T)}{\int_0^T P(t)\,dt}$	no simple formula	$\dfrac{\int_0^T P(t)F(t)\,dt}{\int_0^T P(t)\,dt}$
r	$-P'(0)$	$R(0)$	$F(0)$

Table 5.1: Relations between term structure representations under continuous compounding. The symbols $P(T)$, $R(T)$, $F(T)$, and $S(T)$ refer to discount factors, yields, forward rates, and swap rates respectively; r denotes the short rate.

curve is used. The Vasicek parametrization, associated to the Vasicek model, is given by

$$F^{\text{Vas}}(T) = re^{-aT} + b\big(1-e^{-aT}\big) - \frac{\sigma^2}{2a^2}\big(1-e^{-aT}\big)^2. \tag{5.16}$$

The model has one parameter that is allowed to vary in time to accommodate changes in the yield curve; this parameter is denoted by r since it coincides with the short rate. In the Nelson-Siegel parametrization, proposed by Charles Nelson and Andrew Siegel in 1987, the forward curve is of the form

$$F^{\text{NS}}(T) = \beta_0 + \beta_1 e^{-aT} + \beta_2 aT\, e^{-aT}. \tag{5.17}$$

The three parameters β_0, β_1 and β_2 are measured in the same units as interest rates. The parameter a is positive; it can be thought of as defining a unit of time since it always occurs in the combination aT. For a fixed value of a, the expression in (5.17) can be viewed as a linear combination of three basis functions, namely the constant function 1, the exponential function e^{-aT} which shows a declining behavior, and the function aTe^{-aT} which is "hump shaped" since it starts from the value 0 at $T=0$, increases to a maximum at $T=1/a$, and then asymptotically decreases to 0 again. An extension of the Nelson-Siegel parametrization was proposed by Lars Svensson in 1995:

$$F^{\text{NSS}}(T) = \beta_0 + \beta_1 e^{-a_1 T} + \beta_2 a_1 T\, e^{-a_1 T} + \beta_3 a_2 T\, e^{-a_2 T}. \tag{5.18}$$

This is called the Nelson-Siegel-Svensson model, or just the Svensson model.

It should be noted that the term "model" is used here as another word for "parametrization". Both the Nelson-Siegel parametrization and its extension by Svensson were originally formulated just as ways of capturing the term structure at a given moment in time, without a specification of a model for the evolution of the parameters in time. However (see Exc. 5.8.9), the affine term structure models that are discussed in Section 5.6.2 below can lead to representations that are similar to (5.17) or (5.18) with constant parameters a_i and parameters β_i that can vary in time.

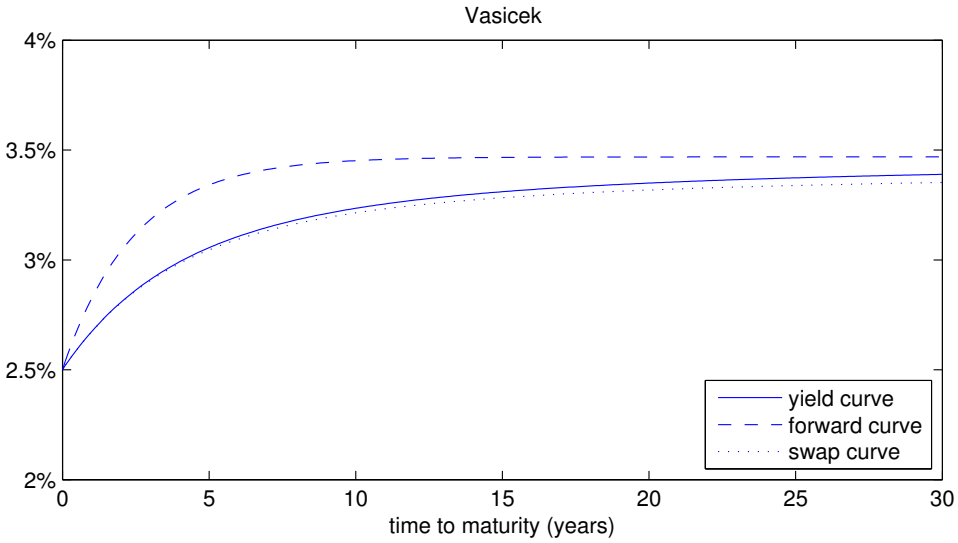

Figure 5.1: Yield curve in a Vasicek parametrization, with corresponding forward curve and swap curve. The parameter values are: $a = 0.4$, $b = 0.035$, $r = 0.025$, $\sigma = 0.01$.

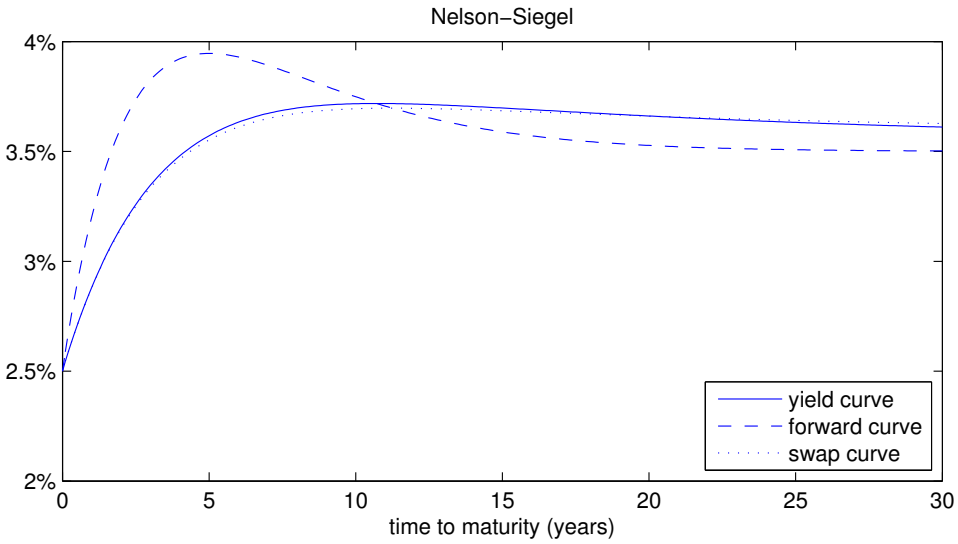

Figure 5.2: Yield curve in a Nelson-Siegel parametrization, with corresponding forward curve and swap curve. The parameter values are: $a = 0.3$, $\beta_0 = 0.035$, $\beta_1 = -0.01$, $\beta_2 = 0.02$.

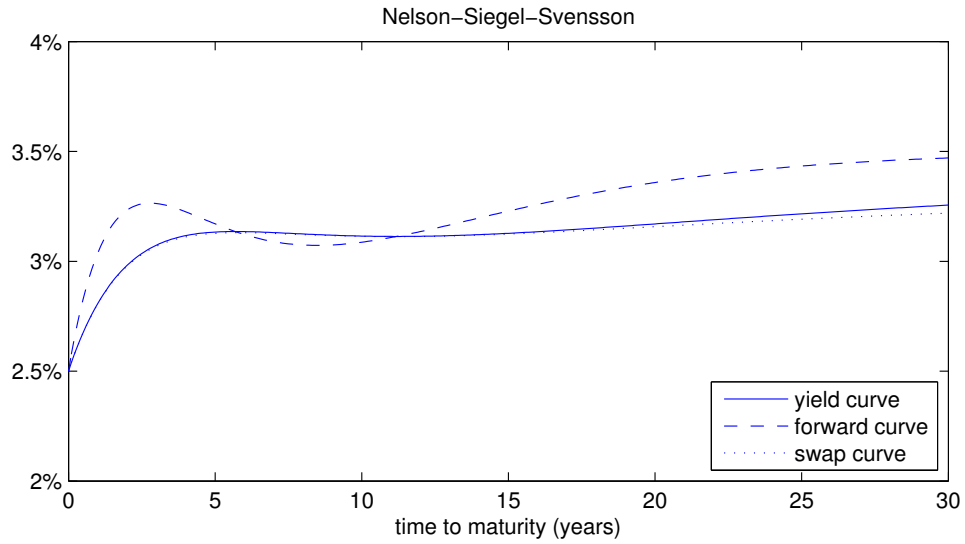

Figure 5.3: Yield curve in a Nelson-Siegel-Svensson parametrization, with corresponding forward curve and swap curve. The parameter values are: $a_1 = 0.4$, $a_2 = 0.2$, $\beta_0 = 0.035$, $\beta_1 = -0.01$, $\beta_2 = 0.02$, $\beta_3 = -0.02$.

5.3 Model-free relationships

The previous section has focused on what might be called the "cross-sectional" dimension of the term structure; we looked at interest rates as a function of maturity at a given point in time. To enable valuation of term structure derivatives such as swaps and caps, one also needs to consider the "time series" dimension. The terminology is borrowed from econometric panel data analysis. Unlike what is usually the case in panel data, in term structure models the two dimensions are in fact closely intertwined due to the assumption of absence of arbitrage. In this section, we concentrate on relationships that can be proved purely on the basis of absence of arbitrage, without any further model assumptions.

It is sometimes stated that the instantaneous forward rate for maturity T reflects market expectations concerning the level of the short rate at time T, so that in fact the forward curve gives the market's prediction for the evolution of the short rate. This is called the *expectations hypothesis*. As always, one has to be careful in interpreting prices as predictions; the purpose of markets is to equilibrate demand and supply, rather than to implement a statistical procedure. Risk premia will often generate a bias with respect to expectations under the real-world measure.

All relations obtained in this section are consequences of the Fundamental Theorem of Asset Pricing as expressed through the numéraire-dependent pricing formula (3.25), with different choices of the numéraire. Let us start with what might be called the "generic" choice of a numéraire, namely the money market account. By

definition, the value of the money market account M_t follows the process given by

$$dM_t = r_t M_t \, dt \tag{5.19}$$

where r_t is the short rate at time t. By the numéraire-dependent pricing formula, the price at time 0 of a bond that pays one unit of currency at time T satisfies

$$\frac{P_0(T)}{M_0} = E^{\mathbb{Q}_M}\left[\frac{P_T(T)}{M_T}\right] = E^{\mathbb{Q}_M}\left[\frac{1}{M_T}\right]$$

since $P_T(T) = 1$ by definition. The measure $\mathbb{Q}_M$ is the risk-neutral measure, i.e. the martingale measure that corresponds to taking the money market account as a numéraire. From the relation above, we have

$$P_0(T) = E^{\mathbb{Q}_M}\left[\frac{M_0}{M_T}\right]. \tag{5.20}$$

An expression that is stated more explicitly in terms of the short rate can be obtained from (5.19) in the following way. Writing $m_t := \log M_t$, we have from Itô's rule

$$dm_t = r_t \, dt \tag{5.21}$$

so that

$$m_t = m_0 + \int_0^t r_s \, ds.$$

From this, one finds

$$M_t = M_0 \exp\Big(\int_0^t r_s \, ds\Big). \tag{5.22}$$

In this way, the following relation is established:

$$\boxed{P_0(T) = E^{\mathbb{Q}_M}\left[\exp\Big(-\int_0^T r_t \, dt\Big)\right].} \tag{5.23}$$

Here we have a relationship between the current term structure on the one hand, and on the other hand expectations (under $\mathbb{Q}_M$) relating to the future evolution of the short rate. In general however it is not just the expectation of future short rates which plays a role, but also their variance and in fact their entire distribution, because expectation is taken of a nonlinear function (namely the exponential) of the random variable $-\int_0^T r_s \, ds$. More generally than (5.23), one can also write

$$P_t(T) = E_t^{\mathbb{Q}_M}\left[\exp\Big(-\int_t^T r_s \, ds\Big)\right]. \tag{5.24}$$

to give the term structure at an arbitrary time t. The expression above shows that any model under $\mathbb{Q}_M$ for the evolution of the short rate in fact completely defines a model for the evolution of the term structure as a whole.

Instead of the money market account, one can also take the bond that matures at a given time T as a numéraire. The corresponding martingale measure is called the *T-forward measure*; this measure is denoted by $\mathbb{Q}_T$. For any positive number ΔT, we can write the following relation:

$$\frac{P_0(T+\Delta T)}{P_0(T)} = E_0^{\mathbb{Q}_T}\frac{P_T(T+\Delta T)}{P_T(T)} = E_0^{\mathbb{Q}_T}P_T(T+\Delta T) \tag{5.25}$$

since $P_T(T) = 1$ by definition. The above relation is an application of the numéraire-dependent pricing formula, with $P_t(T+\Delta T)$ as the asset and $P_t(T)$ as the numéraire. Subtract 1 from both sides of the equation above, divide by ΔT, and take the limit as ΔT tends to zero. We obtain on the left hand side

$$\lim_{\Delta T\downarrow 0}\frac{P_0(T+\Delta T)-P_0(T)}{P_0(T)\Delta T} = \frac{P_0'(T)}{P_0(T)} = -F_0(T)$$

and on the right hand side

$$\lim_{\Delta T\downarrow 0} E_0^{\mathbb{Q}_T}\frac{P_T(T+\Delta T)-1}{\Delta T} = E_0^{\mathbb{Q}_T}P_T'(T) = -E_0^{\mathbb{Q}_T}r_T.$$

Therefore, we arrive at the following relation:

$$F_0(T) = E_0^{\mathbb{Q}_T}r_T. \tag{5.26}$$

In other words, the forward rate for maturity T is equal to the expected value of the short rate at time T under the martingale measure corresponding to the bond maturing at time T. It has to be considered that this martingale measure incorporates in general a risk adjustment, so that the expectation under the objective ("real-world") measure is not necessarily the same. The relation (5.26), like the bond price equation (5.23), establishes a connection between the shape of the term structure and its stochastic evolution.

Another relation that can be obtained using the T-forward measure is the following. For given maturities T_1 and T_2 with $T_2 > T_1$, use the T_2-bond as a numéraire and consider the price at time 0 of the T_1-bond. On the basis of the numéraire-dependent pricing formula, we can write

$$\frac{P_0(T_1)}{P_0(T_2)} = E^{\mathbb{Q}_{T_2}}\frac{P_{T_1}(T_1)}{P_{T_1}(T_2)} = E^{\mathbb{Q}_{T_2}}\frac{1}{P_{T_1}(T_2)}\,. \tag{5.27}$$

Consider now a floating-rate payment as would occur in a swap contract, when T_1

and T_2 refer to successive tenor dates. The payment that is to be made at time T_2 is the interest that is received at time T_2 when the notional principal is invested at time T_1 in bonds that mature at time T_2. The total amount received at time T_2 for an investment of V at time T_1 is $(1/P_{T_1}(T_2))V$, and the part of this which represents interest is $C_{T_2} := (1/P_{T_1}(T_2))V - V$. The value of this uncertain payoff at time 0 can be found from the numéraire-dependent pricing formula:

$$\frac{C_0}{P_0(T_2)} = E^{\mathbb{Q}_{T_2}}\Big(\frac{1}{P_{T_1}(T_2)} - 1\Big)V.$$

According to (5.27), the right hand side is equal to $(P_0(T_1)/P_0(T_2) - 1)V$. Therefore we find

$$C_0 = (P_0(T_1) - P_0(T_2))V.$$

This simple expression was also found in Section 5.1 on the basis of a replication argument. The expression above applies to a single floating-rate payment. The value of a series of floating-rate payments in connection with a number of successive tenor dates $T_0 = 0, T_1, \ldots, T_n$ is given by

$$(P_0(0) - P_0(T_1))V + (P_0(T_1) - P_0(T_2))V + \cdots + (P_0(T_{n-1} - P_0(T_n))V \\ = (1 - P(T_n))V$$

where use is made of the relation $P_0(0) = 1$ and of the telescope rule. Again this derivation on the basis of the numéraire-dependent pricing formula, with a suitable choice of the numéraire, confirms the expression that was already derived above on the basis of a replication argument.

5.4 Requirements for term structure models

Dynamic term structure models are used to generate scenarios for the future development of interest rates. When these models are formulated under the real-world measure, they can be used for risk management purposes. Term structure models that are formulated under a martingale measure that corresponds to a chosen numéraire are used for pricing purposes. Such models can also be obtained from models under the real-world measure by incorporating a market price of risk; in practice however, one often goes immediately to a specification under a martingale measure. The number of traded assets in the interest-rate market is theoretically infinite, since zero-coupon bonds with different maturity dates are independent assets.[2] Term structure models may aim at describing the evolution of the term

[2]The term "independent assets" is used here in the sense that the price of any one cannot be written as a deterministic function of the prices of the others using only absence of arbitrage.

structure as a whole, or they may have the more modest ambition of describing only the evolution of certain term structure products. The latter models will be referred to below as "partial models".

One way to formulate a term structure model is to specify a state equation

$$dX_t \;=\; \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t \tag{5.28}$$

where W_t is a Brownian motion under the real-world measure, together with a bond pricing function $\pi_T(t, x)$ corresponding to each maturity $T \geq t$:

$$P_t(T) \;=\; \pi_T(t, X_t). \tag{5.29}$$

To prevent arbitrage in this model, there should exist a function $\lambda(t, x)$ such that for all T the equality $\mu_T - r\pi_T = \sigma_T \lambda$ is satisfied. Here, μ_T and σ_T refer to the drift and volatility of the price of the bond with maturity T. One then has a fully dressed-up model that can be used for pricing term structure products as well as for risk management purposes.

An alternative way to arrive at a specification of a term structure model is to start with a specification of the process of the short rate. When bond prices are subsequently obtained from the formula (5.23), absence of arbitrage is automatically guaranteed. Given that the formula (5.23) calls for the computation of an expectation under the risk-neutral measure $\mathbb{Q}_M$, it can be convenient to start from a specification under that measure. Depending on the specification of the short rate, it may or may not be possible to obtain an explicit expression from (5.24). A well-known example of a model for which it is indeed possible to get such an expression is the Vasicek model that was already introduced in Chapter 3. The process for the short rate may be specified by means of a state differential equation of the usual form $dX_t = \mu_X\, dt + \sigma_X\, dW$, in combination with a specification of the short rate at time t of the form $r_t = r(t, X_t)$ where $r(\cdot, \cdot)$ is a function of $1 + n$ variables. This is a general class of models. The term "short rate model" is used for models in which the short rate acts as the *only* state variable.

In addition to absence of arbitrage, there are several other conditions which a good term structure model should fulfill. Here are some of the questions that may be asked:

(i) is the model capable of representing the current term structure

(ii) is the model capable of reproducing current prices of term structure derivatives

(iii) does the model produce plausible future term structures

(iv) is it easy to to compute prices of popular term structure derivatives such as caps and swaptions.

It is not easy to reconcile all these requirements, due in part to the fact that the condition of absence of arbitrage provides a connection between the shape of the term structure on the one hand and its dynamics on the other hand. There is no single term structure model that has such a dominant position as the standard

Requirements (i) and (ii) in the above list are often met (at least approximately) by *calibration* of a given term structure model, that is, by adapting the parameters of the model in such a way that the prices of term structure products as produced by the model match prices that are observed in the market as closely as possible. A good term structure model is one that is able to match observed prices to a high degree of accuracy with a relatively small number of calibrated parameters, and which is such that the parameter values do not change dramatically over time when they are recalibrated to new data. Large jumps of parameter values under recalibration are a sign of overparametrization.

If the purpose of calibration is just to match observed bond prices (rather than to reproduce also the prices of term structure derivatives such as caps and swaptions), then there is a relatively simple way to achieve exact calibration, starting from any given term structure model. This method is based on the expression

$$F_0(T) = -\frac{d}{dT} \log E^{\mathbb{Q}_M}\Big[\exp\Big(-\int_0^T r_t\, dt\Big)\Big] \tag{5.30}$$

which follows from (5.23) and (5.8). Note that this expression has a certain symmetry; the operations on the left hand side of the expectation symbol are the inverses of those that appear on the right hand side. In particular, if $f(t)$ is a deterministic function of time, then

$$-\frac{d}{dT} \log E^{\mathbb{Q}_M}\Big[\exp\Big(-\int_0^T (r_t + f(t))\, dt\Big)\Big] = F_0(T) + f(T). \tag{5.31}$$

This means that any forward curve can be reproduced by adding a suitable deterministic function of time to the short rate process. In this way, the "cross-sectional" properties of a given model are adjusted by manipulating its "time series" behavior, based on the connection between the two dimensions as expressed by (5.23) or (5.30). In particular it is seen that bond prices can already be reproduced exactly by a very simple model, namely the one that prescribes the short rate deterministically by $r_t = F_0(t)$, where $F_0(\cdot)$ is the current forward curve.

5.5 Short rate models

The simplest interest rate models are the ones that have only one state variable. The standard choice for the state variable in such models is the short rate itself. We then speak of a *short rate model*. Models of this type assume that the short rate

follows a Markovian process; in other words, the statistics of the future evolution of the term structure are determined completely by the current value of the short rate. This is strong simplification of reality. However, it should be kept in mind that any model is a simplification of reality, and the choice of a model always involves a balance between realism and simplicity.

Bond prices in a short rate model can be determined by the general formula (5.24), or from the Black-Scholes equation. In the latter approach, when the equation $\mu_T - r\pi_T = \sigma_T\lambda$ is written in explicit form, the following PDE for the functions $\pi_T(t,r)$ emerges:

$$-\frac{\partial \pi_T}{\partial t} = (\mu - \sigma\lambda)\frac{\partial \pi_T}{\partial r} + \tfrac{1}{2}\sigma^2\frac{\partial^2 \pi_T}{\partial r^2} - r\pi_T, \qquad \pi_T(T,r) = 1 \tag{5.32}$$

where we write μ instead of μ_X and σ instead of σ_X to make the notation a bit lighter. All bond prices satisfy the same partial differential equation (the "term structure equation"); the difference between the equations is only in the boundary condition.

The term structure equation is specified by the three functions $\mu = \mu(t,r)$, $\sigma = \sigma(t,r)$, and $\lambda = \lambda(t,r)$. In fact the functions μ and λ only occur in the combination $\mu - \lambda\sigma$ so that there are many combinations of μ and λ that give rise to the same term structure equation. The model still leaves freedom to consider particular functional forms; for instance the well-known Vasicek model is obtained by choosing

$$\mu(t,r) - \sigma(t,r)\lambda(t,r) = a(b-r), \qquad \sigma(t,r) = c \tag{5.33}$$

where a, b, and c are constants. Perhaps the most popular alternative short rate model is the one that was proposed by John Cox, Jonathan Ingersoll and Stephen Ross in 1985, and that is hence known as the Cox-Ingersoll-Ross model or as the CIR model. The evolution of the short rate as specified in this model is

$$dr_t = a(b - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t. \tag{5.34}$$

For pricing purposes, this is usually taken as a model under the risk-neutral measure. If (5.34) is interpreted as a real-world model, then it can be rewritten as a model of the same form under $\mathbb{Q}_M$ if a suitable assumption is made concerning the price of risk; for instance, it could be assumed that the price of risk is proportional to $\sqrt{r_t}$.

Several possible choices for the risk-adjusted drift and volatility parameters in short rate models have been proposed, which give rise to different shapes of the term structure as expressed for instance by the yield curve. Given the drift and the volatility parameters, the term structure in a short rate model is completely determined by the current level of the short rate. For the evolution of the term structure this means that the interest rates for all maturities are fully correlated;

any given movement of the one-year rate, for instance, can only be accompanied by one particular movement of the 20-year rate. More flexible models can be obtained by increasing the number of state variables.

5.6 Affine models

A term structure model is said to be *affine* if the yield curves that it produces are of the form

$$R_t(T) = \alpha(t,T) + \beta(t,T)X_t \tag{5.35}$$

where $\alpha(t,T)$ is a scalar-valued function, $\beta(t,T)$ is a row vector of length n, and X_t is an n-vector of state variables. If the model is time-homogeneous, then the functions $\alpha(t,T)$ and $\beta(t,T)$ depend on t and T only through the difference $T-t$. The yield curve at any given time is then a linear combination of the functions $\alpha(T)$ and $\beta_1(T),\dots,\beta_n(T)$ (writing T now for the time *to* maturity), of which the first always has coefficient 1, while the coefficients of the components of the row vector $\beta(T)$ are stochastically varying state variables. Yield curves that are structured in this way are obtained for instance from models of the form

$$dX_t = (AX_t + g)\,dt + B\,dW_t, \qquad r_t = h^\top X_t \tag{5.36}$$

where A and B are constant matrices of sizes $n\times n$ and $n\times k$ respectively, g and h are constant n-vectors, and W_t is a Brownian motion under the risk-neutral measure $\mathbb{Q}_M$. If the volatility matrix B is made dependent on the state variable X_t in such a way that $BB^\top$ is affine in X_t, then the associated term structure is still affine; an example is the CIR model (5.34). In this section, however, the focus is on models in which the volatility matrix is constant, as in (5.36).

5.6.1 Single state variable

If the dimension of the state variable in (5.36) is 1, then, by linear scaling, the state variable can be made equal to the short rate. The resulting model is the Vasicek model that was already discussed in Chapter 3:

$$dr_t = a(b-r_t)\,dt + \sigma\,dW_t \tag{5.37}$$

where W_t is a Brownian motion under the risk-neutral measure $\mathbb{Q}_M$. An extended version of the model will be considered in Section 5.6.3 below.

In Chapter 3 we started in fact with a specification of the Vasicek model under the real-world measure. If it is assumed that the price of interest rate risk λ is

constant, then the real-world model

$$dr_t = a(b - r_t)\,dt + \sigma\,dW_t^{\mathbb{P}} \tag{5.38}$$

corresponds via the formula $dW_t^{\mathbb{Q}_M} = \lambda\,dt + dW_t^{\mathbb{P}}$ to the model under $\mathbb{Q}_M$ that is given by

$$\begin{aligned} dr_t &= a(b - r_t)\,dt + \sigma(-\lambda\,dt + dW_t^{\mathbb{Q}_M}) \\ &= a(b - \frac{\lambda\sigma}{a} - r_t)\,dt + \sigma\,dW_t^{\mathbb{Q}_M}. \end{aligned}$$

The model that is obtained in this way is indeed of the form (5.37), with the parameter b in the model under $\mathbb{Q}_M$ corresponding to the parameter $b - \lambda\sigma/a$ in the model under $\mathbb{P}$.

In principle it is possible (in an idealized world in which the Vasicek model gives an accurate description of the short rate dynamics, and the price of risk is indeed constant) to infer the value of the price of interest rate risk λ by combining pricing data with time series data. Empirical work in this direction usually leads to negative values of λ. As discussed in Section 3.6, this would mean that a positive increment of the Brownian motion $W^{\mathbb{P}}$ that appears in the equation (5.38) is associated to bad news for the general investor. For instance, it can be noted that an increase of interest rates implies a decrease of bond values.

The yield curve in the Vasicek model was already obtained in (4.67) and is repeated here:

$$R_0(T) = b - \frac{\sigma^2}{2a^2} + \left(r_0 - b + \frac{\sigma^2}{a^2}\right)\frac{1 - e^{-aT}}{aT} - \frac{\sigma^2}{2a^2}\frac{1 - e^{-2aT}}{2aT}\,. \tag{5.39}$$

The function $(1 - e^{-aT})/(aT)$ represents the average of the exponential function e^{-at} over the interval from 0 to T (that is to say, we have $(1/T)\int_0^T e^{-at}\,dt = (1 - e^{-aT})/(aT)$). The function therefore tends to 1 if $T > 0$ tends to zero, and its limit as T tends to infinity is 0. From the formula above, it consequently follows that

$$\lim_{T\to\infty} R_0(T) = b - \frac{\sigma^2}{2a^2}\,. \tag{5.40}$$

In other words, the Vasicek model predicts that, while the yield curve varies in time, it always converges to the same value for long maturities.[3] When the limit value

[3]The same property is generally found in arbitrage-free term structure models. Reasons why this should be so are discussed by Philip Dybvig, Jonathan Ingersoll and Stephen Ross in their paper "Long forward and zero-coupon rates can never fall", *Journal of Business* 69 (1996), 1–25.

above is denoted by $R_0(\infty)$, we can write

$$R_0(T) = R_0(\infty) + (r_0 - R_0(\infty)) \frac{1 - e^{-aT}}{aT} + \frac{\sigma^2}{2a^2} \frac{(1 - e^{-aT})^2}{2aT}. \tag{5.41}$$

For different values of the current short rate r_0, different yield curves are obtained as shown in Fig. 4.4.

The Vasicek model represents an attempt to explain the prices of all term structure products (not only bonds, but also swaptions, caps, and so on) at all times as a function of just one variable, namely the short rate. Moreover, the form of this function is determined by only three parameters. Calibration of the model therefore requires some courage. On the basis of the theory of affine processes (see Section 2.6.3), it can be established that the covariance of r_t and r_{t+h} in the model (5.37) for fixed h is given in the stationary situation by

$$\operatorname{cov}(r_t, r_{t+h}) = \frac{e^{-ah}}{2a} \sigma^2. \tag{5.42}$$

In particular, the standard deviation of the stationary distribution is $\sigma/\sqrt{2a}$. It then follows from the above that the correlation coefficient between r_t and r_{t+h} in the stationary situation is equal to e^{-ah}. These quantities do not depend on the parameter b, and so they are not affected by change of measure; in other words, the real-world variance and covariance can be used to calibrate the parameters a and σ. One possible approach therefore could be to work with interest rate data for a particular period, and to use the observed standard deviation of the short rates and the correlation coefficient for instance for $h = 1$ to find estimates of a and σ. Alternatively, the parameter σ might be determined on the basis of the observed short-term volatility of the short rate. In order to obtain an estimate for the parameter b in (5.37), which incorporates the price of risk, one has to use pricing information. For instance the relation (5.40) could be used. Many other approaches are possible, and in a given concrete situation a choice should be made with an eye on the purpose for which the model is going to be used. In applications with a short time horizon, it is important to take the current shape of the term structure into account, and then the parameters in the Vasicek model may be chosen on the basis of curve fitting to ensure that a reasonable match with the current yield curve is obtained.

5.6.2 Higher-dimensional models

To find a formula for bond prices under the general affine model (5.36), one can use the formula (5.20) where $M_0 = 1$. The process $m_t := \log M_t$ satisfies the stochastic differential equation (5.21). Since the short rate process satisfies $r_t = h^\top X_t$ as

specified in (5.36), the joint process of X_t and m_t is described by

$$d\begin{bmatrix} X_t \\ m_t \end{bmatrix} = \left(\begin{bmatrix} A & 0 \\ h^\top & 0 \end{bmatrix} \begin{bmatrix} X_t \\ m_t \end{bmatrix} + \begin{bmatrix} g \\ 0 \end{bmatrix} \right) dt + \begin{bmatrix} B \\ 0 \end{bmatrix} dW_t.$$

This is a linear stochastic differential equation, with deterministic initial conditions. Consequently, the joint distribution of X_T and m_T is normal. Moreover, it follows from the solution formulas in Section 2.6.3 (see (2.80) and (2.85)) that the expected value of the vector consisting of X_T and m_T depends linearly on X_0 and m_0, whereas the variance depends on T and on the system parameters, but not on the initial conditions. Since $m_0 = 0$ by definition, the expected values of X_T and m_T actually depend on X_0 only. Therefore, the expected value of m_T is a linear function of X_0 with coefficients that may depend on T, while the variance of m_T is a function of T and does not depend on X_0. We also know that m_T is normal; therefore the money market account at time T, $M_T = \exp(m_T)$, follows a lognormal distribution. Applying the standard formula for the expectation of a lognormal variable, one finds from (5.20) that the bond price $P_0(T)$ in the model (5.36) is of the form

$$P_0(T) = \exp\Big(f_0(T) + \sum_{i=1}^{n} f_i(T) X_{0,i} \Big) \tag{5.43}$$

for certain functions $f_0(T), f_1(T), \ldots, f_n(T)$. The corresponding yield curve depends in an affine way on the state variables:

$$R_0(T) = -\frac{1}{T}\Big(f_0(T) + \sum_{i=1}^{n} f_i(T) X_{0,i} \Big). \tag{5.44}$$

Since nothing in the model depends on calendar time, time 0 is in fact a generic point in time and T represents time to maturity. It is therefore confirmed that (5.36) is an affine term structure model as defined at the beginning of this section. The SDE that drives the coefficients and the functions that together form the yield curve cannot be chosen independently of each other; quite to the contrary, the basis functions for the yield curve are fixed once the SDE (under the risk-neutral measure) has been chosen.

The solution formulas in Section 2.6.3 can be used to obtain explicit expressions for the functions $f_i(T)$ appearing in (5.43), using the fact that

$$\exp\left(\begin{bmatrix} A & 0 \\ h^\top & 0 \end{bmatrix} t \right) = \begin{bmatrix} \exp(At) & 0 \\ h^\top \int_0^t \exp(As)\, ds & 1 \end{bmatrix}. \tag{5.45}$$

An alternative approach is to note that (5.43) gives expressions for arbitrage-free asset prices, so that the no-arbitrage condition (3.80) must hold. The subscript T

will be used to refer to the bond with maturity T. From (5.43) we have

$$\pi_T(t,x) = \exp\Big(f_0(\tau) + \sum_{i=1}^{n} f_i(\tau)x_i\Big), \qquad \tau = T - t$$

where x_i is the i-th component of the column vector x, and τ is the time to maturity. The corresponding drift under the risk-neutral measure can be obtained from (3.7) together with (5.36). We have

$$\begin{aligned}
\frac{\partial \pi_T}{\partial t}(t,x) &= \big(-f_0'(\tau) - \sum_{i=1}^{n} f_i'(\tau)x_i\big)\,\pi_T(t,x)\\
\frac{\partial \pi_T}{\partial x_i}(t,x) &= f_i(\tau)\,\pi_T(t,x)\\
\frac{\partial^2 \pi_T}{\partial x_i \partial x_j}(t,x) &= f_i(\tau)\,f_j(\tau)\,\pi_T(t,x).
\end{aligned}$$

Let $\bar{f}(t)$ denote the column vector with entries $f_i(t)$. One can then write, using (3.7),

$$\mu_T(t,x) = \big[-f_0'(\tau) - \bar{f}'(\tau)^\top x + \bar{f}(\tau)^\top(Ax+g) + \tfrac{1}{2}\bar{f}(\tau)^\top BB^\top \bar{f}(\tau)\big]\,\pi_T(t,x).$$

The condition (3.80) therefore becomes in this particular case:

$$-f_0'(\tau) - \bar{f}'(\tau)^\top x + \bar{f}(\tau)^\top(Ax+g) + \bar{f}(\tau)^\top BB^\top \bar{f}(\tau) = h^\top x.$$

The relation must hold for all x. Consequently, the following two equations must both be satisfied:

$$\bar{f}'(\tau) = A^\top \bar{f}(\tau) - h \tag{5.46a}$$

$$f_0'(\tau) = \bar{f}(\tau)^\top g + \tfrac{1}{2}\bar{f}(\tau)^\top BB^\top \bar{f}(\tau). \tag{5.46b}$$

Moreover, the condition $\pi_T(T,x) = 1$ for all x implies that $f_0(0) = 0$ and $\bar{f}(0) = 0$. It is seen that the vector function $\bar{f}$ can be found by solving a system of n linear differential equations. Afterwards, the scalar function f_0 is obtained by integration. The two equations (5.46) describe the link between the "cross-sectional" functions f_0 and $\bar{f}$ and the "time series" model (5.36) for the state variables which, as seen in (5.43), serve as weights associated to the functions.

Since the coefficients follow a normal distribution, the possibility that interest rates take arbitrarily large negative or positive values is not excluded by affine models, although the associated probabilities are small for usual choices of the model parameters. Perhaps a bigger concern is that the number of parameters is usually not enough to get an accurate match with the term structure as observed in the

market at a given point in time. One possibility to bring more flexibility into the model is to allow parameters to be time-dependent. This approach is discussed next.

5.6.3 The Hull-White model

The Vasicek model can produce only a limited range of term structures. Typically it is not possible to choose the parameters in the model in such a way that bond prices as determined from the model match the bond prices that are observed in the market. The reason is simply that the model has only three parameters, while there are bond prices for many different maturities. To accommodate all of those bond prices, more parameters are required. An extension of the Vasicek model was proposed in 1990 by John Hull and Alan White. In fact they proposed several extensions, but the one that has become most popular has the form

$$dr_t = (\theta(t) - ar_t)\,dt + \sigma\,dW_t \tag{5.47}$$

where $\theta(t)$ is a deterministic function of time, a and σ_r are constants, and W_t denotes a process that is a Brownian motion under the risk-neutral measure. The model can be looked at as an OU model with time-dependent reversion level given by $\theta(t)/a$. Since the short rate is the only state variable, the model can be categorized as a short rate model.

Alternatively one can extend the Vasicek model by allowing a deterministic function of time to be added to the short rate process, as discussed in Section 5.4. This leads to a model of the form

$$dX_t = -aX_t\,dt + \sigma\,dW_t, \quad X_0 = 0 \tag{5.48a}$$

$$r_t = X_t + f(t) \tag{5.48b}$$

where W_t is as before a Brownian motion under the risk-neutral measure, and $f(t)$ is a differentiable deterministic function of time. The state variable is now not the same as the short rate; therefore it is denoted by X_t rather than by r_t. The parameter b has been set equal to 0 in the above formulation, and the initial value of the state variable has been set to 0 as well; this can be done because, as is seen from the solution formula (2.79), nonzero values of these quantities only give rise to a deterministic additive term in the solution, and such a term is already covered by the function $f(t)$. Computing the differential of r_t from the above equations, we obtain

$$\begin{aligned} dr_t &= dX_t + f'(t)\,dt = (f'(t) - aX_t)\,dt + \sigma\,dW_t = \\ &= (f'(t) + af(t) - ar_t)\,dt + \sigma\,dW_t \end{aligned}$$

which shows that the model (5.48) is the same as the model (5.47) provided that

$$f'(t) + af(t) = \theta(t) \quad (t \geq 0), \qquad f(0) = r_0. \tag{5.49}$$

More explicitly, this means that the relation between $\theta(t)$ and $f(t)$ is given by

$$f(t) = e^{-at} r_0 + \int_0^t e^{-a(t-s)} \theta(s)\, ds. \tag{5.50}$$

Therefore we can find $\theta(\cdot)$ from $f(\cdot)$ or vice versa.

In order to see in what way a particular choice of the function $f(t)$ in the model (5.48) has an impact on the term structure associated to the model at time 0, we can for instance compute the forward curve. From (5.30) and (5.48) it follows that the forward curve that is produced by the model (5.48) at time 0 is given by

$$F_0(T) = f(T) - \frac{d}{dT} \log \Big[E^{\mathbb{Q}_M} \Big[\exp\Big(- \int_0^T X_s\, ds \Big) \Big] \Big]. \tag{5.51}$$

To obtain a more explicit expression on the basis of the model specification (5.48), introduce a new random variable Z_t by

$$Z_t = \int_0^t X_s\, ds$$

and write just E instead of $E^{\mathbb{Q}_M}$ for brevity. The evolution of the processes X_t and Z_t is given by (5.48a) and by the equation $dZ_t = X_t\, dt$; consequently, the pair consisting of X_t and Z_t follows a multivariate Ornstein-Uhlenbeck process. In particular, Z_t is normally distributed for all t. Since $EX_t = 0$ for all $t \geq 0$ as follows from (5.48a), we also have $EZ_t = 0$ for all t. By the standard rule

$$E\big[e^{\mu + \sigma Z}\big] = e^{\mu + \frac{1}{2}\sigma^2} \qquad (Z \sim N(0,1)). \tag{5.52}$$

for the expectation of a lognormal variable and because $EZ_T = 0$, we have

$$\log E\big[\exp(-Z_T)\big] = \tfrac{1}{2} \operatorname{var}(Z_T).$$

The variance of Z_T can be computed by means of the formulas for linear SDEs that have been discussed in Section 2.6.3. The joint equations for X_t and Z_t are

$$d \begin{bmatrix} X_t \\ Z_t \end{bmatrix} = \begin{bmatrix} -a & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ Z_t \end{bmatrix} dt + \begin{bmatrix} \sigma \\ 0 \end{bmatrix} dW_t. \tag{5.53}$$

Define

$$A = \begin{bmatrix} -a & 0 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} \sigma \\ 0 \end{bmatrix}. \tag{5.54}$$

Since $X_0 = 0$ as well as $Z_0 = 0$, the expression in formula (2.79) becomes

$$\begin{bmatrix} X_t \\ Z_t \end{bmatrix} = \int_0^t e^{A(t-s)} B \, dW_s. \tag{5.55}$$

Computation shows that, with A as given in (5.54), we have

$$e^{At} = \begin{bmatrix} e^{-at} & 0 \\ \frac{1}{a}(1 - e^{-at}) & 1 \end{bmatrix}.$$

Therefore we have from (5.55)

$$Z_T = \int_0^T \frac{1}{a}\big(1 - e^{-a(T-s)}\big)\sigma \, dW_s.$$

From the standard expression (2.49) for the variance of a stochastic integral with deterministic integrand, we find that

$$\operatorname{var} Z_T = \int_0^T \Big(\frac{1}{a}\big(1 - e^{-a(T-s)}\big)\sigma\Big)^2 ds = \frac{\sigma^2}{a^2}\int_0^T \big(1 - e^{-as}\big)^2 ds$$

where in the second expression the integral has been rewritten using a change of variable from s to $T - s$. It follows that

$$\frac{d}{dT} \log E\big[\exp(-Z_T)\big] = \frac{\sigma^2}{2a^2}\big(1 - e^{-aT}\big)^2.$$

The expression (5.51) can now be written more explicitly:

$$F_0(T) = f(T) - \frac{\sigma^2}{2a^2}\big(1 - e^{-aT}\big)^2. \tag{5.56}$$

The expression shows how to construct the model in such a way that a given forward curve is matched; namely, the function $f(t)$ in the model should be taken equal to

$$f(t) = F_0^m(t) + \frac{\sigma^2}{2a^2}\big(1 - e^{-at}\big)^2$$

where $F_0^m(t)$ is the forward curve that is observed in the market. The constraint $f(0) = r_0$ is satisfied automatically, since the left end point of the forward curve $F_0^m(0)$ is equal to the short rate (see (5.7)). The observed forward curve $F_0^m(T)$ is taken here as an *input*, and the model function $f(t)$ is taken as an *output*.

In terms of the original model formulation (5.47), the calibration of the model function $\theta(t)$ becomes

$$\theta(t) = af(t) + f'(t) = aF_0^m(t) + (F_0^m)'(t) + \frac{\sigma^2}{2a}(1 - e^{-2at}). \tag{5.57}$$

It may be advisable to use the formulation (5.48) rather than (5.47), since in practice the forward curve is obtained from market data for a finite set of maturities, so that the differentiation required in (5.57) is not a completely straightforward operation.

The Vasicek model is recovered from the Hull-White model by taking $\theta(t) = ab$. From (5.50) and (5.56) one quickly obtains the forward rate formula according to the Vasicek model:

$$F_0(T) = r_0 e^{-aT} + b(1 - e^{-aT}) - \frac{\sigma_r^2}{2a^2}\big(1 - e^{-aT}\big)^2 \tag{5.58}$$

as already stated in (5.16).

The parameter a that appears in the Hull-White model is subject to the effects of risk aversion, just as $\theta(t)$ is. This parameter is therefore also usually calibrated on the basis of market prices. When bond prices are already matched by a suitable choice of $\theta(t)$, other term structure products need to be used to calibrate a. There are plenty of such products available; for instance swaptions can be used. The parameter σ_r is often calibrated in this way as well, even though it is a volatility parameter so that it is possible to relate it directly to the variability of interest rates. Given that many swaption prices are available, one cannot expect to match these all by tuning only two parameters. One might consider replacing a and σ by functions $a(t)$ and $\gamma(t)$ that depend deterministically on time, but the robustness of such a procedure may be questioned. Alternatively one can choose a and σ in such a way that the prices of a number of selected products are matched as closely as possible according to some reasonable criterion (for instance sum of squares of absolute errors, sum of absolute values of absolute errors, sum of squares of relative errors, and so on).

As an extension of their model (5.47), Hull and White have also proposed the following model in which a stochastic element is introduced into the drift of the short rate:

$$dr_t = (\theta(t) + u_t - ar_t)\,dt + \sigma_1\,dW_{1,t} \tag{5.59a}$$

$$du_t = -bu_t\,dt + \sigma_2\,dW_{2,t} \tag{5.59b}$$

where $W_{1,t}$ and $W_{2,t}$ are Brownian motions (possibly correlated) under the risk-neutral measure. This is known as the *two-factor Hull-White model* to distinguish it from the model (5.47) which is then referred to as the *one-factor Hull-White model*. The model is affine, with two state variables (not counting the log money market account m_t) and two driving Brownian motions. It can be analyzed in a similar way as the one-factor model; in particular, the function $\theta(t)$ can be chosen to match observed bond prices. The two-factor model has some additional parameters with respect to the one-factor model, namely b, σ_2, and the correlation coefficient ρ of

the two Brownian motions. These parameters can be used to obtain closer matches with for instance swaption prices than would be possible in the one-factor model.

5.6.4 The Heath-Jarrow-Morton model

A role for calendar time in economic models could be envisaged for instance in the agricultural sector, where there is an annual cycle in the availability of crops. Within an interest rate model, appearance of calendar time as one of the determinants of future developments seems less natural. In other words, models are expected to be *time-homogeneous*. This property is however not satisfied by models such as (5.47) and (5.59), due to the appearance of a time-dependent parameter which is in fact essential to achieve the goal of obtaining a match with currently observed bond prices. In defense of these models, it may be argued that if one wants to work with models that have a low state space dimension, then something needs to be adjusted to obtain a close match with the term structure of interest rates as observed in the real world, since one cannot expect that a many-faceted process such as the evolution of the term structure could be adequately described in terms of a model with only one or two state variables.

A modeling framework that is capable of producing time-homogeneous models that still fully match the currently observed term structure was proposed by David Heath, Robert Jarrow and Andrew Morton in 1987.[4] The approach relies on the introduction of an infinity of state variables, namely the forward rates for all maturities. In such a model, the current term structure can be matched exactly by including the current forward curve in the initial condition of the infinite-dimensional state vector.

The Heath-Jarrow-Morton (HJM) model does not fit into the framework of the generic state space model (3.1) that is used in this book, since in (3.1) it is assumed that the number of state variables is finite. Also, the standard formulation of the HJM model parametrizes forward rates in terms of time of maturity, rather than in terms of time to maturity, which means that these rates can be used as state variables only for a finite amount of time, namely up to maturity. A form of the HJM model that is similar to (3.1) could be given as follows:

$$dF_t(T) = \alpha_T(t, X_t)\,dt + \sigma_T(t, X_t)\,dW_t, \qquad t \leq T, \qquad F_0(T) = F_0^m(T) \tag{5.60}$$

where X_t is a state variable that in general includes all of the forward rates as well as other variables that one might want to incorporate (for instance, macro-economic variables). If such additional variables are indeed included, then their dynamics

[4] "Bond pricing and the term structure of interest rates: A new methodology", working paper, Cornell University. A revised version was published in *Econometrica* in 1992 under the title "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation".

should be specified as well. Simulation of forward rates will be simplified if the dependence of the model function $\alpha_T(t,x)$ and $\sigma_T(t,x)$ on the state variable x is limited. The simplest case occurs when these functions do not depend on x at all. In that case, the model (5.60) implies that the forward rates follow correlated Gaussian processes; therefore this version is called the Gaussian HJM model. Due to the parametrization in terms of time of maturity, the requirement for time homogeneity is not that the model functions should not depend on t, but rather that they should depend on calendar time t only through the difference $T-t$. The Brownian motion W_t in the model (5.60) can be a vector, usually taken to be of finite length. The specification of the model is under the risk-neutral measure, so that W_t is a Brownian motion with respect to that measure.

Prices of bonds in terms of forward rates are given by (5.9). Generalizing that formula to a general time t, we can write

$$P_t(T) = \exp\Big(-\int_t^T F_t(s)\,ds\Big). \tag{5.61}$$

This is a specification of prices of traded assets in terms of state variables with dynamics given by (5.60). To ensure absence of arbitrage, the criterion (3.80) can be applied. To make the notation shorter as well as more in line with the original publication, write $\alpha_t(T)$ for the process $\alpha_T(t,X_t)$, and define $\sigma_t(T)$ likewise. Recall (see Section 5.2.3) that $F_t(t) = r_t$, the short rate at time t. Freely using the Leibniz rule[5] for differentiation of integrals, and interchanging integrals as well, one can write

$$\begin{aligned} d(\log P_t(T)) &= -d\Big(\int_t^T F_t(s)\,ds\Big) = F_t(t)\,dt - \int_t^T dF_t(s)\,ds \\ &= \Big(r_t - \int_t^T \alpha_t(s)\,ds\Big)\,dt - \Big(\int_t^T \sigma_t(s)\,ds\Big)\,dW_t. \end{aligned}$$

It then follows from the Itô rule that the drift term of the asset with price $\pi_T = P_t(T)$ $(T \geq t)$ is given by

$$\mu_T = \Big[r_t - \int_t^T \alpha_t(s)\,ds + \tfrac{1}{2}\Big(\int_t^T \sigma_t(s)\,ds\Big)\Big(\int_t^T \sigma_t(s)\,ds\Big)^\top\Big]\pi_T.$$

The condition for absence of arbitrage is therefore

$$\int_t^T \alpha_t(s)\,ds = \tfrac{1}{2}\Big(\int_t^T \sigma_t(s)\,ds\Big)\Big(\int_t^T \sigma_t(s)\,ds\Big)^\top.$$

The condition certainly holds for $T = t$. Therefore, the condition above is satisfied

[5] Gottfried Leibniz (1646–1716), German philosopher, diplomat, mathematician, lawyer, and historian.

when the derivatives with respect to T of the left hand side and the right hand side are equal, i.e.

$$\alpha_t(T) = \sigma_t(T) \int_t^T \sigma_t(s)\, ds, \qquad t \leq T.$$

In terms of the notation employed in (5.60), this means that, to ensure absence of arbitrage, the model functions in (5.60) must satisfy, for all $t \leq T$ and all x,

$$\alpha_T(t,x) = \sigma_T(t,x) \int_t^T \sigma_s(t,x)\, ds. \tag{5.62}$$

A similar result is derived in Section 5.7.2 below in a context in which only rates for a finite number of maturities are involved.

The HJM model is not a model in the same sense as, for instance, the Vasicek model; rather, it is a modeling framework, or, one might say, a recipe for arriving at arbitrage-free term structure models. Its generality is both a strength and a weakness; many specific cases can be covered, but the modeling framework as such does not provide much in terms of guidelines beyond the no-arbitrage condition (5.62). One can also derive models in other ways and use the condition (3.80) for absence of arbitrage, or one of the other equivalent conditions discussed in Chapter 3 as may be convenient in a specific case.

5.7 Partial models

The term structure models discussed so far all aim at describing the evolution of interest rates for all maturities. On the basis of such models, one can price any term structure derivative. It is quite ambitious to try to construct a model that is capable of producing reliable prices and hedge strategies across such a broad range of products. Most of the contracts that are actually traded in the market belong to certain standard types, and one wonder whether it would not be possible to price such contracts on the basis of simplified models. Indeed it is possible to do that; some examples are discussed in this section.

5.7.1 The Black (1976) model

A popular term structure product is the *swaption*. Briefly stated, a swaption is an option on a swap. More in detail, a swaption contract gives the holder the right to enter, at a specified time in the future, a swap contract at a specified swap rate. As discussed in Section 5.2.4, the swap rate in a swap contract is usually set in such a way that the value of the floating leg of the contract, at the time the contract is agreed upon, is equal to the value of the fixed leg at the same time. In this way, at the time that the swap contract is entered, its value to both parties is zero. A

swaption has the nature of an option which may or may not be taken at the holder's discretion, and therefore its value must be positive. The swap rate that is contracted in a swaption is a free parameter that is similar to the strike in a call option contract. The value of the swaption in general depends strongly on the contracted swap rate.

As discussed in Section 5.2.4, a swap contract requires specification of the *tenor dates* $T_0, T_1, \ldots, T_n$, the notional principal V, and the swap rate. It also needs to be specified which party is the "payer" (i.e. pays the fixed rate) and which party is the "receiver" (i.e. pays the floating rate). According to the formulas derived in Section 5.2.4, the value of a payer swap with swap rate r^{c} at time T_0 is given by

$$(1 - P_{T_0}(T_n))V - \sum_{i=1}^{n} P_{T_0}(T_i) r^{\mathrm{c}} V.$$

Therefore, if C_t denotes the value at time $t \leq T_0$ of a payer swaption maturing at T_0, with swap rate r^{c} and notional principal V, then at the time of maturity we have

$$\begin{aligned} C_{T_0} &= \max\left((1 - P_{T_0}(T_n))V - \sum_{i=1}^{n} P_{T_0}(T_i)\, r^{\mathrm{c}} V, 0\right) \\ &= \max(r^{\mathrm{sw}}_{T_0} - r^{\mathrm{c}}, 0) \sum_{i=1}^{n} P_{T_0}(T_i)\, V. \end{aligned} \tag{5.63}$$

An expression for the price at time $t < T_0$ is given by the numéraire-dependent pricing formula

$$\frac{C_t}{N_t} = E_t^{\mathbb{Q}_N}\left[\frac{C_{T_0}}{N_{T_0}}\right]$$

where N_t refers to the value of a numéraire that may be chosen arbitrarily, and $\mathbb{Q}_N$ is the corresponding martingale measure. It is possible to take

$$N_t = \sum_{i=1}^{n} P_t(T_i) \tag{5.64}$$

because the right hand side is the price of a traded asset, namely a portfolio of bonds which all have face value 1 and which mature at times $T_1, \ldots, T_n$. Since the price of each bond is positive, the price of the portfolio is positive as well. This numéraire is called the *annuity factor numéraire*, because the right hand side of (5.64) represents the present value of an annuity that pays one unit of currency at each of the time points $T_1, \ldots, T_n$. The equivalent martingale measure that corresponds to the annuity factor numéraire is called the *swap measure*. The use of the annuity factor numéraire is convenient here, because the NDPF becomes

$$\frac{C_t}{N_t} = E^{\mathbb{Q}_N}\left[\max\left(\frac{1 - P_{T_0}(T_n)}{\sum_{i=1}^{n} P_{T_0}(T_i)} - r^{\mathrm{c}}, 0\right)\right] V = E_t^{\mathbb{Q}_N}[\max(r^{\mathrm{fs}}_{T_0} - r^{\mathrm{c}}, 0)]V \tag{5.65}$$

where r_t^{fs} (the "forward swap rate" for a swap contract to be entered into at time T_0) is defined for $t \leq T_0$ by

$$r_t^{\mathrm{fs}} = \frac{P_t(T_0) - P_t(T_n)}{\sum_{i=1}^{n} P_t(T_i)} \tag{5.66}$$

(compare (5.1)). The expression (5.65) shows that to obtain the swaption price it is not necessary to have a full term structure model. Actually it is enough to have the distribution of $r_{T_0}^{\mathrm{fs}}$ under the martingale measure that corresponds to the numéraire defined in (5.64). As is shown by (5.66), the forward swap rate r_t^{fs} is equal to the price of a traded asset divided by the numéraire defined in (5.64). Therefore, under the corresponding martingale measure $\mathbb{Q}_N$, the process $\{r_t^{\mathrm{fs}}\}$ must be a martingale. Under the assumption that interest rates are positive, so that $P_t(T_n) < P_t(T_0)$, the forward swap rate is positive as well. The lognormal model may then come to mind:

$$dr_t^{\mathrm{fs}} = \sigma r_t^{\mathrm{fs}}\, dW_t \tag{5.67}$$

where W_t is a Brownian motion under $\mathbb{Q}_N$, and σ is a constant. Under this model, the evaluation of the expression in (5.65) is a special case of the calculations that lead to the Black-Scholes formula for option prices. We obtain

$$E_t^{\mathbb{Q}_N}[\max(r_{T_0}^{\mathrm{fs}} - r^{\mathrm{c}}, 0)] = r_t^{\mathrm{fs}}\Phi(d_1) - r^{\mathrm{c}}\Phi(d_2) \tag{5.68a}$$

where

$$d_1 = \frac{\log(r_t^{\mathrm{fs}}/r^{\mathrm{c}}) + \frac{1}{2}\sigma^2(T_0 - t)}{\sigma\sqrt{T_0 - t}}\,, \quad d_2 = \frac{\log(r_t^{\mathrm{fs}}/r^{\mathrm{c}}) - \frac{1}{2}\sigma^2(T_0 - t)}{\sigma\sqrt{T_0 - t}}\,. \tag{5.68b}$$

The formula for the swaption price that is obtained from (5.68) and (5.65) is usually called the *Black (1976) formula*, because Black has published a formula similar to (5.68) in a 1976 paper on the pricing of options on futures contracts.

Given the contract specifics (time of maturity T_0, tenor dates $T_1, \ldots, T_n$, contracted swap rate r^{c}) and the current term structure, the expression (5.68) gives a one-to-one relation between the parameter σ (volatility of the forward swap rate) and the price of the swaption. It has become customary in the market to quote the swaption price in terms of the parameter σ, which is referred to as the *Black (1976) implied volatility*. Usually one works with a standardized frequency of payments so that the tenor dates are specified just by T_n. The implied volatilities are quoted in the market on a two-dimensional grid, made up from a range of values for T_0 and a range of values for T_n, with the contract swap rate set to the current par swap rate.

5.7.2 LIBOR market models

LIBOR market models[6] focus on products that generate payments at a limited set of dates (typically tenor dates). They are called "market models" because they use state variables that are directly observable in the market. These models therefore allow the use of relatively large numbers of state variables (10 or more), which generates flexibility in a different way than by the device, used for instance in the Hull-White model, of allowing model parameters to depend on calendar time.

An example of a typical observable variable is the rate at which London banks can agree to borrow funds on January 1 of next year and to return these, with interest, on the following July 1. Such forward rates, for particular tenor dates, are available on an ongoing basis. Let $L_t(S,T)$ denote the simply compounded forward rate that is quoted at time t for loans to be made at a tenor date S and to be returned at date T. By an argument similar to the one used above for continuous compounding, the relation of the forward rate $L(S,T)$ to the prices of bonds maturing at times T and S respectively is given by

$$L(S,T) = \frac{P(S) - P(T)}{(T-S)P(T)}. \tag{5.69}$$

The actual value of the factor $T - S$ is determined by a "day count convention".

To formulate a LIBOR market model, one starts by selecting a number of successive tenor dates for which LIBOR rates are available, say $T_1, \ldots, T_{n+1}$. Let ΔT_i $(i = 1, \ldots, n)$ denote the difference between T_{i+1} and T_i according to a given day count convention. The state variables in LIBOR market models are the LIBOR forward rates corresponding to the periods from T_i to T_{i+1}. Let us denote these rates by L_i $(i = 1, \ldots, n)$; then we have an n-dimensional state variable L_t. The LIBOR rates themselves are not prices of tradable assets, but they can be related to the prices of bonds by the formula (5.69). Writing $P_{i,t}$ for the price at time t of the bond maturing at the tenor date T_i, we have (suppressing the subscript t, but showing the index i that relates to the tenor date):

$$L_i = \frac{P_i - P_{i+1}}{\Delta T_i \, P_{i+1}} \tag{5.70}$$

or in other words

$$P_i = (1 + \Delta T_i \, L_i) P_{i+1}. \tag{5.71}$$

[6] The term "LIBOR" means London InterBank Offered Rate, i.e. the interest rate paid by London commercial banks on deposits by other banks. This rate is a frequently used benchmark for floating rates.

It follows that we can write successively

$$\frac{P_n}{P_{n+1}} = 1 + \Delta T_n \, L_n$$
$$\frac{P_{n-1}}{P_{n+1}} = \frac{P_{n-1}}{P_n}\frac{P_n}{P_{n+1}} = (1 + \Delta T_{n-1} \, L_{n-1})(1 + \Delta T_n \, L_n)$$

and in general

$$\frac{P_i}{P_{n+1}} = \Pi_{j=i}^{n}(1 + \Delta T_j \, L_j). \tag{5.72}$$

So we can express at least fractions of asset prices in terms of the state variables. Here we took the price of the bond maturing at time T_{n+1} as a common denominator; other choices are possible as well. The equivalent martingale measure that corresponds to taking P_{n+1} as the numéraire is known as the *terminal measure*. This choice of numéraire is convenient for contracts that mature at the tenor date T_{n+1}, since obviously $P_{n+1,T_{n+1}} = 1$ so that the pricing formula for the value at time t of a contract C that pays $F(L_{T_{n+1}})$ at time T_{n+1} becomes

$$\begin{aligned}\pi_C(t,\ell) &= P_{n+1,t} E^{\mathbb{Q}}\Big[\frac{F(L_{T_{n+1}})}{P_{n+1,T_{n+1}}} \,\Big|\, L_t = \ell\Big] \\ &= P_{n+1,t} E^{\mathbb{Q}}[F(L_{T_{n+1}}) \mid L_t = \ell]\end{aligned} \tag{5.73}$$

where $P_{n+1,t}$ is at time t a known quantity. Also contracts maturing at tenor dates T_i other than T_{n+1} can be priced in terms of the LIBOR rates $L_1, \ldots, L_n$ since

$$\frac{1}{P_{n+1,T_i}} = \frac{P_{i,T_i}}{P_{n+1,T_i}} = \Pi_{j=i}^{n}(1 + \Delta T_j \, L_{j,T_i}).$$

So, at least for contracts maturing at tenor dates, it is sufficient to use the LIBOR rates as state variables, and to model only *relative* prices.

Let us now consider models for the joint evolution of the LIBOR rates that take the standard state-space form

$$dL = \mu_L dt + \sigma_L dW. \tag{5.74}$$

The variable W that appears here represents a k-dimensional Brownian motion under the terminal measure. The choice of k (the *number of factors*) is a modeling decision; in principle k can have any value between 1 and n, but in practice often fairly low values of k are chosen (one to three). We can price contracts maturing at tenor dates as soon as we specify the functions μ_L and σ_L, which in general may depend on calendar time and on the vector of LIBOR rates itself; but of course these functions have to be chosen in such a way as to preclude arbitrage. Since we have a model here that determines relative asset prices rather than absolute

prices, we may apply Thm. 3.2.2 rather than Thm. 3.2.3; moreover, since we already assume that W is a Brownian motion under the terminal measure that corresponds to the numéraire P_{n+1}, the condition of the theorem simply becomes that the relative asset prices should be martingales. The question is what conditions this requirement imposes on the functions μ_L and σ_L.

The relative asset prices are given by (5.72); in particular they do not depend directly on time. So the situation we have is of the form

$$\begin{aligned} dX &= \mu_X dt + \sigma_X dW \\ Y/N &= \pi(X) \end{aligned}$$

where we write simply $\pi(X)$ (with component functions $\pi_i(X)$) instead of $\pi_{Y/N}(X)$. Because the state variables in the LIBOR market models are defined as quantities that are in close relationship with traded assets (cf. (5.70)), the function $\pi(X)$ is already given in the model, rather than that it needs to be computed on the basis of the absence-of-arbitrage requirements such as for instance in the case of the Vasicek model. Instead, the requirements of absence of arbitrage are reflected in a certain relation that must hold between the drift and the volatility parameters μ_X and σ_X in a LIBOR model.

By the fundamental theorem of asset pricing, absence of arbitrage holds if the vector process Y/N is a $\mathbb{Q}_N$-martingale; this means that all of its components Y_i/N are martingales. The condition for Y_i/N to be a martingale is

$$\frac{\partial \pi_i}{\partial x}\,\mu_X + \tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial^2 \pi_i}{\partial x^2}\,\sigma_X \sigma_X^\top = 0. \tag{5.75}$$

In the situation of the LIBOR market model, the length of the vector Y is the same as the length of the vector X. Moreover, it follows from (5.72) that the $n \times n$ matrix $\partial\pi/\partial x$ with rows $\partial\pi_i/\partial x$ is upper triangular, with nonzero entries on the main diagonal. Therefore, the equations (5.75) for $i = 1, \ldots, n$ can be summarized in the statement that the vector process Y/N is a martingale if and only if the functions μ_X and σ_X are related by

$$\mu_X = -\Big(\frac{\partial \pi}{\partial x}\Big)^{-1} \Big[\tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial^2 \pi_1}{\partial x^2}\,\sigma_X \sigma_X^\top \quad \cdots \quad \tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial^2 \pi_n}{\partial x^2}\,\sigma_X \sigma_X^\top\Big]^\top. \tag{5.76}$$

This shows in particular that the volatitility σ_X can be freely chosen, and that the drift μ_X is completely determined by the martingale requirement once σ_X has been selected.

The equation (5.76) can be used to determine the relation between μ_L and σ_L in the LIBOR market model, but the calculations are a bit complicated. One may also reason as follows. In general, if Y and Z are processes that are expressed in terms

of a state process by functions π_Y and π_Z, and if both Y and Z are martingales, then

$$\mu_{Y/Z} = -\sigma_{Y/Z}\,\frac{\sigma_Z^\top}{\pi_Z} \tag{5.77}$$

(this is a special case of formula (3.33)). Now note from (5.70) that

$$L_i = \frac{1}{\Delta T_i}\Big(\frac{P_i}{P_{i+1}} - 1\Big) = \frac{1}{\Delta T_i}\Big(\frac{P_i/P_{n+1}}{P_{i+1}/P_{n+1}} - 1\Big) \tag{5.78}$$

and that both P_i/P_{n+1} and P_{i+1}/P_{n+1} are martingales under the terminal measure. Let us write π_i for the (relative) pricing function of P_i/P_{n+1} and σ_i for the corresponding volatility function. We have

$$\sigma_i = \frac{\partial \pi_i}{\partial L}\,\sigma_L = \sum_{j=1}^{n} \frac{\partial \pi_i}{\partial L_j}\,\sigma_{L_j} = \sum_{j=i}^{n} \frac{\partial \pi_i}{\partial L_j}\,\sigma_{L_j} \tag{5.79}$$

where the last equality holds because P_i/P_{n+1} does not depend on L_j for $j < i$. From the specific form of π_i as given in (5.72), it follows that, for $j \geq i$,

$$\frac{\partial \pi_i}{\partial L_j} = \frac{\Delta T_j}{1 + \Delta T_j\, L_j}\,\pi_i. \tag{5.80}$$

Combining (5.77), (5.79), and (5.80), we find

$$\mu_{L_i} = -\sigma_{L_i} \sum_{j=i+1}^{n} \frac{\Delta T_j}{1 + \Delta T_j\, L_j}\,\sigma_{L_j}^\top. \tag{5.81}$$

This determines μ_L once σ_L has been given. It follows from (5.76) that there are no other conditions to be fulfilled, so the condition (5.81) is both necessary and sufficient for the LIBOR market model to be arbitrage-free.

The LIBOR market model still leaves a great deal of freedom, since the number of factors k and the $n \times k$ matrix $\sigma_X = \sigma_X(t, L)$ can in principle be arbitrarily specified. Often a special form is chosen of the type

$$\sigma_{L_i}(t, L) = \sigma_i(t) L_i R_i \tag{5.82}$$

where $\sigma_i(t)$ is a scalar function of calendar time and R_i is a constant row vector. The time function $\sigma_i(t)$ is often chosen to be piecewise constant with jumps at the tenor dates. The model is then finitely parametrized, and the parameters can be determined by a combination of estimation on the basis of time series data and calibration by matching observed prices of derivatives.

5.8 Exercises

1. Show that the discount curve $P(T)$ can be expressed in terms of the swap curve $S(T)$ by means of the differential equation

$$P'(T) = -\Big[P(T)S(T) + (1 - P(T)) \frac{S'(T)}{S(T)} \Big] \tag{5.83}$$

with initial condition $P(0) = 1$.

2. Prove that, for $T_2 > T_1$, one can write

$$F_0(T_1, T_2) = \frac{1}{T_2 - T_1} \log E^{\mathbb{Q}_{T_2}} \big[\exp\big((T_2 - T_1) R_{T_1}(T_2) \big) \big] \tag{5.84}$$

where $\mathbb{Q}_{T_2}$ denotes the T_2-forward measure.

3. At time 0, mr. Balding and ms. Young decide to form a joint investment fund. The purpose of the fund is to provide them with money (in the form of a lump sum) at retirement. Mr. Balding and ms. Young will retire at time T_1 and time T_2, respectively, with $T_1 < T_2$. The amounts that are contributed by them at time 0 are denoted by A_0^1 and A_0^2. The value of the joint fund at time t is denoted by A_t. Up to time T_1, the fund's administrator communicates to mr. Balding and ms. Young a number called the "accounting value of the benefit" b_t^i. This number is continually adjusted in response to investment returns and changes in interest rates, in such a way that at all times t the following relation holds:

$$P_t(T_1) b_t^1 + P_t(T_2) b_t^2 = A_t. \tag{5.85}$$

At time 0, the accounting values are defined by $b_0^i = A_0^i / P_0(T_i)$ so that (5.85) is satisfied at $t = 0$. At later times, the accounting values are defined by $b_t^i = \alpha_t b_0^i$ where α_t is a common adjustment factor that is chosen such that (5.85) holds. At time T_1, the amount that mr. Balding receives is $b_{T_1}^1$.

a. Show that, if $P_{T_1}(T_2)$ as seen from time 0 is a deterministic quantity (i.e. the discount factor for maturity T_2 at time T_1 can be exactly predicted at time 0), then the time-0 value of the amount received by mr. Balding at time T_1 is equal to his contribution, namely A_0^1.

b. Now assume that $P_{T_1}(T_2)$ is a nondegenerate random variable, and that $P_{T_1}(T_2)$ and A_{T_1}/A_0 are independent under the T_1-forward measure. Prove that in this case the time-0 value of the amount received by mr. Balding is larger than A_0^1. In other words, the decision by mr. Balding and ms. Young to form a joint fund under the stated conditions entails a net transfer of value from ms. Young to mr. Balding.

[*Hint*: use the property $E^{\mathbb{Q}_{T_1}} P_{T_1}(T_2) = P_0(T_2)/P_0(T_1)$ and Jensen's inequality.][7]

4. **a**. Show that, if the short rate is constant in time, then the yield curve is constant across maturities (and hence also constant in time).

b. Conversely, show that if in an arbitrage-free model at all times the yields are constant across maturities, then the yields for all maturities are constant in time. [*Hint*: a model in which at all times the yield curve is constant across maturities produces bond prices that depend on calendar time t and maturity date T through

$$P_t(T) = \exp\big(-r_t(T-t)\big)$$

where r_t is a scalar process. Assume that r_t satisfies an equation of the form $dr_t = \mu_r\, dt + \sigma_r\, dW_t$ under the risk-neutral measure, and apply the criterion (3.80) for absence of arbitrage.]

5. Suppose that a regulatory authority asks an institution to report on the evolution of its assets within a single scenario for interest rates. Current time is 0, the projection date is T_1 (for instance one year from now). The authority might for example prescribe a scenario in which $R_{T_1}(T_2) = R_0(T_2 - T_1)$ for all $T_2 \geq T_1$; this would mean that the yield curve stays the same. Show that, if it assumed that this scenario indeed occurs with probability 1, then an arbitrage opportunity exists unless the prescribed scenario is such that

$$R_{T_1}(T_2) = F_0(T_1, T_2). \tag{5.86}$$

Also show that no arbitrage is possible when, for $t \geq 0$, the projected yield curve is defined by $R_t(T) = F_0(t, T)$ where $T \geq t$.

6. Suppose that the short rate follows a deterministic function of time, say $r_t = r(t)$ where $r(\cdot)$ is a given function. On the basis of (5.23), compute the corresponding yield curve and the corresponding forward curve.

7. Fig. 4.4 suggests that the Vasicek model is capable of producing term structures that are approximately flat. Is the model (with nonzero volatility) capable of generating a yield curve that is exactly flat (same yield for all maturities)?

8. **a.** The stochastic differential equation (4.11) that is satisfied by the short rate in the Vasicek model generates a *stationary distribution*, which does not depend on the initial condition. Since the SDE in (4.11) is linear, the stationary distribution is normal. Find the mean and the standard deviation of this distribution when $a = 0.5$, $b = 0.06$ (parameter value in the specification under the risk-neutral measure), and

[7] Johan Jensen (1859-1925), Danish engineer. Jensen did not have a formal degree in mathematics, but still wrote mathematical papers in his spare time while working as a telephone engineer (a booming technology in his days) in Copenhagen.

$\sigma = 0.02$. [*Hint*: see Section 2.6.] To what extent does the value of the parameter b play a role?

b. Another parameter that can be associated to the model (4.11) is the so called *half-life*, which is the time it would take for a deviation of the equilibrium value to be reduced to one-half of its original size, if there would be no disturbances (i.e. if σ would be zero). Determine the half-life of the model with parameters as given in part a. of this exercise. Is the value of the half-life affected by the parameter b?

c. The Vasicek model for the short rate is an affine model and so it is possible that in this model the short rate becomes negative. If the parameter values are as in part a., how often does this happen on average in the long run?

9. Suppose that an affine term structure model of the form (5.36) is given with

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -a & a \\ 0 & 0 & a \end{bmatrix}, \quad h = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Compute the functions that have variable weights in the corresponding yield curve, and compare to the Nelson-Siegel model.

10. A portfolio strategy can be defined at follows. Starting with a given capital at time 0, use the money to buy bonds with maturity T. At time ΔT (where $\Delta T < T$), sell these bonds and buy new ones with maturity $T + \Delta T$. At time $2\Delta T$, sell those and buy new ones with maturity $T + 2\Delta T$; and so on. This is called "rolling over the bonds". Theoretically, one can let ΔT tend to zero and create in this way what might be called a *constant-maturity bond*.

a. Let V_t denote the value at time t of a constant-maturity bond. Suppose that a term structure model is given in the form

$$\begin{aligned} dX_t &= \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t \\ P_t(T) &= \pi_T(t, X_t) \qquad (T \geq t) \\ r_t &= h(t, X_t) \end{aligned}$$

where W_t is a Brownian motion under the risk-neutral measure $\mathbb{Q}_M$. Write a stochastic differential equation for V_t under $\mathbb{Q}_M$. [*Hint*: the relation (3.11) implies that the volatility of a self-financing portfolio at time t is determined by the portfolio composition at time t and the volatility at time t of the assets from which the portfolio is constructed. Feel free to use the same relation even in the present case of a portfolio strategy that uses infinitely many assets.]

b. In particular, find the SDE that describes the evolution of V_t under $\mathbb{Q}_M$ in the

case of the Vasicek model. Under the assumption that the price of risk is constant, also find the SDE that describes the evolution of V_t under $\mathbb{P}$.

11. Consider a swap with swap rate 5.5%, annual payments, a principal of 100, and 5 years to maturity. Assume that the term structure is as prescribed by the Vasicek model with parameters as in Exc. 8.a, and that the short rate is 5%. Compute the current value of the swap to the party that pays the floating rate.

12. The derivation of the formula for the yield curve under the Vasicek model is a somewhat tedious and error-prone affair, and this becomes only worse in higher-dimensional affine models. To get confirmation that a proposed formula is indeed correct, bond prices can also be obtained by simulation, on the basis of formula (4.22) in the course notes. It has to be taken into account that prices obtained from simulation are subject to Monte Carlo error as well as to discretization error. Both issues are addressed in the exercise below.

a. Simulate 10^4 joint trajectories of the short rate and the money market account according to the Vasicek model, with parameters as in Exc. 3.a and initial short rate $r_0 = 0.02$. Take $T_{\text{hor}} = 30$ as the simulation horizon, and use time step $\Delta t = 0.1$ for the Euler discretization. Compute bond prices for maturities $T = \Delta t$, $T = 2\Delta t$, ..., $T = T_{\text{hor}}$ on the basis of the formula $P(T) = E^{\mathbb{Q}_M}[1/M_T]$. Plot the corresponding yield curve together with the yield curve as given by (4.34). [The amount of 10^4 scenarios should be enough to give you results that are accurate up to a few basis points.[8]]

b. For each maturity $T \in \{\Delta t, 2\Delta t, \ldots, T_{\text{hor}}\}$, compute the standard deviation of the results obtained by simulation, and use this to compute an approximate confidence interval for the yields obtained from simulation. (Different confidence intervals are associated to different maturities.) Plot the *difference* of the yield curve obtained from simulation and the yield curve according to (4.34), and in this plot also show the bounds that are obtained from the confidence intervals for all maturities.

To improve the behavior of the simulation for small maturities, an adaptation to the Euler scheme can be made as follows. First of all, simulation of the money market account may be replaced by simulation of the log money market account $L_t := \log M_t$, which satisfies the stochastic differential equation

$$dL_t = r_t\, dt.$$

The usual Euler discretization for this SDE is

$$L_{t+\Delta t} = L_t + r_t \Delta t$$

[8] A basis point is 1/100th of a percentage point.

but, if we first update r and then update L, we can replace this by

$$L_{t+\Delta t} = L_t + \tfrac{1}{2}(r_t + r_{t+\Delta t})\Delta t.$$

c. Redo part b. using the new variable L_t and the new discretization scheme.

13. A *swaption* is a contract that gives the holder the right to enter, at a given time T in the future, a swap contract with specified parameters. We consider here the valuation of a swaption under the following conditions:

- the swaption has an option maturity of $T = 1$ year, and the current short interest rate is $r_0 = 0.05$;
- the swap contract that the swaption refers to is a five-year swap with yearly payments in which a fixed rate of 6% is received, and the value of the principal is one million euro;
- valuation is done under the Vasicek model with the following parameters: $a = 0.2$, $b = 0.06$ (parameter value in the specification under the risk-neutral measure), and $\sigma = 0.02$.

Several approaches can be used to determine the price of the swaption. Parts a. and b. of this exercise call for the construction of auxiliary functions that are of use in several different methods.

a. Write code that will compute bond prices in a Vasicek model with given parameters, for a range of maturities $T_1, \ldots, T_m$ and for a range of values of the short rate $r_1, \ldots, r_m$. The code should allow for vector inputs $[T_1 \cdots T_n]$ and $[r_1 \cdots r_m]$, and should produce as output a matrix which has as its (i, j)-th element the bond price for maturity T_j when the short rate is r_i.

b. Write code that computes swap values in a Vasicek model with given parameters, for a range of values of the short rate $r_1, \ldots, r_m$. Assume that current time is the initiation date T_0, and that the tenor dates are equally spaced. Take as input parameters the model parameters of the Vasicek model, the notional principal, the number n of future tenor dates, the distance between successive tenor dates, the fixed rate in the swap contract, and a vector $[r_1 \cdots r_m]$ of possible values of the short rate at the time at which the swap contract is initiated. The output is an m-vector of corresponding swap values.

The swaption payoff at time $T = 1$ is given by

$$C_T = \max(S_T, 0) \tag{5.87}$$

where S_T is the value at time T of the swap that may be initiated at that time

if the swaption holder chooses to do so. In the Vasicek model, the only uncertain factor that influences the swap value is the short rate r_T, so that the swaption can be viewed as an option on the short rate. The random variable S_T can be written as $S_T = f(r_T)$ where $f(\cdot)$ is the function that is implemented in part b. above. The swaption can be priced by the formula

$$\frac{C_0}{M_0} = E^{\mathbb{Q}_M}\left[\frac{C_T}{M_T}\right] = E^{\mathbb{Q}_M}\left[\frac{\max\big(f(r_T), 0\big)}{M_T}\right]. \tag{5.88}$$

This can be evaluated numerically by the Monte Carlo method, on the basis of an Euler discretization of the joint stochastic differential equations (5.37) and (5.19).

c. Determine the value of the swaption by the Monte Carlo method as suggested above. Use 10^4 trajectories and take step size 0.01 for the Euler discretization. Determine an estimated value as well as a confidence interval.

An alternative formula for the swaption value can be obtained by using the T-bond as a numéraire, rather than the money market account:

$$C_0 = P_0(T)\, E^{\mathbb{Q}_T}\left[\frac{C_T}{P_T(T)}\right] = P_0(T)\, E^{\mathbb{Q}_T}[\max(f(r_T), 0)]. \tag{5.89}$$

The advantage of this formula is that it requires only the distribution of r_T and not of any other variables. To achieve this, we do need to switch from the specification under the risk-neutral measure to the specification under the forward measure.

d. Use the change-of-numéraire formula (3.71) to write a stochastic differential equation for r_t under the T-forward measure. You may find it convenient to use the expression (3.75).

e. Use an Euler discretization of the SDE that you found under d. to compute a Monte Carlo approximation of the swaption price on the basis of the formula (5.89). Use 10^4 trajectories and take step size 0.01 for the Euler discretization. Determine an estimated value as well as a confidence interval.

With a bit more effort, the Euler discretization can be discarded.

f. Use the formulas in Section 2.6.3 to give an explicit description of the distribution of r_T under the T-forward measure. [*Hint*: the distribution is normal, so it is sufficient to give the mean and the variance.]

g. Determine the price of the swaption again by the Monte Carlo method on the basis of the expression (5.89), but this time using draws from the distribution of r_T directly. Use 10^4 samples; determine an estimated value as well as a confidence interval.

Since the swaption value is given in (5.89) in terms of the expectation of a function

of a variable with known distribution, it is possible to use a numerical quadrature method (cf. Exc. 4.5.7) instead of a Monte Carlo method. In the present case this comes down to replacing an expectation of the form $E[g(Z)]$, where Z is a standard normal variable and g is a given function, by an expectation of the form $E[g(\hat{Z})]$ where $\hat{Z}$ is a discrete random variable that has approximately the same distribution as Z.

h. Evaluate the expectation in (5.89) by numerical quadrature. To approximate the normal random variable, construct an evenly spaced grid of 100 points ranging from -4 to 4, and define a discrete random variable that takes values in the grid points with probabilities chosen such as to approximate the standard normal distribution. Compute the corresponding swaption value. To assess the accuracy of the answer, also compute the value when the grid range is extended from $[-4, 4]$ to $[-5, 5]$, and when 1000 grid points are used instead of 100.

14. A bank has a traditional mortgage contract available for customers in which interest rate payments are fixed for ten years. Payments are to be made monthly, and the monthly amount to be paid is determined as one-twelfth of the ten-year interest rate times the amount of the loan. The ten-year rate that is used is the one that holds at the starting time of the mortgage. The bank managers are considering to introduce an additional feature to this contract: after five years, customers may choose to reset the interest rate to the five-year rate that holds at that time. In return for this privilege, customers should accept a certain surcharge; for instance, if the ten-year rate at the initiation of the contract is 5% and the surcharge is 1%, the monthly payments will be one-twelfth of 6% of the amount of the loan, rather than one-twelfth of 5%. If after five years the reset option is not used, the surcharge will remain in force for the rest of the duration of the contract. When a reset does take place, the surcharge is no longer paid.

The managers want to know which value of the surcharge is such that the value of the contract with the reset option is the same as the value of the traditional contract. Determine the surcharge that satisfies this condition. Use the following assumptions:

- only payments during ten years are considered (the mortgage will be renegotiated after that time);
- in each version of the contract, the first monthly payment is made one month after the rate has been (re)set;
- the evolution of the term structure is described by the Vasicek model with $a = 0.5$, $b = 0.04$, $\sigma = 0.03$, and the price-of-risk parameter is equal to $\lambda = -0.2$;

- the current short rate is $r = 0.04$.

Use a Monte Carlo method in combination with the numéraire-dependent pricing formula; take the money market account as a numéraire. Find the surcharge by trial and error, at the level of accuracy of basis points. You may find it useful to write separate routines that perform the following operations:

- compute, for a given value of the short rate r, the interest rate for a given maturity according to the Vasicek model;
- compute, for a given value of r, the current value of a series of constant monthly payments on the basis of the term structure that is connected to r through the Vasicek model.

15. Consider the following contracts for interest payments on a loan. The borrower is referred to below as the customer, and the lender (the institution that provides the loan) is called the bank.

- Contract C_0, with parameters A ("principal"), r_p ("rate paid"), and T_f ("final payment date"): monthly payments of $(r_p/12)A$ until time T_f. The last payment is made at time T_f.
- Contract C_1, with parameters A, r_p, T_f, and T_1 ("reset date"): same as above, but at time T_1 the customer has the option of replacing the rate r_p, for payments after time T_1, by the rate that holds at time T_1 for loans that mature at time T_f. If this option is chosen, the customer has to pay 1% commission (i.e. the amount $0.01A$) at time T_1 to the bank.
- Contract C_2, with parameters A, r_p, T_f, T_1 ("first reset date"), and T_2 ("second reset date"). The contract is the same as above, except that the interest rate may now be reset both at time T_1 and at time T_2. The new rate is in both cases given by the rate that holds at the time of reset for loans that mature at time T_f. At each time at which the reset option is used, the customer pays 1% commission.
- Contract C_3. Same as above, except that at most one reset is allowed; in other words, if the reset option is used at time T_1, then there cannot be a second reset at time T_2. One percent commission is paid if the option is used; this payment is then made at the time of use (either T_1 or T_2).

Assume that a term structure model is available which describes the dynamics of the vector of state variables X_t under an equivalent martingale measure $\mathbb{Q}_N$ which corresponds to a chosen numéraire $N_t = \pi_N(t, X_t)$. The model leads to a function

π_T such that the value at time t of a default-free zero-coupon bond that pays 1 at time T is given by

$$P_t(T) = \pi_T(t, X_t).$$

Let the function $\pi_{C_0}(A, r_p, T_f; t, X_t)$ denote the value at a given time $t < T_f$ of *future* payments to the bank resulting from contract C_0 (i.e. payments that will be made after time t). For brevity the parameters may sometimes be omitted, so that we can write simply $\pi_{C_0}(t, X_t)$. The functions π_{C_1}, π_{C_2}, and π_{C_3} are defined analogously.

a. Give an expression for the function $\pi_{C_0}(t, x)$, for $t < T_f$, in terms of the function $\pi_T(t, x)$.

b. Give an expression for the function $\pi_{C_1}(t, x)$, for $t < T_1$, in terms of the functions $\pi_T(t, x)$, $\pi_N(t, x)$, and $\pi_{C_0}(t, x)$.

c. Give expressions for the functions $\pi_{C_2}(t, x)$ and $\pi_{C_3}(t, x)$, for $t < T_1$, in terms of the functions $\pi_T(t, x)$, $\pi_N(t, x)$, $\pi_{C_0}(t, x)$, and $\pi_{C_1}(t, x)$.

d. Assume that the Vasicek model holds with the same parameter values as in Exc. 14, and that the current short rate is 4%. Let r_p be equal to the 15-year rate that is given by the model. Compute the value at time 0 of all four contracts mentioned above, assuming the following parameter values: $A = 100$, $T_1 = 5$, $T_2 = 10$, $T_f = 15$.

16. In this exercise we consider a risk management problem for a firm that for its activities needs to keep a constant amount funded. Say that, for the coming ten years, the firm wants to have the amount 1000 available. Among the many ways of achieving this, we may in particular consider the following alternatives.

(i) Long-term funding: the firm borrows 1000 for ten years and pays a constant amount of interest each month, where the interest rate is set such that the amount to be returned after ten years is exactly 1000.

(ii) Short-term funding: each month, the firm borrows 1000 for one month and pays the one-month interest rate at the end of the month. The amount of 1000 is returned at the end of the month but is immediately borrowed again (refinancing of the loan). The firm keeps on doing this for ten years.

All payments are discounted to current time on the basis of the current term structure; in this way one can compute the total cost of interest payments, which is deterministic in the case of long-term financing but stochastic when short-term funding is used. Assume that the Vasicek model holds with the following parameter values: $a = 0.8$; $b = 0.04$ (parameter value under the real-world measure); $\sigma = 0.01$. The current short rate is given by $r_0 = 0.04$. Under the assumption that the price of interest rate risk λ is equal to 0, compute

(i) the total cost of interest payments in the case of long-term funding

(ii) the expectation and the standard deviation (under the real-world measure) of the total cost of interest payments in the case of short-term funding.

Also produce a histogram to show the distribution of the total cost of interest payments under short-term funding. For the numerical evaluations, you can use the Monte Carlo method with 10^4 scenarios; give an approximate confidence interval for the expected total costs of interest payments.

Repeat the computations under the assumption that $\lambda = -0.5$, and once more under the assumption that $\lambda = 0.5$. Comment on the impact of the value of λ on the decision whether to rely on short-term funding or long-term funding.

Chapter 6

Finite-difference methods

As we have seen, there are several general option pricing methods, based on either the equivalent martingale measure, the Black-Scholes partial differential equation, or the pricing kernel. Depending on the dynamics of the given model and the nature of the option to be priced, it is possible in a number of cases to obtain analytical solutions of the pricing equations as discussed in Chapter 4. However, in many cases no analytical solution is known and consequently one has to apply numerical methods to find prices and hedging strategies. Each of the option pricing methods suggests a computational approach. The Black-Scholes partial differential equation is

$$\frac{\partial \pi}{\partial t} + \frac{\partial \pi}{\partial x}\mu_X + \tfrac{1}{2}\,\mathrm{tr}\,\frac{\partial^2 \pi}{\partial x^2}\sigma_X \sigma_X^\top - r\pi = \frac{\partial \pi}{\partial x}\sigma_X \lambda \tag{6.1}$$

together with specific boundary conditions for each given contract. This is a partial differential equation of a type that also occurs frequently in problems of physics and engineering. A large body of knowledge exists concerning numerical methods for solving such PDEs. The pricing formula based on the equivalent martingale measure is

$$\frac{C_t}{N_t} = E_t^{\mathbb{Q}_N}\left[\frac{C_T}{N_T}\right] \tag{6.2}$$

where C_t denotes the claim price at time t, and N_t is a numéraire. If we can sample from the distribution of C_T/N_T under the risk-neutral measure $\mathbb{Q}$ associated to the numéraire, then a simple way to obtain an approximation of the expected value is to average over a large number of samples. This is a basic version of the Monte Carlo method. The same method can be applied to the characterization of the price in terms of the pricing kernel.

In this chapter we concentrate on PDE-based numerical methods; the Monte Carlo method and its ramifications will be discussed in Chapter 7. Discretization of partial differential equations is a subject of long-standing interest in physics and engineering. Generally speaking there are two approaches that start from different philosophies. In the *finite-difference method*, emphasis is on the approximation

of the differential operators that occur in a given partial differential equation. The *finite-element method* starts from a representation of approximate solutions in terms of suitably chosen basis functions. The latter approach is the method of choice in many engineering problems where complex geometries play a role, for instance in computing the effects of disturbances on mechanical structures such as trucks or suspension bridges. Financial applications usually are not complicated in this respect, although they may very well be complicated in other respects (such as high dimension, separation into stages, free boundaries). Below we concentrate on the finite-difference method.

Two important issues that may arise in the application of finite difference methods are the following:

- instability
- the "curse of dimensionality".

When instability arises, finite difference methods may go completely astray and produce results that are very far from the truth. The issue can be handled however, as will be discussed below. The quick increase of complexity with dimension is an intrinsic property of finite-difference methods, and is a notable difference with Monte Carlo methods. In this chapter, the discussion of finite difference method will be limited to the one-dimensional case, that is, concerning models that have only one state variable. Even if the techniques presented below apply in principle also to higher-dimensional situations, the size of the matrices involved tends to become prohibitive when these techniques are used in three or more dimensions.

6.1 Discretization of differential operators

The finite-difference method is based on the approximation of derivatives by finite difference quotients. The approximation formulas that are needed may be derived from the Taylor[1] expansion. To illustrate this in the case of a function of one variable, note that for a sufficiently smooth function f, one may write

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \tfrac{1}{2}f''(x)(\Delta x)^2 + \tfrac{1}{6}f^{(3)}(x)(\Delta x)^3 + \cdots . \tag{6.3}$$

Replacing Δx by $-\Delta x$, we also have

$$f(x-\Delta x) = f(x) - f'(x)\Delta x + \tfrac{1}{2}f''(x)(\Delta x)^2 - \tfrac{1}{6}f^{(3)}(x)(\Delta x)^3 + \cdots . \tag{6.4}$$

[1] Brook Taylor (1685–1731), British mathematician.

From (6.3), we can write

$$f'(x) = \frac{f(x+\Delta x) - f(x)}{\Delta x} + O(\Delta x). \tag{6.5}$$

This formula is a straightforward approximation of a differential quotient by a difference quotient. A more accurate approximation may be obtained by subtracting (6.3) and (6.4) from each other:

$$f'(x) = \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} + O\big((\Delta x)^2\big) \tag{6.6}$$

where now the error term is of order $(\Delta x)^2$ rather than of order Δx as in (6.5), since the quadratic terms in (6.3) and (6.4) cancel when the two formulas are subtracted. The formula (6.6) is called the *symmetric difference formula*. Formulas that are asymptotically (as Δx tends to zero) even more accurate can be obtained by using more values of f, for instance $f(x+2\Delta x)$ and $f(x-2\Delta x)$; usually however the formulas (6.5) or (6.6) already provide enough accuracy.

A formula for the second derivative may be obtained by adding the expressions (6.3) and (6.4). One gets

$$f''(x) = \frac{f(x+\Delta x) - 2f(x) + f(x-\Delta x)}{(\Delta x)^2} + O\big((\Delta x)^2\big) \tag{6.7}$$

where the error term has order $(\Delta x)^2$ because the cubic terms cancel when (6.3) and (6.4) are added.

The same technique may be applied to obtain partial derivatives of functions of several variables. For instance, if f is a function of x_1 and x_2, then

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{f(x_1+\Delta x_1, x_2) - f(x_1, x_2)}{\Delta x_1} + O(\Delta x_1) \tag{6.8}$$

and so on.

Finite-difference formulas are typically *linear* in the function values that they use. This is natural, since the operation of differentiation is linear itself (that is to say, $(af_1 + bf_2)' = af_1' + bf_2'$ when a and b are constants), and the same holds for higher-order derivatives, as well as for partial derivatives.

6.2 Space discretization for the BS equation

The Black-Scholes equation for the unknown price $\pi(t,x)$ of a traded asset is (6.1) together with contract-specific boundary conditions. In the pricing of derivative contracts, typically the value of the contract at the time of expiration is given and the process of numerically solving the above equation can be viewed as a procedure

that works backward from the time of maturity to the initial time. The PDE (6.1) may be rewritten in the form

$$-\frac{\partial \pi}{\partial t} = \frac{\partial \pi}{\partial x}(\mu_X - \sigma_X \lambda) + \tfrac{1}{2} \operatorname{tr} \frac{\partial^2 \pi}{\partial x^2} \sigma_X \sigma_X^\top - r\pi \tag{6.9}$$

to bring out this point of view more clearly. The right hand side in (6.9) contains only derivatives with respect to the state variables, whereas the left hand side is simply a derivative with respect to time; the minus sign is in some sense appropriate since the equation is to be solved backward in time. The discretization process for the above equation can be split into two stages; first, discretization of the right hand side containing the differential operators in the direction of the variable x, followed by discretization in time.

We now first consider the discretization in space. For any given t, the pricing function $\pi(t, x)$ is a function of the state variables x; so it is a function on (a subset of) $\mathbb{R}^n$. For computational purposes, some finite representation of the function needs to be used. Suppose that a grid has been chosen in the state space and that the function $\pi(t, x)$ is represented by a vector $z(t)$ of values on grid points; so the vector $z(t)$ has length N where N is the number of points in the chosen grid. By making use of a finite-difference scheme, the right hand side of (6.9) can then be expressed as the result of an operation on $z(t)$, and since in standard finite-difference applications this operation is linear, it can in fact be represented by a matrix.

Let us consider this process in more detail for the case in which we have a one-dimensional state variable, and a one-dimensional Brownian motion. We focus first of all on the second derivative that appears in (6.9). This operator may be approximated by the symmetric formula (6.7), where it is natural to let the step Δx be equal to the distance between two adjacent grid points. The formula (6.7) can be applied everywhere at the grid except at the boundary points, where only one neighbor is available. How to deal with boundary points is a general problem in the finite-difference method, which gains in importance as the dimension of the state space increases. Usually some information about the nature of the solution has to be used to arrive at an appropriate treatment of boundary points in a specific application. In many financial applications with a one-dimensional state, the state variable represents the price of an underlying asset, and the problem is to price an option that has a piecewise linear payoff function. In such cases, a grid will be chosen in such a way that the boundary points are far away from the location of the kinks in the payoff function. The presence of the kinks will then only have a small influence on the value of the solution at the boundary points, and we may act (for the purposes of computing the solution at the boundary points) as if the kinks are not present and the payoff is linear in the underlying. Then the pricing function is also linear as a function of the underlying, and as a result the second derivative

$$A_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 & & & & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & & & & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}. \tag{6.10}$$

Display 6.1: Approximation of first-order differential operator.

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 1 & -2 & 1 & 0 & & & & 0 \\ 0 & 1 & -2 & 1 & 0 & & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & & 0 & 1 & -2 & 1 & 0 \\ 0 & & & & 0 & 1 & -2 & 1 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix}. \tag{6.11}$$

Display 6.2: Approximation of second-order differential operator for functions that are approximately linear at the boundaries. Replace first and last row by $[-1 \;\; 1 \;\; 0 \;\; \cdots]$ and $[\cdots \;\; 0 \;\; -1 \;\; 1]$ for functions that are approximately exponential at the boundaries.

is zero. Using this as an approximation, the matrix representation for the second derivative is $\frac{1}{(\Delta x)^2} A_2$ where A_2 is the matrix shown in Display 6.2. The modifications at the boundary may not be satisfactory for all applications; as an alternative one can try to determine approximate solutions at the boundary by a different method, and then use this information to correct the results of the calculations on the basis of finite-difference matrices.

To approximate the first derivative, we may use the symmetric difference formula (6.6) at all points except the boundary points; at the boundary points a one-sided formula of the type (6.5) may be used. This leads to a matrix representation for the first derivative of the form $\frac{1}{\Delta x} A_1$ where A_1 is given as in Display 6.1, where it is assumed that the grid points $x_1, x_2, \ldots, x_N$ are ordered such that $x_1 < x_2 < \cdots < x_N$.

For a complete representation of the right hand side of (6.9), we also need to take care of the multiplication operators appearing in this equation. We have assumed that the model we consider is driven by a single Brownian motion, so that not only μ_X but also λ and σ_X are scalar functions. We assume furthermore that these

functions depend only on x, not on t. Writing $\bar{\mu}(x) := \mu_X(x) - \sigma_X(x)\lambda(x)$ for brevity, the operator of multiplication by $\bar{\mu}(x)$ is represented by the diagonal matrix with entries $\bar{\mu}(x_1), \cdots, \bar{\mu}(x_N)$ on the diagonal. Call this matrix A_μ. The operation of multiplying by σ_X^2 is represented in the same way by a diagonal matrix A_σ. Finally, the complete operator on the right hand side of (6.9) is represented by the matrix

$$A := -rI + \frac{1}{\Delta x} A_\mu A_1 + \tfrac{1}{2} \frac{1}{(\Delta x)^2} A_\sigma A_2. \tag{6.12}$$

6.3 Preliminary transformation of variables

The above is a straightforward discretization applied to the Black-Scholes equation as given in (6.9). It may be advantageous, however, to apply some transformation of independent and/or independent variables before applying a discretization scheme. As an example, take the PDE associated to the standard Black-Scholes model:

$$\frac{\partial \pi}{\partial t} + rS \frac{\partial \pi}{\partial S} + \tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 \pi}{\partial S^2} - r\pi = 0. \tag{6.13}$$

Define a new function $\tilde{\pi}$ by $\tilde{\pi}(t, x) = \pi(t, \exp(x))$; then we have

$$\frac{\partial \tilde{\pi}}{\partial x}(t, x) = \exp(x) \frac{\partial \pi}{\partial S}(t, \exp(x))$$

and

$$\frac{\partial^2 \tilde{\pi}}{\partial x^2}(t, x) = \exp(x) \frac{\partial \pi}{\partial S}(t, \exp(x)) + \exp(2x) \frac{\partial^2 \pi}{\partial S^2}(t, \exp(x)).$$

Setting $S = \exp(x)$, i.e. $x = \log S$, one can write in shorthand notation

$$S \frac{\partial \pi}{\partial S} = \frac{\partial \tilde{\pi}}{\partial x}, \qquad S \frac{\partial \pi}{\partial S} + S^2 \frac{\partial^2 \pi}{\partial S^2} = \frac{\partial^2 \tilde{\pi}}{\partial x^2}.$$

The equation (6.13)) is then rewritten as follows:

$$\frac{\partial \tilde{\pi}}{\partial t} + (r - \tfrac{1}{2}\sigma^2) \frac{\partial \tilde{\pi}}{\partial x} + \tfrac{1}{2}\sigma^2 \frac{\partial^2 \tilde{\pi}}{\partial x^2} - r\tilde{\pi} = 0. \tag{6.14}$$

This new version does not have state-dependent coefficients, in contrast to the original equation (6.13). Note that a constant-distance grid in the variable y corresponds to a constant-ratio grid in the variable x; this fits in with the intuition that the relative sizes of values of the underlying are more important than the absolute sizes.

The transformation to logarithmic variables turns functions that are linear in the original variable x to functions that are exponential in the new variable y. This fact has some implications for the approximation of the second-order differential operator at the boundary points. While for a function that is approximately linear the second derivative is approximately 0, the second derivative of a function that is

approximately exponential is approximately equal to its first derivative. So, after logarithmic transformation, it makes sense to replace the zero first and last row in the matrix A_2 as given in Display 6.2 by the first and last row of the matrix A_1 in Display 6.1.

6.4 Time stepping

After discretization with respect to the state variables, the equation (6.9) is replaced by the linear ordinary differential equation

$$-\frac{dz}{dt}(t) = Az(t). \tag{6.15}$$

Here it is assumed that the coefficients appearing in (6.9), such as μ_X, σ_X and so on, do not depend on time directly, i.e. otherwise than through the state variables. If such dependence does appear, then the expression at the right hand side of (6.15) is replaced by $A(t)z(t)$ where $A(t)$ is a deterministic matrix function of time. This complicates the implementation of the numerical method to be discussed below only to a small extent, but for the analysis of the method it is easier to suppose (as it is the case in many applications) that we are in the time-homogeneous case in which the matrix A is constant.

For pricing purposes, the equation (6.15) is to be solved backwards in time; that is to say, $z(T)$ is given and we want to compute $z(0)$. We can change to the more standard setting of solving differential equations forward in time by introducing a new unknown function $v(t) := z(\tau(t))$ with $\tau(t) := T - t$. Applying the chain rule of deterministic calculus, we can write

$$\frac{dv}{dt}(t) = \frac{dz}{d\tau}(\tau(t))\,\frac{d\tau}{dt}(t) = -Az(\tau(t))\,(-1) = Az(\tau(t)) = Av(t).$$

Therefore the equation (6.15) becomes

$$\frac{dv}{dt}(t) = Av(t) \tag{6.16}$$

where now $v(0)$ is given and we want to compute $v(T)$. The solution can be written, in terms of the matrix exponential that was introduced in (2.76), as $v(T) = e^{AT}v(0)$. The size of the matrix A is given by the number of grid points in the state space, so this matrix could be quite large. Computation of the exponential of a large matrix is a challenge for numerical software. While many software packages have built-in routines for computing the matrix exponential, it is nevertheless worthwile to have a look at the time discretization methods that can be used for this purpose, in the particular case of space-discretized versions of equations of the form (6.1). One of the reasons is that these methods are also of use in applying the finite-difference method

to the computation of the values of American options, as discussed in Section 6.6 below.

So, let us choose a time step Δt. Since $v(0)$ is given, a natural approach is to compute $v(\Delta t)$ from $v(0)$, then $v(2\Delta t)$ from $v(\Delta t)$, and so on until we arrive at $v(T)$. In general, suppose that we have computed $v(t)$, and we want to compute $v(t+\Delta t)$. A straightforward application of the finite-difference formula (6.5) leads to the approximation

$$\frac{v(t+\Delta t)-v(t)}{\Delta t} = Av(t) \tag{6.17}$$

from which we get

$$v(t+\Delta t) = (I+\Delta t A)v(t). \tag{6.18}$$

However, the expression at the left hand side of (6.17) may just as well be viewed as an approximation of $(dv/dt)(t+\Delta t)$. The approximation

$$\frac{v(t+\Delta t)-v(t)}{\Delta t} = Av(t+\Delta t) \tag{6.19}$$

leads to the formula

$$v(t+\Delta t) = (I-\Delta t\, A)^{-1}v(t). \tag{6.20}$$

This is called an *implicit* scheme, since it requires an equation to be solved, namely $(I-\Delta t\, A)v(t+\Delta t) = v(t)$. Solving such an equation is a numerical problem of its own. On the positive side, the computation is facilitated by the fact that a good initial guess for the solution is available; indeed, the solution vector obtained in the preceding step can be used as such. The method (6.18), which does not require solving a linear equation, is called an *explicit* scheme. More generally, we can use an approximation of the form

$$\frac{v(t+\Delta t)-v(t)}{\Delta t} = \theta Av(t+\Delta t) + (1-\theta)Av(t) \tag{6.21}$$

where θ is a number between 0 and 1. For $\theta = 0$, the explicit scheme is recovered, whereas taking $\theta = 1$ produces the implicit scheme. The general scheme obtained from (6.21) is

$$v(t+\Delta t) = (I-\theta\Delta t\, A)^{-1}(I+(1-\theta)\Delta t\, A)v(t). \tag{6.22}$$

To assess the error that is incurred here, the expression above may be compared to the exact solution, which is given by

$$v(t+\Delta t) = e^{\Delta t A}v(t). \tag{6.23}$$

The matrix exponential may be expanded in terms of powers of Δt:

$$e^{\Delta t A} = I + A\Delta t + \tfrac{1}{2}A^2(\Delta t)^2 + \tfrac{1}{6}A^3(\Delta t)^3 + \cdots . \tag{6.24}$$

This may be compared to the power series expansion of the expression in (6.22):

$$\begin{aligned}(I - \theta\Delta t A)^{-1}(I + (1-\theta)\Delta t A) = \\ = (I + \theta A\Delta t + \theta^2 A^2(\Delta t)^2 + \cdots)(I + (1-\theta)A\Delta t) = \\ = I + A\Delta t + \theta A^2(\Delta t)^2 + \theta^2 A^3(\Delta t)^3 + \cdots .\end{aligned} \tag{6.25}$$

It is seen that the error is of order $(\Delta t)^2$ for all θ, except when $\theta = 0.5$; in the latter case the error is of order $(\Delta t)^3$. The method that is obtained from (6.21) by taking $\theta = 0.5$ is called the *Crank-Nicolson scheme*.[2] All schemes of the form (6.21) with $\theta \neq 0$ require the solution of a matrix equation and so in this sense all these schemes are implicit; therefore the method (6.19) is also sometimes referred to as the "fully implicit" scheme.

6.5 Stability analysis

The numerical procedure that was developed above eventually comes down to a recursion of the form $v(t + \Delta t) = Mv(t)$ where M is a certain matrix. At each step of the recursion there is an approximation error. Of course the scheme has been constructed in such a way that this error is small; however we need to analyze what happens to the error as the recursion proceeds. Denote the "exact" solution at time $k\Delta t$ by v_k, and let $\tilde{v}_k$ denote the actually computed solution. We then have (ideally)

$$v_{k+1} = Mv_k \tag{6.26}$$

while on the other hand

$$\tilde{v}_{k+1} = M\tilde{v}_k + \varepsilon_k \tag{6.27}$$

where ε_k is the error incurred at step k. The sequence of errors $e_k := \tilde{v}_k - v_k$ then satisfies

$$e_{k+1} = Me_k + \varepsilon_k. \tag{6.28}$$

Assuming that $\tilde{v}_0 = v_0$ so that $e_0 = 0$, we have $e_1 = \varepsilon_1$, $e_2 = M\varepsilon_1 + \varepsilon_2$, and in general

$$e_k = M^{k-1}\varepsilon_1 + M^{k-2}\varepsilon_2 + \cdots + M\varepsilon_{k-1} + \varepsilon_k. \tag{6.29}$$

[2] John Crank (1916–2006), British mathematician. Phyllis Nicolson (1917–1968), British mathematician. The Crank-Nicolson scheme was developed during the Second World War and was published in 1947.

It is seen that there is a multiplicative effect on errors. In case M has eigenvalues that have absolute value larger than 1, the error after a number of iterations may be much larger than the error in each individual step. Therefore it is important to consider the eigenvalues of the recursion matrix.

When considering various time stepping schemes, we always relate a certain matrix A representing a continuous-time evolution to a matrix M that is used in a discrete-time recursion. The matrix M can be described as a function of the matrix A; in fact, in all cases, it is a function of the form $p(A)q(A)^{-1}$ where $p(x)$ and $q(x)$ are polynomials. Generally speaking, given a polynomial $p(x) = p_k x^k + p_{k-1} + \cdots p_1 x + p_0$, the matrix $p(A)$ is defined, for any square matrix A, as $p(A) = p_k A^k + p_{k-1} + \cdots + p_1 A + p_0 I$. Some useful properties are the following:

(i) if p and q are both polynomials, then the matrices $p(A)$ and $q(A)$ commute, i.e. $p(A)q(A) = q(A)p(A)$;

(ii) (as a consequence of the above) if p and q are polynomials and $q(A)$ is invertible, then $p(A)q(A)^{-1} = q(A)^{-1}p(A)$;

(iii) if λ is an eigenvalue of A, then $p(\lambda)$ is an eigenvalue of $p(A)$, and conversely, if μ is an eigenvalue of $p(A)$, then there is an eigenvalue λ of A such that $p(\lambda) = \mu$;

(iv) if λ is an eigenvalue of A, then $p(\lambda)/q(\lambda)$ is an eigenvalue of $p(A)q(A)^{-1}$, and conversely, if μ is an eigenvalue of $p(A)q(A)^{-1}$, then there is an eigenvalue λ of A such that $p(\lambda)/q(\lambda) = \mu$.

The statement in (i) is proved simply by expanding both $p(A)q(A)$ and $q(A)p(A)$ in terms of powers of A, and noting that these expansions are the same. To prove the first fact stated in (iii), let x be an eigenvector of A associated to the eigenvalue λ, and note that $p(A)x = p_k A^k x + \cdots + p_0 x = p_k \lambda^k x + \cdots p_0 x = p(\lambda)x$ so that x is also an eigenvector of $p(A)$, with eigenvalue $p(\lambda)$. The converse statement follows from this in case the matrix A has a complete set of eigenvectors, but holds as well in the general case.[3]

From the theory of differential equations, it needs to be remembered that the solutions of a vector differential equation of the form

$$\frac{dx}{dt}(t) = Ax(t), \qquad x(0) = x_0$$

(i.e. a system of linear differential equations) converge to 0 as $t \to \infty$ if and only if all of the eigenvalues of the matrix A are in the open left half of the complex plane

[3]For instance, one may use the argument, although it may be considered inelegant, that every square matrix can be approximated arbitrarily closely by a matrix that has a complete set of eigenvectors, and combine this with the fact that the eigenvalues of a matrix are continuous functions of its entries.

(i.e. their real parts are negative). It is then said that the matrix A is "stable". For discrete-time recursions, the analogous condition is that all eigenvalues of the recursion matrix should have absolute value less than 1, i.e. they should be inside the unit circle in the complex plane. In the application discussed here, the matrix A is derived as an approximation on a finite grid of the differential operator that appears on the right hand side of (6.9)), or a log-transformed version of it. As long as the interest rate r is nonnegative, this operator should not generate any exponential growth.[4] Consequently, we can focus on the question when a recursion matrix of the form $M = p(A)q(A)^{-1}$ is stable, given that the matrix A is stable. According to the facts from linear algebra discussed above, this happens exactly when the real part of $p(\lambda)/q(\lambda)$ is negative whenever the absolute vale of λ is less than 1. More precise conditions can be given if more is known about the eigenvalues of A.

Consider in particular the explicit scheme (6.18), which constructs the recursion matrix M from the continuous-time matrix A by the formula $M = I + \Delta t A$. The range of points in the complex plane that are taken into the unit circle by the mapping $\lambda \mapsto 1 + \lambda\Delta t$ consists of the interior of a circle in the complex plane with radius $-1/\Delta t$ and radius $1/\Delta t$. In particular, eigenvalues that lie on the real axis to the left of $-2/\Delta t$ will be mapped *outside* the unit circle. So if we apply the explicit scheme to the equation (6.16), it may happen that the discrete-time approximation is unstable even when the original equation is stable.

It depends on the choice of the time step Δt and on the eigenvalues of the matrix A whether or not a numerical stability problem as described above may arise. The eigenvalues of A themselves depend on the discretization step that has been applied in the state space. To get some idea of the conditions under which a loss of stability may arise, let us analyze a *model problem* (that is, a problem that is simple enough to be completely analyzed and that can be taken as more or less representative of a large model class). Suppose that we have a one-dimensional state space which has been discretized using a fixed step Δx after a logarithmic transformation, and that, as above, a symmetric difference has been used for the second derivative. For small Δx, it is seen from formula (6.12) that the second derivative is the dominant term in A. Assume that volatility (after the log transformation) is constant, as it is the case in the standard Black-Scholes model. For convenience, we also replace the matrix A_2 (associated to the second derivative) by the symmetric matrix D shown in Display 6.3 which differs from A_2 only in the first and last row. It can be shown (see Exc. 2) that, independent of the number of grid points that is chosen, the eigenvalues of D always lie between 0 and -4; and there are eigenvalues close

[4] When r is negative, then for instance the time-0 price of a constant payoff at time T is exponentially increasing with T. To treat such cases numerically, it is advisable to apply first a time-dependent transformation of variables that removes this behavior.

$$D = \begin{bmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{bmatrix} \tag{6.30}$$

Display 6.3: Matrix for model problem.

to 0 and close to -4. Given Δx, one should therefore take Δt such that

$$-2\,\frac{1}{\Delta t} \leq -4\,\frac{\frac{1}{2}\sigma^2}{(\Delta x)^2}$$

or in other words

$$\sigma\sqrt{\Delta t} \leq \Delta x. \tag{6.31}$$

The quantity that appears at the left hand side can be interpreted as a measure of the "typical change of the state variable" corresponding to the time interval Δt.

The constraint (6.31) may be unpleasant. If for instance the step in the space direction is halved in order to improve the accuracy of the approximation in the x-direction, then the number of time steps must be quadrupled to keep the ratio $\sigma\sqrt{\Delta t}/\Delta x$ the same. This by itself causes the computation time to be multiplied by four, in addition to the effects of working with a larger matrix A. The presence of the stability condition (6.31), which appears in similar form also for higher-dimensional state spaces, is therefore a disadvantage of the explicit method.

The analysis may be carried out similarly for the more general method (6.21). According to the properties of eigenvalues of matrices of the form $q(A)^{-1}p(A)$ as discussed above, the eigenvalues of the matrix $(I-\theta\Delta t\,A)^{-1}(I+(1-\theta)\Delta t\,A)$ which appears in (6.21) are obtained from the eigenvalues of A by applying the transformation

$$\lambda \mapsto \frac{1+(1-\theta)\Delta t\,\lambda}{1-\theta\Delta t\,\lambda}. \tag{6.32}$$

In the model problem analyzed above, the eigenvalues of the matrix A range from a little bit less than 0 to approximately $-2\sigma^2/\Delta x^2$. This leads to the stability condition

$$\left|\frac{1-2(1-\theta)\kappa}{1+2\theta\kappa}\right| < 1 \qquad \text{with } \kappa := \frac{\sigma^2\Delta t}{\Delta x^2}\,. \tag{6.33}$$

The condition above can be rewritten as

$$(1-2\theta)\kappa < 1. \tag{6.34}$$

For $\theta = 0$ (the explicit method), we arrive at the condition $\kappa < 1$ that was already found above. The constraint on κ becomes less strict for larger values of θ, and no constraint on κ remains if $\theta \geq \frac{1}{2}$. So in the Crank-Nicolson scheme, and also in the fully implicit scheme, there is no condition on the time step that is needed to guarantee stability; these schemes are therefore called *unconditionally stable*. For values $\theta \in \left[\frac{1}{2}, 1\right]$, the mapping (6.32) takes the open left half of the complex plane (i.e. the set of complex numbers with negative real part) to the interior of the unit disc in the complex plane. Therefore, if the matrix A is stable in the continuous-time sense (all eigenvalues in the open left half of the complex plane), then the associated matrix $(I - \theta\Delta t\, A)^{-1}(I + (1-\theta)\Delta t\, A)$ is stable in the discrete-time sense (all eigenvalues inside the unit disc).

6.6 American options

The simplest computational approach to the pricing of American options is to replace them by a *Bermudan approximation*. After discretization in space and in time, the option value at each point in the time grid is determined as the maximum of the value of immediate exercise and the continuation value. The latter value is determined as the value of the European option whose time of maturity is the point in the time grid that corresponds to the foregoing computational step, and whose payoff is the approximation to the American option value that has been computed for that point. In this approach, the American option is treated as a Bermudan option, which can only be exercised at a limited number of points in time instead of at any point in time as in the American case. This should reduce the value of the option somewhat, but the error that is induced in this way tends to zero as the time step becomes smaller and smaller.

To illustrate the Bermudan approximation method, let the value of immediate exercise at time t be given by the vector $g(t)$, where the entries of this vector correspond to the points of a grid in the state space. The recursion (6.18) for European options is now replaced by

$$v(t + \Delta t) = \max\big(g(t + \Delta t),\, (I + \Delta t\, A)v(t)\big) \tag{6.35}$$

where the max refers to the *componentwise* maximum. Instead of the explicit form, one might for instance also use the implicit or Crank-Nicolson form.

To get a wider array of choices in computation, start from the basic pricing equation (3.68) for American options, which takes the place of the equation (6.9) that applies for European options. We can write (3.68) in a form akin to (6.9), but instead of the single equation (6.9) we get that, for every t and x, at least one of

the following sets of equalities and inequalities should be satisfied:

$$-\frac{\partial \pi}{\partial t} \geq \frac{\partial \pi}{\partial x}(\mu_X - \sigma_X \lambda) + \tfrac{1}{2} \operatorname{tr} \frac{\partial^2 \pi}{\partial x^2} \sigma_X \sigma_X^\top - r\pi \quad \text{and} \quad \pi = F \tag{6.36a}$$

$$-\frac{\partial \pi}{\partial t} = \frac{\partial \pi}{\partial x}(\mu_X - \sigma_X \lambda) + \tfrac{1}{2} \operatorname{tr} \frac{\partial^2 \pi}{\partial x^2} \sigma_X \sigma_X^\top - r\pi \quad \text{and} \quad \pi \geq F. \tag{6.36b}$$

As before, we may first do a transformation of variables, then discretize in space and reverse the parametrization of time. The conditions (6.36) are then replaced by the requirement that, for every t and i, at least one of the following should be satisfied (compare (6.16)):

$$\frac{dv_i}{dt}(t) \geq (Av(t))_i \quad \text{and} \quad v_i(t) = g_i(t) \tag{6.37a}$$

$$\frac{dv_i}{dt}(t) = (Av(t))_i \quad \text{and} \quad v_i(t) \geq g_i(t) \tag{6.37b}$$

where $g(t)$ is a discretized version of the payoff function $F(t, \cdot)$, suitably transformed in case a transformation of variables has been done. The size of the matrix A and the length of the vector $g(t)$ are equal to the number of grid points used. The equations (6.37) can be written more concisely in vector notation, as follows:

$$0 \leq \frac{dv}{dt}(t) - Av(t) \perp v(t) - g(t) \geq 0 \tag{6.38}$$

with componentwise inequalities; the zero symbols on the left hand side and the right hand side represent n-vectors. This is indeed the same as the conditions stated in (6.37), because the condition $x \perp y$ is satisfied for nonnegative vectors x and y in $\mathbb{R}^n$ if and only if for each $i = 1, \ldots, n$, either $x_i = 0$ or $y_i = 0$ or both. An equivalent reformulation is

$$\min\left(\frac{dv}{dt}(t) - Av(t), v(t) - g(t)\right) = 0 \tag{6.39}$$

where the componentwise minimum is used and again the symbol 0 on the right hand side is actually a vector of zeros. One may introduce so-called "slack variables" by defining

$$z(t) = \frac{dv}{dt}(t) - Av(t), \qquad w(t) = v(t) - g(t) \tag{6.40}$$

which leads to a representation of (6.37) in the form

$$\frac{dv}{dt}(t) = Av(t) + z(t) \tag{6.41a}$$

$$w(t) = v(t) - g(t) \tag{6.41b}$$

$$0 \leq w(t) \perp z(t) \geq 0. \tag{6.41c}$$

The last line may also be written as $\min(w(t), z(t)) = 0$. The conditions that appear in (6.41c) are called *complementarity conditions* since they express that the sets of indices i at which $w_i = 0$ and of indices i at which $z_i = 0$ are complements, at least in the typical situation in which there are no indices i for which one has both $w_i = 0$ and $z_i = 0$. Indices with $w_i = 0$ correspond to grid points in the exercise region, whereas indices such that $z_i = 0$ refer to grid points in the continuation region. Complementarity conditions form a typical structure that occurs in optimization problems subject to constraints (slack variables vs. Lagrange[5] multipliers) as well as in many situations in physics and engineering that involve unilateral constraints.

Starting from (6.39) or one of the equivalent formulations, time-stepping schemes can be constructed by replacing the continuous-time expressions by expressions in terms of variables $v(t)$ and $v(t + \Delta t)$, which can be solved for $v(t + \Delta t)$ when $v(t)$ is given. To take a time step, it is natural to assume that the complementarity conditions are already satisfied at time t, and then to derive equations for $v(t + \Delta t)$ from the requirement that these conditions should also hold at time $t + \Delta t$. A scheme that is similar to the explicit scheme for European options is the following:

$$\min\left(\frac{v(t + \Delta t) - v(t)}{\Delta t} - Av(t),\, v(t + \Delta t) - g(t + \Delta t)\right) = 0.$$

This condition is equivalent to[6]

$$\min\big(v(t + \Delta t) - (I + \Delta tA)v(t),\, v(t + \Delta t) - g(t + \Delta t)\big) = 0.$$

The unknown $v(t + \Delta t)$ can be taken out:

$$v(t + \Delta t) + \min\big(-(I + \Delta tA)v(t),\, -g(t + \Delta t)\big) = 0.$$

In other words,

$$v(t + \Delta t) = \max\big((I + \Delta tA)v(t),\, g(t + \Delta t)\big). \tag{6.42}$$

This scheme can be interpreted as resulting from an approximation of the American option as a Bermudan option which can be exercised as the sample times $0, \Delta t, 2\Delta t, \ldots$, with $(I + \Delta tA)v(t)$ as an approximation of the continuation value at time point $t + \Delta t$ (compare (6.18)),[7] while $g(t + \Delta t)$ represents the value of immediate exercise. Although the error propagation in the American case is more

[5] Joseph-Louis Lagrange (1736–1813), Italian/French mathematician and astronomer.

[6] Note that the condition $\min(x, y) = 0$, for $x, y \in \mathbb{R}^n$, is equivalent to $\min(c_1 x, c_2 y) = 0$ for any $c_1, c_2 > 0$.

[7] Recall that the time parameter used in the computation is in reverse direction, so that time point $t + \Delta t$ is $T - t - \Delta t$ in actual time.

complicated than what is shown by (6.29), it is still advisable to keep the eigenvalues of the recursion matrix, which is equal to $I+\Delta t A$ in this case, within the unit circle. This means that the same restrictions apply as in the case of the explicit method for European options.

A more general scheme can be obtained by using an expression for the discretized time derivative as in (6.21). One can write the set of conditions to be satisfied by $v(t+\Delta t)$ as follows:

$$\min\big(\eta\big((I-\theta\Delta t A)v(t+\Delta t)-v_0\big),\, v(t+\Delta t)-g(t+\Delta t)\big)=0 \tag{6.43}$$

where $v_0 := (I+(1-\theta)\Delta t A)v(t)$ is used as a shorthand, and where a positive parameter η is introduced to create an additional degree of freedom in the numerical procedure.[8] The condition (6.43) is equivalent to

$$v(t+\Delta t)=\max\big((I-\eta(I-\theta\Delta t A))v(t+\Delta t)+\eta v_0,\, g(t+\Delta t)\big). \tag{6.44}$$

In contrast to the case of the explicit method, it is not possible to write down the solution immediately. The equation (6.44) is of the form of a fixed-point equation, and a natural way to solve it would be iteration: start with an initial guess for $v(t+\Delta t)$, insert that into the right hand side of (6.44) to obtain a new and hopefully improved guess, insert the new guess into the right hand side of (6.44) to obtain a second new guess, and so on. Due to the presence of the max operator, this is a nonlinear iteration and its convergence behavior is not easily analyzed. It is advisable though to make sure that the eigenvalues of the matrix $I-\eta(I-\theta\Delta t A)$ are inside the unit circle. If we assume as before that the eigenvalues of the matrix A range from 0 to -2κ, this means that both $1-\eta$ and $1-\eta(1+2\theta\kappa)$ should be less than 1 in absolute value. The stability criterion (6.34) is to be satisfied as well; note that this criterion can also be written as $1+2\theta\kappa>\kappa$. It follows that η should be less than $2/\kappa$; in particular, no single value of η is good enough to accommodate all values that κ might take.

An alternative route is to consider not $v(t+\Delta t)$ directly as the unknown, but rather $(I-\theta\Delta t A)v(t+\Delta t)$. Instead of (6.43), one then obtains the condition

$$\min\big(y-v_0,\, \eta\big((I-\theta\Delta t A)^{-1}y-g(t+\Delta t)\big)\big)=0$$

where again a parameter $\eta>0$ is inserted, although on the other side since that is more convenient in this case. The condition above can be rewritten as

$$y=\max\big(v_0,(I-\eta(I-\theta\Delta t A)^{-1})y+\eta g(t+\Delta t)\big).$$

[8] For even more degrees of freedom, one can note that, for any diagonal matrix H with positive elements on the diagonal, $\min(Hx,y)=0$ is also equivalent to $\min(x,y)=0$.

The conditions associated with the eigenvalues of the matrix that appears in the iteration are now $|1-\eta| < 1$ and $|1-\eta/(1+2\theta\kappa)| < 1$. These conditions are satisfied when $\eta < 2$, irrespective of the value of κ.

The procedures that are suggested here to find numerical solutions of (6.41) by time stepping are by no means the only ones possible. In the mathematical programming literature, extensive studies have been made of problems of the form

$$0 \leq w \perp q + Mw \geq 0 \tag{6.45}$$

where $w \in \mathbb{R}^n$ is the unknown, $q \in \mathbb{R}^n$ is a given vector, and M is a given $n \times n$ matrix. A problem of this type is called a "Linear Complementarity Problem" (LCP). It is known that the problem (6.45) is uniquely solvable for each given q if and only if all principal minors[9] of M are positive. Many algorithms have been devised to solve LCPs; the literature is somewhat analogous to the literature on linear programming and contains pivoting methods as well as iterative methods. Pivoting methods aim at finding the complementary index sets $\{i \,|\, w_i > 0\}$ and $\{i \,|\, w_i = 0\}$. In applications to pricing of American options, these index sets correspond to the exercise region and the continuation region. In specific cases, it may be possible to parametrize these regions in a simple way; for instance, in the case of a put option in the Black-Scholes model, it is clear that one only needs to find the separation point between the exercise region which contains low values of the underlying and the continuation region, which contains high values. Using this information, one might construct a very efficient pivoting algorithm.

A simple approximation scheme is shown in Code Example 6.1. The scheme uses a wide grid in the direction of the current price of the underlying, and a large number of time steps. Recursion is carried out with no transformation of variables; the Bermudan approximation is used, with the continuation value being computed by the Crank-Nicolson method. The result of the calculation is shown in Fig. 6.1.

6.7 Markov chains and tree methods

The finite-difference method applies discretization to a partial differential equation that describes the exact solution to an option pricing problem. The discretization is necessary in cases in which no analytical solution can be found. But, given that approximation is necessary anyway, one might argue that discretization could be applied already at an earlier stage. If the process of the underlying asset S_T is replaced by a process that takes place in discrete time steps and in a discrete

[9] A *principal submatrix* of a given $n \times n$ matrix M is a submatrix of the form $(m_{ij})_{i\in\alpha, j\in\alpha}$ where α is a nonempty subset of the index set $\{1, \ldots, n\}$. The *principal minors* of M are the determinants of the principal submatrices of M.

```
r = 0.03; sigma = 0.2; K = 100; T = 1;
Smin = 50; Smax = 250; Nx = 100; Nt = 100;
dx = (Smax-Smin)/Nx; dt = T/Nt;
S = (Smin:dx:Smax)'; N = length(S);
D1  =   (1/dx)*(diag([-0.5*ones(1,N-2)  -1],-1) + ...
                diag([-1  zeros(1,N-2)   1], 0) + ...
                diag([ 1   0.5*ones(1,N-2)], 1));
D2  = (1/dx^2)*(diag([ones(1,N-2)      0],-1)   + ...
                diag([0 -2*ones(1,N-2) 0], 0)   + ...
                diag([0      ones(1,N-2)], 1))  ;
mS  = diag(S);     % multiplication by S
mS2 = diag(S.^2);  % multiplication by S^2
A   = r*mS*D1 + 0.5*sigma^2*mS2*D2 - r*eye(N); % assemble matrix A
% start recursion
V = max(K-S,0); % initial condition
for i = 1:Nt
  V = (eye(N)-0.5*dt*A)\((eye(N)+0.5*dt*A)*V); % CN method
  V = max(V,K-S); % max of continuation value and exercise value
end
% end recursion
plot(S,[V max(K-S,0)]); axis([70 150 0 30]) % exc value for comparison
xlabel('value of underlying at t = 0'); ylabel('option value')
```

Code Example 6.1: Approximation scheme for the calculation of the value of an American put option in the BS model, as a function of the underlying.

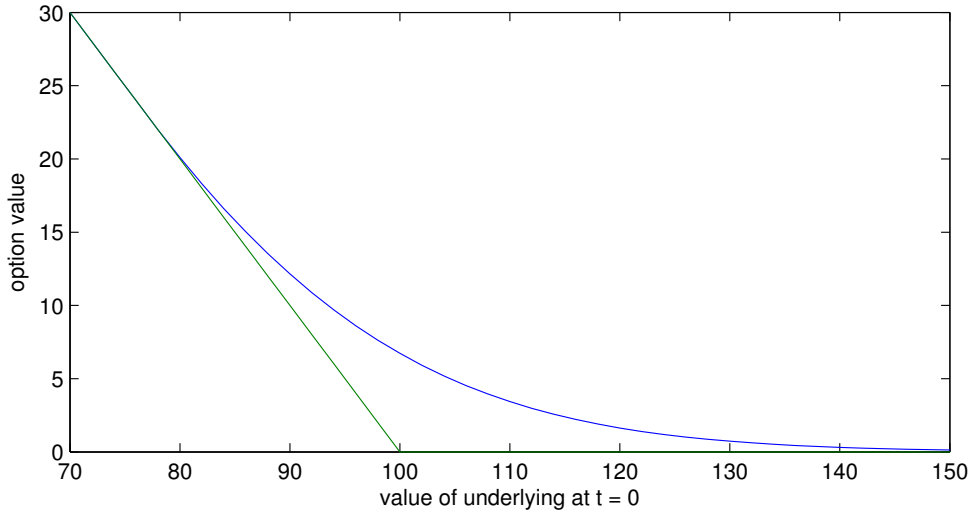

Figure 6.1: Value of an American put option as a function of the value of the underlying at time 0, as computed by Code Example 6.1.

state space (i.e., a finite set) then the problem of computing a quantity of the form $E[F(S_T)]$ can still be stated; moreover, if the number of time steps and/or states is not excessively large, it may be possible to solve this problem exactly. We then have an exact solution of an approximate problem, which may serve just as well as an approximate solution of the exact problem. The underlying philosophy is that, whenever one constructs an approximation scheme, it is helpful if the approximating scheme is not just a numerical construct but has an interpretation by itself. The availability of such an interpretation makes it easier to understand what is happening in the scheme, and to ensure that desired properties will indeed hold. As will be seen below, computational schemes that are obtained from constructing an exact solution to an approximate problem can sometimes be the same as schemes that are designed as approximate solutions to an exact problem. In such cases, the alternative view can be useful to grasp the meaning of approximations that have been made.

6.7.1 Random walks and Markov chains

A simple example of approximation of a continuous-time continuous-state[10] stochastic process by a discrete-time discrete-state process is the approximation of Brownian motion by a random walk. Let a process X_t be defined by

$$X_t = x_0 + \sigma W_t \tag{6.46}$$

where W_t is a Brownian motion. Let Δt be a time step. For $k = 1, 2, \ldots$, we have

$$X_{k\Delta t} = x_0 + \sigma \sum_{i=1}^{k} (W_{i\Delta t} - W_{(i-1)\Delta t}) = x_0 + \sum_{i=0}^{k} Z_i \tag{6.47}$$

where the random variables $Z_i := \sigma(W_{i\Delta t} - W_{(i-1)\Delta t})$ form an i.i.d. sequence of normally distributed variables with expectation 0 and variance $\sigma^2 \Delta t$. The values of the process X_t at times $k\Delta t$ therefore can be described as a cumulative sum of independent random variables. Suppose now that the i.i.d. random variables Z_i are replaced by i.i.d. variables $\hat{Z}_i$ that can only take two values, namely Δx and $-\Delta x$, with equal probabilities. The expectation of $\hat{Z}_i$ is 0 and the variance is Δx^2, which is equal to the variance of the normal increments in (6.47) if we take

$$\Delta x = \sigma \sqrt{\Delta t}. \tag{6.48}$$

[10]The term "state" is used here in the general sense of stochastic processes, as a possible value that the process may take, rather than in the specific sense of the state space model (3.1).

The central limit theorem implies that, for large N, the distribution of $\sum_{i=1}^{N} \hat{Z}_i$ approximates the distribution of $\sum_{i=1}^{N} Z_i$. The discrete-time process given by

$$\hat{X}_k = x_0 + \sum_{i=1}^{k} \hat{Z}_i \qquad (k = 0, 1, 2, \dots). \tag{6.49}$$

is called a *random walk*. Due to the fact that the increments in (6.49) are independent, all statistical properties of the process $\hat{X}_k$ after a given index k_0 depend on the history of the process up to k_0 only through $\hat{X}_{k_0}$; what happened before k_0 doesn't matter. This is the "Markovian property" that was already discussed in Section 3.1.1 for the case of state processes described by SDEs: the current value of the state contains all information from the past that is relevant to the future. Another property of the random walk is that, unlike the process X_t which can take all real values, the values it can take are confined to the countably infinite set $\{x_0 + i\Delta x \mid i \in \mathbb{Z}\}$. The Markovian property and the restriction of values that the process may take to a finite or countably infinite set are the defining characteristics of what are called *Markov chains*. The set of values that can be taken is called the *state space* of the chain.

Due to the Markov property, the statistical properties of a Markov chain are determined fully by only specifying the *transition probabilities*, which are the probabilities that the process will take the value j at step $k+1$, given that it is in state i at time k. Usually, the assumption is made that these probabilities depend only on i and j, and not on the time index k. Markov chains that satisfy this property are said to be *homogeneous*. The random walk can be described in terms of a homogeneous Markov chain by defining the state space to consist of the set $\{x_0 + i\Delta x \mid i \in \mathbb{Z}\}$, with transition probabilities defined by

$$p_{ij} = \begin{cases} \frac{1}{2} & \text{if } j = i+1 \text{ or } j = i-1 \\ 0 & \text{else.} \end{cases} \tag{6.50}$$

Specifically, the random walk is the process that is obtained from this chain by imposing the initial condition $\hat{X}_0 = x_0$.

For the purpose of numerical calculation, we like to work with Markov chains that have a finite state space. The approximation of the process X_t by means of the random walk $\hat{X}_k$ as described above would require an infinite state space if considered on intervals of arbitrary length, but fortunately most problems that are encountered in contingent claims analysis are equipped with a finite end date. In particular, in a European option pricing problem one is asked to compute a quantity of the form $E[F(X_T)]$ where F is a given function and T is a fixed time. In such a situation, it is enough to have an approximation of the process X_t on the interval $[0, T]$. The time step would then be taken equal to $\Delta t = T/N$ where N is a number

that is sufficiently large to ensure that the approximation of X_T by $\hat{X}_N$ is good enough for the desired accuracy. The state space can be restricted to the finite set $\{x_0 - N\Delta x, \ldots, x_0 - \Delta x, x_0, x_0 + \Delta x, \ldots, x_0 + N\Delta x\}$. The transition probabilities for the Markov chain to be defined on this set can be defined as in (6.50), except at the boundaries $i = -N$ and $i = N$. The corresponding states can only be reached at time T so it could be said that it is irrelevant how the transition probabilities are defined here, but for completeness one might for instance specify that the boundary states are absorbing, i.e., $p_{ii} = 1$ for $i = -N$ and $i = N$.

To obtain an approximate value for a quantity $E[f(X_T)]$ from the random walk approximation, it is natural to replace X_T by $\hat{X}_N$ and to define

$$\pi_k^i = E[F(\hat{X}_N) \,|\, \hat{X}_k = x_0 + i\Delta x] \qquad (i = -k, \ldots, k).$$

From the tower law of conditional expectations, it follows that $\pi_0^0 = E[F(\hat{X}_N)]$ can be computed recursively as follows:

$$\pi_N^i = F(x_0 + i\Delta x) \qquad (i = -N, \ldots, N) \tag{6.51a}$$

$$\pi_k^i = E[\pi_{k+1}^j \,|\, \hat{X}_k = x_0 + i\Delta x] = \tfrac{1}{2}\pi_{k+1}^{i+1} + \tfrac{1}{2}\pi_{k+1}^{i-1} \qquad (k = N-1, \ldots, 0;\ i = -k, \ldots, k). \tag{6.51b}$$

This computational scheme can be compared to the one that would be obtained from the PDE associated to the continuous-time continuous-state process X_t. Define

$$\pi(t, x) = E[F(X_T) \,|\, X_t = x].$$

Assuming that this function is sufficiently smooth, one can apply the Itô formula:

$$d\pi(t, X_t) = \frac{\partial \pi}{\partial t}(t, X_t)\, dt + \frac{\partial \pi}{\partial x}(t, X_t)\, dW_t + \tfrac{1}{2}\,\sigma^2 \frac{\partial^2 \pi}{\partial x^2}(t, X_t)\, dt.$$

From this it follows that

$$\begin{aligned} \pi(t, x) &= E[\pi(t + \Delta t, X_{t+\Delta t}) \,|\, X_t = x] \\ &= \pi(t, x) + \Big(\frac{\partial \pi}{\partial t}(t, X_t) + \tfrac{1}{2}\,\sigma^2 \frac{\partial^2 \pi}{\partial x^2}(t, X_t)\Big)\, dt + o(\Delta t) \end{aligned}$$

where $o(\Delta t)$ denotes a term that is small relative to Δt.[11] Subtracting $\pi(t, x)$ from both sides, dividing by Δt, and taking the limit as Δt tends to 0, one finds that

[11] The notation $f(x) = o(g(x))$ $(x \to a)$ in general means that $\lim_{x \to a} f(x)/g(x) = 0$. In particular, a term $o(x^k)$ $(x \downarrow 0)$ represents a quantity that approaches 0 more rapidly than x^k, as x tends to 0.

$\pi(t,x)$ satisfies the partial differential equation

$$-\frac{d\pi}{dt}(t,x) = \tfrac{1}{2}\,\sigma^2\,\frac{\partial^2\pi}{\partial x^2}(t,x) \tag{6.52}$$

which is known as the Kolmogorov backward equation in the theory of stochastic processes.[12] The standard explicit approximation scheme with time step Δt and space step Δx, based on the formula (6.7) for the second derivative with respect to x, leads to the computational scheme

$$\hat\pi_k^i = \hat\pi_{k+1}^i + \tfrac{1}{2}\sigma^2\,\frac{\hat\pi_{k+1}^{i+1} - 2\hat\pi_{k+1}^i + \hat\pi_{k+1}^{i-1}}{\Delta x^2}\,\Delta t$$

where $\hat\pi_k^i$ is an approximation to $\pi(k\Delta t, x_0 + i\Delta x)$. If one takes $\Delta x = \sigma\sqrt{\Delta t}$ as in (6.48), the scheme simplifies to

$$\hat\pi_k^i = \tfrac{1}{2}\hat\pi_{k+1}^{i+1} + \tfrac{1}{2}\hat\pi_{k+1}^{i-1} \tag{6.53}$$

which is the same as (6.51b). The computational scheme that is derived from the random walk approximation can therefore be viewed as a special case of the explicit approximation scheme for the PDE (6.52), with the central difference formula applied to approximate the second-order derivative in the space direction.

The computational process expressed by (6.51b) or (6.53) may be visualized as follows. Imagine the states of the Markov chain as bins that contain marbles. The bins are numbered from $-N$ to N. The process is initialized by placing $F(x_0+i\Delta x)$ marbles in bin i.[13] At each step, one half of the marbles in each bin are moved to the bin to the right, while the other half are moved to the bin to the left. The number of marbles that end up at bin 0 after N steps is the sought after quantity π_0^0. While carrying out this process, one does not need to bother about the marbles that are moved to bins from which bin 0 can no longer be reached in the remaining stages. Therefore, while the process starts with bins $-N$ to N, after k steps one only needs to be concerned with bins $-N+k$ to $N-k$. The process that is reduced in this way takes place on a triangle-shaped domain, which motivates the term "tree method". Another reduction of the effort in moving marbles can be achieved if it is noted that, after each step, all of the marbles that were in even-numbered bins before are now in odd-numbered bins, and vice versa. If the number N of steps is even, this means that the final outcome π_0^0 of the process is entirely determined by the numbers of marbles that were in even-numbered bins at the start of the process, while the result is determined entirely by the marbles initially in the odd-numbered

[12] Andrey Nikolaevich Kolmogorov (1903–1987), Russian mathematician.

[13] It is assumed here that the marbles can be cut, to accommodate non-integer quantities; if the payoff function F can take negative values, then also antimarbles need to be brought into play.

bins in case N is odd. Supposing for instance that one takes N to be even, the odd-numbered bins can be dispensed with by doubling the step; the computational scheme becomes

$$\pi_k^i = \tfrac{1}{4}\pi_{k+2}^{i+2} + \tfrac{1}{2}\pi_{k+2}^{i} + \tfrac{1}{4}\pi_{k+2}^{i-2} \quad (k = N-2, \dots, 2, 0; i = -k, \dots, k). \tag{6.54}$$

The same computational scheme is obtained from the explicit finite-difference scheme based on central differences, if one takes the time step and the space step equal to $2\Delta t$ and $2\Delta x$ respectively. While the choice $\Delta x = \sigma\sqrt{\Delta t}$ puts the single-step scheme right on the boundary of the safe region in terms of stability analysis (see (6.31)), in the two-step scheme one has

$$\frac{\sigma^2(2\Delta t)}{(2\Delta x)^2} = \tfrac{1}{2}$$

so that the stability criterion is satisfied with a considerable margin.

As we have seen, the computational method obtained from the random walk approximation of Brownian motion can be interpreted as a special case of the explicit finite-difference method. Conversely, one can say that the explicit finite-difference based on the central-difference formula has a Markov chain interpretation if the space step Δx and the time step Δt are related by $\Delta x = \sigma\sqrt{\Delta t}$. If instead the space step and the time step are related by $\Delta x = \sqrt{2}\,\sigma\sqrt{\Delta t}$, then there is a Markov chain interpretation as well, since the two-step random walk is also a Markov chain, with transition probabilities $p_{ij} = \frac{1}{4}$ if $j = i+1$ or $j = i-1$, and $p_{ii} = \frac{1}{2}$. Generally, a computational scheme for the PDE (6.52) can be given a Markov chain interpretation (in the sense that the scheme, which is defined on a certain grid in the state space, is the same as the one that would be obtained from a Markov chain defined on the same grid by the recursion $\pi_k^i = E[\pi_{k+1}^j \,|\, \hat{X}_k = x_i]$) if the recursion matrix of the scheme can be interpreted as a matrix of transition probabilities. This means that the elements of the matrix must be nonnegative, and that the sum of the entries in each row must be equal to 1. If A_2 is a matrix that represents a numerical approximation to the second order derivative, then a typical recursion matrix in the explicit method is $I + cA_2$, where $c = \frac{1}{2}\sigma^2\Delta t/\Delta x^2 \geq 0$, while in the implicit method and in the Crank-Nicolson method the corresponding matrices are $(I - cA_2)^{-1}$ and $(I - \frac{1}{2}cA_2)^{-1}(I + \frac{1}{2}cA_2)$ respectively. The condition for the row sums of a matrix M to be equal to 1 can be written in the form $M\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the vector all of whose entries are equal to 1. For any reasonable approximation A_2 of the second-order differential operator one should have $A_2\mathbb{1} = 0$, since the second derivative of a constant function is 0. This implies that $(I + \Delta t A_2)\mathbb{1} = \mathbb{1}$. From $(I - \Delta t A_2)\mathbb{1} = \mathbb{1}$ it follows that $(I - \Delta t A)^{-1}\mathbb{1} = \mathbb{1}$ as well, and likewise it can be shown that the row sum criterion is satisfied also by the recursion matrix

corresponding to the Crank-Nicolson scheme. However, the nonnegativity criterion is not always satisfied. For instance, in the explicit scheme with matrix A_2 as in (6.11), the condition for all entries of $I + cA_2$ to be nonnegative is $1 - 2c \geq 0$, or in other words $\sigma^2 \Delta t / \Delta x^2 \leq 1$. This condition is satisfied with equality in the case of the random walk interpretation (6.49). Looking at $I + cA_2$, with A_2 given by (6.11) and $c \leq \frac{1}{2}$, as a matrix of transition probabilities, the fact that the first and last rows of the matrix A_2 are filled with zeros is seen to correspond to the assumption that the outmost states are absorbing.

6.7.2 Binomial and trinomial trees

Now, let us turn to the point where many basic courses on option pricing start, namely the binomial tree method. Consider an option that will expire at time T, and take a fixed time step $\Delta t = T/N$; N is the number of time steps. Write $t_i = i\Delta t$ for $i = 0, \ldots, n$, and denote the price S_{i+1} of the underlying at time t_i by S_i. In the binomial model, it is assumed that the price of the underlying at time t_{i+1}, given the price at time t_i, can only take two values which relate to the value at time t_i by constant factors that are traditionally denoted by u (for "up") and d (for "down"). The probabilities of these two asset price changes under the risk-neutral measure are likewise constant in time; the risk-neutral probability of an "up" move is denoted by q, so that the probability under $\mathbb{Q}$ of a "down" move is $1 - q$. The model also includes a riskless asset that follows the evolution $B_{i+1} = (1 + r)B_i$. To prevent confusion between the interest rate r that appears here, which is a discretely compounded interest rate with the time step as the unit of time, and the interest rate parameter in the Black-Scholes model (which is a continuously compounded interest rate with one year as the unit of time), the latter will be denoted by r_{BS} in this section. The martingale condition for the asset S is[14]

$$S_i = \frac{1}{1+r}\big(quS_i + (1-q)dS_i\big).$$

This implies

$$q = \frac{1 + r - d}{u - d}\,. \tag{6.55}$$

The value of q that is determined in this way lies strictly between 0 and 1 under the condition

$$u > 1 + r > d$$

[14] It is actually the condition for the *relative* asset price S_i/B_i to be a martingale with respect to the probabilities q and $1 - q$.

which is required to prevent arbitrage between the riskless asset and the risky asset. The value of the underlying at the time of expiry can be written as

$$S_N = u^J d^{N-J} S_0 = S_0 \exp\big(N \log d + (\log u - \log d)J\big), \quad J \sim B(N,q) \tag{6.56}$$

where $B(N,q)$ refers to the binomial distribution with number of trials N and success probability q. The above expression can be compared to the one that holds in the Black-Scholes model under the risk-neutral measure:

$$S_T = S_0 \exp\big((r_{\text{BS}} - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,Z\big), \quad Z \sim N(0,1). \tag{6.57}$$

As is well known, the binomial distribution with parameters N and q and the normal distribution with expectation Nq and variance $Nq(1-q)$ can be looked at as approximations of each other, in particular in cases where N is large and q is not close to either 0 or 1. To make the binomial model converge to the BS model, the parameters u, d, and r should be taken such that

$$\lim_{N\to\infty} \big(Nq \log u + N(1-q)\log d\big) = (r_{\text{BS}} - \tfrac{1}{2}\sigma^2)T$$

$$\lim_{N\to\infty} Nq(1-q)(\log u - \log d)^2 = \sigma^2 T.$$

After dividing by N (i.e., multiplying by $\Delta t/T$) this can be written as

$$q \log u + (1-q)\log d = (r_{\text{BS}} - \tfrac{1}{2}\sigma^2)\Delta t + o(\Delta t) \tag{6.58}$$

$$q(1-q)(\log u - \log d)^2 = \sigma^2 \Delta t + o(\Delta t) \tag{6.59}$$

as $\Delta t \downarrow 0$. We can show the following.

Lemma 6.7.1 *The conditions (6.58) and (6.59) are satisfied if the parameters u, d, and r in the binomial model are chosen as functions of Δt in such a way that*

$$u = 1 + \sigma\sqrt{\Delta t} + o(\sqrt{\Delta t}) \tag{6.60}$$

$$d = 1 - \sigma\sqrt{\Delta t} + o(\sqrt{\Delta t}) \tag{6.61}$$

$$r = r_{BS}\Delta t + o(\Delta t) \tag{6.62}$$

and the parameter q is determined by the martingale condition (6.55).

Proof Using the standard expansion $\log(1+x) = x - \frac{1}{2}x^2 + o(x^2)$, one obtains from (6.60) and (6.61) that

$$\log u = u - 1 - \tfrac{1}{2}\sigma^2\Delta t + o(\Delta t), \qquad \log d = d - 1 - \tfrac{1}{2}\sigma^2\Delta t + o(\Delta t).$$

The martingale condition (6.55) implies that $q(u-1)+(1-q)(d-1)=r$, so that from the above and (6.62) it can be concluded that (6.58) holds. Furthermore, from (6.55) and (6.60–6.62) it follows that

$$q = \frac{\sigma\sqrt{\Delta t} + o(\sqrt{\Delta t})}{2\sigma\sqrt{\Delta t} + o(\sqrt{\Delta t})} = \frac{1+o(1)}{2+o(1)} = \tfrac{1}{2} + o(1)$$

so that $q(1-q) = \frac{1}{4} + o(1)$. Since $(\log u - \log d)^2 = (2\sigma\sqrt{\Delta t}) + o(\sqrt{\Delta t}))^2 = 4\sigma^2\Delta t + o(\Delta t)$, it is seen that (6.59) holds as well. □

The conditions above allow for various approximation schemes. In a sense, the simplest choice would be to take

$$u = 1+\sigma\sqrt{\Delta t},\ d = 1-\sigma\sqrt{\Delta t},\ r = r_{\mathrm{BS}}\Delta t,\ q = \frac{\sigma\sqrt{\Delta t} + r_{\mathrm{BS}}\Delta t}{2\sigma\sqrt{\Delta t}}$$

but one can also use

$$u = 1+\sigma\sqrt{\Delta t} + r_{\mathrm{BS}}\Delta t,\ d = 1-\sigma\sqrt{\Delta t} + r_{\mathrm{BS}}\Delta t,\ r = r_{\mathrm{BS}}\Delta t,\ q = \tfrac{1}{2}$$

or

$$u = e^{\sigma\sqrt{\Delta t}},\ d = e^{-\sigma\sqrt{\Delta t}},\ r = e^{r_{\mathrm{BS}}\Delta t} - 1,\ q = \frac{e^{r_{\mathrm{BS}}\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}\,. \tag{6.63}$$

The specification of r as in (6.63) guarantees that not only the stock, but also the bond is priced exactly in the tree model for every value of Δt, and not just asymptotically as Δt tends to 0.

Consider now the pricing of a European option whose value at time T is given by $C_T = F(S_T)$, where F is a given payoff function. Let π_k^j denote the option price at the point in the tree where k forward time steps have been taken and j "up" moves have occurred, with $0 \le j \le k$. In the binomial tree model, the price π_0^0 at time 0 is computed by setting $\pi_N^j = F(u^j d^{N-j} S_0)$ for $0 \le j \le N$, and calculating recursively

$$\pi_k^j = \frac{1}{1+r}\left(q\pi_{k+1}^{j+1} + (1-q)\pi_{k+1}^j\right) \quad (k = N-1,\dots,0;\ j = 0,\dots,k). \tag{6.64}$$

The grid points $S_k^j = u^j d^{k-j} S_0$ become equally spaced when a log transformation is applied, so it is natural to compare the binomial tree recursion to a finite-difference scheme for the Black-Scholes model written in terms of the logarithm of the stock price, as in (6.14). The difference schemes that were considered above were based on a fixed grid in the underlying; such a grid is also obtained in the binomial tree model if the approximation scheme (6.63) is adopted, which satisfies the condition

$ud = 1$.[15] The binomial tree method then falls in the category of explicit finite-difference methods which express the option price at time t in grid point i in terms of the option price at time $t + \Delta t$ in grid points $i - 1$, i, and $i + 1$. The grid points obtained from the tree method at step k under the condition $ud = 1$, after log transformation, are given by

$$\log S_0 + j \log u + (k - j) \log d = \log S_0 + (2j - k) \log u \quad (0 \leq k \leq N,\ 0 \leq j \leq k)$$

which fit into the grid

$$x_i = \log S_0 + i\Delta x \qquad (-N \leq i \leq N) \tag{6.65}$$

if one takes $\Delta x = \log u$. The tree method uses the even-numbered grid points x_i when k is even and the odd-numbered ones when k is odd. In particular, note that if index j in the tree method corresponds to index i in the fixed grid (6.65), then index $j + 1$ corresponds to $i + 2$, rather than to $i + 1$.

The explicit finite difference methods discussed above were constructed with the aim of providing approximations that are of first order in the time step Δt and of second order in the space step Δx. Writing $\hat{\pi}_k^i$ for the approximate option value at time $t = k\Delta t$ and grid point $x_i = \log S_0 + i\Delta x$, one finds the following expression for the recursion implied by an explicit scheme based on central difference formulas for the log-transformed BS equation (6.14):

$$\begin{aligned}
\hat{\pi}_k^i &= (1 - r_{\text{BS}}\Delta t)\hat{\pi}_{k+1}^i + (r_{\text{BS}} - \tfrac{1}{2}\sigma^2)\frac{\Delta t}{2\Delta x}\big(\hat{\pi}_{k+1}^{i+1} - \hat{\pi}_{k+1}^{i-1}\big) \\
&\qquad + \tfrac{1}{2}\sigma^2 \frac{\Delta t}{\Delta x^2}\big(\hat{\pi}_{k+1}^{i+1} - 2\hat{\pi}_{k+1}^{i} + \hat{\pi}_{k+1}^{i-1}\big) \\
&= \Big(1 - r_{\text{BS}}\Delta t - \sigma^2 \frac{\Delta t}{\Delta x^2}\Big)\hat{\pi}_{k+1}^i + \tfrac{1}{2}\sigma^2 \frac{\Delta t}{\Delta x^2}\big(\hat{\pi}_{k+1}^{i+1} + \hat{\pi}_{k+1}^{i-1}\big) \\
&\qquad + \tfrac{1}{2}(r_{\text{BS}} - \tfrac{1}{2}\sigma^2)\frac{\Delta t}{\Delta x}\big(\hat{\pi}_{k+1}^{i+1} - \hat{\pi}_{k+1}^{i-1}\big) \\
&= \Big(1 - r_{\text{BS}}\Delta t - \sigma^2 \frac{\Delta t}{\Delta x^2}\Big)\big(\hat{\pi}_{k+1}^i - \tfrac{1}{2}(\hat{\pi}_{k+1}^{i+1} + \hat{\pi}_{k+1}^{i-1})\big) \\
&\qquad + \tfrac{1}{2}(1 - r_{\text{BS}}\Delta t)\big(\hat{\pi}_{k+1}^{i+1} + \hat{\pi}_{k+1}^{i-1}\big) + \tfrac{1}{2}(r_{\text{BS}} - \tfrac{1}{2}\sigma^2)\frac{\Delta t}{\Delta x}(\hat{\pi}_{k+1}^{i+1} - \hat{\pi}_{k+1}^{i-1}).
\end{aligned} \tag{6.66}$$

[15] Finite-difference schemes can also be constructed on grids that shift in time. In fact, for equations of the form (6.14) this is recommended practice in cases in which the "convection term" (i.e. the term associated to the first-order derivative) is important relative to the "diffusion term" (the term associated to the second derivative).

Suppose now that Δx is chosen as a function of Δt in such a way that $\sigma^2\Delta t/\Delta x^2 \to 1$ as Δt tends to zero, or in other words

$$\frac{\sigma^2\Delta t}{\Delta x^2} = 1 + o(1) \qquad (\Delta t \downarrow 0). \tag{6.67}$$

We then have

$$(r_{\text{BS}} - \tfrac{1}{2}\sigma^2)\frac{\Delta t}{\Delta x} = \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\,\frac{\sigma\sqrt{\Delta t}}{\Delta x}\,\sqrt{\Delta t} = \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\,\sqrt{\Delta t} + o(\sqrt{\Delta t}).$$

Also note that[16]

$$\hat\pi^i_{k+1} - \tfrac{1}{2}(\hat\pi^{i+1}_{k+1} + \hat\pi^{i-1}_{k+1}) = O(\Delta x^2) = O(\Delta t) \tag{6.68}$$

and

$$\hat\pi^{i+1}_{k+1} - \hat\pi^{i-1}_{k+1} = O(\Delta x) = O(\sqrt{\Delta t}). \tag{6.69}$$

Therefore, under the condition (6.67), we can write

$$\begin{aligned}\hat\pi^i_k = (1 - r_{\text{BS}}\Delta t)\Big[\tfrac{1}{2}\Big(1 + \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t}\Big)\hat\pi^{i+1}_{k+1} \\ + \tfrac{1}{2}\Big(1 - \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t}\Big)\hat\pi^{i-1}_{k+1}\Big] + o(\Delta t).\end{aligned} \tag{6.70}$$

Note that an error of magnitude $o(\Delta t)$ is already incurred by using the first-order difference approximation for the partial derivative with respect to time, so that a modification of the expression for $\hat\pi^i_k$ by a term of this size may be considered acceptable. The equation (6.70) then shows that the recursion that is derived from the finite-difference interpretation can be written in the form of a binomial tree if the condition (6.67) is imposed. Conversely, every binomial tree method that satisfies the conditions of Lemma 6.7.1 and the condition $ud = 1$ agrees with the scheme (6.70) up to a term of size $o(\Delta t)$. This follows by noting that the condition (6.58), in the case $ud = 1$ so that $\log d = -\log u$, can be rewritten as $(2q-1)\log u = (r_{\text{BS}} - \frac{1}{2}\sigma^2)\Delta t + o(\Delta t)$ which implies (using (6.60))

$$\begin{aligned}2q - 1 &= \frac{(r_{\text{BS}} - \frac{1}{2}\sigma^2)\Delta t + o(\Delta t)}{\sigma\sqrt{\Delta t} + o(\sqrt{\Delta t})} = \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2 + o(1)}{\sigma + o(1)}\sqrt{\Delta t} \\ &= \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t} + o(\sqrt{\Delta t})\end{aligned}$$

[16] The notation $f(x) = O(g(x))$ $(x \to a)$ means that there is a constant c such that $|f(x)/g(x)| \le c$ for all x in a sufficiently small neighborhood of a; in other words, the quotient $f(x)/g(x)$ remains bounded as x tends to the limit value a.

so that

$$q = \tfrac{1}{2}\Big(1 + \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t}\Big) + q_1 \tag{6.71}$$

where $q_1 = o(\sqrt{\Delta t})$. From this one finds

$$q\pi_{k+1}^{j+1} + (1-q)\pi_{k+1}^{j} =$$
$$\tfrac{1}{2}\Big(1 + \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t}\Big)\pi_{k+1}^{j+1} + \tfrac{1}{2}\Big(1 - \frac{r_{\text{BS}} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t}\Big)\pi_{k+1}^{j} + \tfrac{1}{2}q_1(\pi_{k+1}^{j+1} - \pi_{k+1}^{j}).$$

The final term on the right hand side is of size $o(\Delta t)$ under the assumption that (6.67) holds with $\Delta x = \log u$; see (6.69). From (6.62) it follows that

$$\frac{1}{1+r} = 1 - r_{\text{BS}} + o(\Delta t).$$

Taking everything together, it is seen that the recursion from the binomial tree method subject to the conditions of Lemma 6.7.1 and the requirement $ud = 1$ is the same as the recursion from the finite-difference method subject to the condition (6.67), up to a term of size $o(\Delta t)$.

The explicit form of the finite-difference method, with the three-point approximation formulas for the differential operator in the space direction, in general leads to a trinomial computational scheme, as seen in (6.66). The trinomial scheme becomes binomial (possibly after modification by a term of order $o(\Delta t)$) under the condition (6.67). From a numerical perspective, however, there is not much to say in favor of choosing the ratio $\sigma^2\Delta t/\Delta x^2$ equal to 1; a smaller value may be preferable. In particular, the analysis in the case of the random walk approximation of Brownian motion suggests that N steps of the trinomial scheme corresponding to the value $\sigma^2\Delta t/\Delta x^2 = \frac{1}{2}$ are comparable to $2N$ steps of the binomial scheme.

The trinomial scheme is still a tree method, so that the number of points in the space domain that need to be taken into account decreases as the computation progresses. This contributes to efficiency, as compared to the implicit and Crank-Nicolson methods which do not admit reduction of the number of grid points in the state space. In the one-dimensional problems for which the tree method is designed, however, computation time is unlikely to be a major issue. Moreover, the tree method concentrates on the option value for one specific value of the underlying, whereas there are also situations in which it is desired to have option values corresponding to a range of values of the underlying; for instance, to obtain values for the sensitivity of the option value to the value of underlying ("delta" and "gamma"), or in cases where the computation is used as an intermediate stage in the valuation of a compound option.

In a tree method, the number of time steps also determines the extent of the grid in the state space at the time of maturity. For instance, when $\Delta x = \sigma\sqrt{\Delta t}$ as

in the case of the binomial method, then the grid extends to $N\Delta x = N\sigma\sqrt{T/N} = \sqrt{N}\,\sigma\sqrt{T}$ which may be considered to be rather far out if N is large, given that $\sigma\sqrt{T}$ is the standard deviation of X_T in the model example (6.46). Using a finite-difference scheme, one can opt for a more narrow grid. More attention must then be paid to possible propagation of errors that arise at the boundaries. If the number of grid points in the state space is chosen to be fairly large in response to this concern, then the stability condition for the explicit method forces the number of time steps to be large as well. By using instead an implicit method or a Crank-Nicolson method, this constraint can be removed.

6.8 Exercises

1. Compute the eigenvalues and eigenvectors of the matrix D defined in (6.30), for instance for matrix size 100. Find the largest and the smallest eigenvalue. Plot some of the eigenvectors against their index (i.e., look at the eigenvector as representing a function of a continuous variable—which is natural since the matrix D is an approximation of the operation of the second derivative on functions). What kind of behavior do you observe? Can you explain what you see? [*Hint*: consider the ordinary differential equation $y'' = \lambda y$ for negative values of λ. The eigenvectors of the matrix D can be viewed as approximate solutions of this differential equation, where λ is equal to the eigenvalue that corresponds to this eigenvector.]

2. Prove that all eigenvalues of the matrix D defined in (6.30) are in the interval $(-4, 0)$. [*Hint*: show that D is negative definite and that $D+4I$ is positive definite.] Also show that

$$\lambda_{\min} \leq -4 + 2/\sqrt{n}, \qquad \lambda_{\max} \geq -2/\sqrt{n}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and the largest eigenvalue of D respectively, and n is the size of D. [*Hint*: use the fact, for a symmetric matrix A, the smallest and the largest eigenvalue can be found as the minimum and the maximum respectively of the function $x^\top Ax/x^\top x$, defined for $x \in \mathbb{R}^n$ with $x \neq 0$.[17]]

3. In this exercise we compute the price of a call option, as a function of the price of the underlying, by means of the finite difference method. We work in the standard Black-Scholes model with the following parameters: $\sigma = 0.2$, $r = 0.04$. The call option has time of expiry $T = 1$ and strike $K = 100$.

a. Implement a finite-difference scheme using the implicit method to compute option values on a grid of stock values ranging from $S_0 = 50$ to $S_0 = 200$. Use 50 steps in the asset price direction and 40 steps in the time direction.To present the results of

[17] This function is called the *Rayleigh quotient* of A, named after John William Strutt (1842–1919), 3rd Baron Rayleigh, usually known as Lord Rayleigh; British physicist, Nobel prize 1904.

the calculation, provide the following output:

(i) a plot of the computed option price as a function of the current price of the underlying (S_0);

(ii) a plot of the difference between the computed solution and the exact solution as given by the Black-Scholes formula, also as a function of S_0;

(iii) the maximum error (i.e. the maximum absolute value of the difference between the computed value and the value obtained from the Black-Scholes value, across the stock values in the grid ranging from $S_0 = 50$ to $S_0 = 200$).

b. Repeat the computation with the explicit method and with the Crank-Nicolson method.

c. For each of the three methods, attempt to find an "economical" (in the sense of computational time) combination of the numbers of steps in the asset price direction and in the time direction, achieving a maximum error less than 0.01.

4. Same questions as in Exc. 3, but this time using logarithmically transformed variables.

5. Write a routine that computes the price of a European option in the standard Black-Scholes model, as a function of the current value of the underlying, for an arbitrary payoff. The payoff is assumed to be given as a column vector of values defined on a logarithmic grid with a fixed step size that is supplied as an input to the routine. Apply a finite-difference method with a time step that is also supplied as an input parameter, and use the Crank-Nicolson method to do the time stepping.

6. Consider a Bermudan put option on an underlying asset with current value $S_0 = 100$, strike $K = 100$, time of maturity $T = 1$, and with six regularly spaced exercise dates until maturity. Assume the Black-Scholes model holds with asset volatility $\sigma = 0.2$ and interest rate $r = 0.04$. Compute numerically the price of the option at $t = 0$. [*Hint*: the routine of Exc. 5 can be of use here.] Compare the solution that you obtain to the value of a perpetual American put with the same parameter values.

This page intentionally left blank

Chapter 7

Monte Carlo methods

The Monte Carlo method for valuation of financial instruments is very popular in practice. Implementation is often straightforward, once a model for the underlying assets under a pricing measure has been given. In this chapter, several issues are discussed that may arise in the use of the method: error analysis, variance reduction, computation of sensitivities, and application to compounded options.

7.1 Basic Monte Carlo

The Monte Carlo method can be applied whenever the quantity that is to be computed takes the form of an expectation of a given random variable, and it is possible to draw arbitrarily many samples from that variable. For instance, the surface area of a circle with diameter 1 is equal to $E[f(U_1, U_2)]$, where U_1 and U_2 are independent random variables which are both uniformly distributed on the interval $[-\frac{1}{2}, \frac{1}{2}]$, and $f(x, y)$ is the function defined by $f(x, y) = 1$ if $x^2 + y^2 \leq \frac{1}{4}$ and $f(x, y) = 0$ otherwise. So the area of the circle can be computed by the Monte Carlo technique.[1] In the field of finance, many quantities that need to be computed appear naturally in the form of expectations. Especially in situations where several state variables are involved, the use of the Monte Carlo method can be attractive.

The Monte Carlo estimate of the quantity EX is simply

$$\text{MCE} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7.1}$$

where the x_i's $(i = 1, \ldots, n)$ are independent draws from the distribution of X. By definition, therefore, the Monte Carlo estimate is a random variable. Its expectation is equal to the quantity EX that needs to be computed; under the assumption that

[1] This example is not one in which the Monte Carlo method would be the method of choice.

$\mathrm{var}(X)$ is finite, the variance of the Monte Carlo estimate is also finite and equals

$$\mathrm{var}(\mathrm{MCE}) = \frac{1}{n^2}\big(n \cdot \mathrm{var}(X)\big) = \frac{1}{n}\,\mathrm{var}(X). \tag{7.2}$$

Therefore, with 95% confidence we can say that the value of the quantity we want to compute lies in the interval

$$[\mathrm{MCE} - 1.96\frac{1}{\sqrt{n}}\,\mathrm{std}(X), \mathrm{MCE} + 1.96\frac{1}{\sqrt{n}}\,\mathrm{std}(X)]. \tag{7.3}$$

Typically we do not know the exact value of the standard deviation of X, but this quantity can be estimated itself by

$$\hat{s} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mathrm{MCE})^2}.$$

When $\mathrm{std}(X)$ is replaced by $\hat{s}$ in (7.3), the resulting confidence interval should be interpreted with care because $\hat{s}$ itself is an estimated quantity.

It follows from (7.3) that the only assumption needed to make the Monte Carlo method converge as n tends to infinity is that $\mathrm{var}(X)$ should be finite. On the other hand, the same expression also shows that the rate of convergence is $\sqrt{n}$. In other words, to get one more decimal of accuracy the number of draws should be increased by a factor 100. From this point of view, the Monte Carlo method is slow indeed. On the other hand, the rate of convergence does not depend on factors that badly affect some other methods, such as the dimension of the space from which the sample points are drawn, and therefore the Monte Carlo method can be a key resource. The method provides an answer, although typically not at a high level of precision, in circumstances where no other methods are feasible. In financial engineering, where often models are complicated but high accuracy is not required, the Monte Carlo method is a very popular tool.

The Monte Carlo method can be applied in a straightforward way to European options which expire at a given date, and with minor modifications it also works in cases where the option expires at a stopping time that is determined by the state variables, such as in the case of a barrier option. However, when exercise is at the discretion of the holder as in the case of American options, then application of Monte Carlo becomes more problematic. The related problem of *nesting* will be discussed in Section 7.4 below. Also, one should keep in mind that convergence of the Monte Carlo method is an asymptotic result, and that there will always be cases in which the asymptotics will only kick in at numbers of samples that are beyond what is practically feasible. Some applications in finance do have a tendency to produce such cases, both in option pricing (far out-of-the-money options) and in

risk management (tail risk). In such cases we have *bad matching* between, on the one hand, the probability density of the samples, and on the other hand, the behavior of the relevant payoff function. A remedy may be found by the technique of importance sampling, discussed in Section 7.2.2 below.

To illustrate the basic Monte Carlo method, suppose that we want to price a European option given by an expiry date T and a payoff function $F(x)$. To apply pricing by the NDPF, first select a numéraire N with pricing function $\pi_N = \pi_N(t, x)$. The pricing function that we are looking for is given by

$$\pi(0,x) = \pi_N(0,x) E^{\mathbb{Q}_N}\Big[\frac{F(X_T)}{\pi_N(T,X_T)} \,\Big|\, X_0 = x\Big]. \tag{7.4}$$

To approximate this, generate n trajectories of the stochastic differential equation

$$dX = (\mu_X - \sigma_X\lambda_N)dt + \sigma_X dW, \quad X_0 = x \tag{7.5}$$

where λ_N is the drift adjustment corresponding to the measure $\mathbb{Q}_N$, and W is Brownian motion. Let $x_T^1, \ldots, x_T^n$ be the values of X_T found in the simulations; then an approximation to (7.4) is

$$\pi(0,x) \simeq \pi_N(0,x)\frac{1}{n}\sum_{i=1}^{n}\frac{F(x_T^i)}{\pi_N(T,x_T^i)}. \tag{7.6}$$

In the pricing kernel method, the formula from which we work is

$$\pi(0,x) = E^{\mathbb{P}}[K_T F(X_T) \mid X_0 = x]. \tag{7.7}$$

To find an approximation by simulation, produce n trajectories of the system of SDEs

$$dX = \mu_X dt + \sigma_X dW \tag{7.8}$$

$$dK = -K(r dt + \lambda^\top dW). \tag{7.9}$$

Let $(x_T^1, k_T^1), \ldots, (x_T^n, k_T^n)$ be the pairs of values of X_T and K_T found in the simulations. An approximation to (7.7) is then

$$\pi(0,x) \simeq \frac{1}{n}\sum_{i=1}^{n} k_T^i F(x_T^i). \tag{7.10}$$

In some cases, it is possible to sample directly from the distribution of the state variables at the time of maturity. For instance, this happens in the Black-Scholes model where we know that the price of the underlying follows a lognormal distribution and the parameters of this distribution are available. More typically

however, the values of X_T need to be computed on the basis of a time-stepping method as described above, which means that the distribution of X_T as given by the model can only be approximated. In addition to the Monte Carlo error there is a simulation error. To get convergence to the true value[2] one needs to increase the number of Monte Carlo samples *and* the number of time steps. An assessment of the relative emphasis that should be placed on each of these can be made on the basis of some assumptions concerning the effect on overall accuracy of increased effort in each of the two directions. Analysis shows that the accuracy of expressions of the form $Ef(X_T)$ improves in proportion to the number of time steps when the Euler method is used.[3] Note that the discretization error creates a bias, whereas the Monte Carlo error causes a variance. The two effects are considered jointly in the root mean squared error (RMSE) which is defined by

$$\text{RMSE} = \sqrt{\text{bias}^2 + \text{variance}}\,.$$

The bias is approximately equal to c_1/N, where N is the number of time steps and c_1 is a constant, and the variance, as discussed above, is approximately equal to c_2/n where n is the number of trajectories generated and c_2 is a constant. It would be reasonable to take the product nN as a measure of the computational effort. For a given level of the computational effort, say α, the RMSE is minimized by selecting n and N such that

$$\frac{c_1^2}{N^2} + \frac{c_2}{n} \to \min \qquad \text{subject to } nN = \alpha.$$

Under the effort constraint, we have

$$\text{RMSE}^2 = \frac{c_1^2 n^2}{\alpha^2} + \frac{c_2}{n}$$

which is minimized (treating n as a continuous variable for convenience) at $n = (\frac{1}{2}c_2/c_1^2)^{\frac{1}{3}}\alpha^{\frac{2}{3}}$. The corresponding value of the number of time steps N is $(\frac{1}{2}c_2/c_1^2)^{-\frac{1}{3}}\alpha^{\frac{1}{3}}$, and

$$\text{RMSE} = c\alpha^{-\frac{1}{3}}$$

where c is a constant determined by c_1 and c_2. In other words, if a certain amount of effort is already divided optimally between the number of trajectories and the number of time steps, and one wants to get one extra decimal of accuracy, one needs to increase the number of time steps by a factor of 10 and the number of

[2]The term "true" is to be interpreted here in the sense of the chosen model. The meaningfulness of the computed value for practical purposes depends on the quality of that model.

[3]In more technical terms: the weak order of convergence of the Euler method is 1. The notion of convergence used here is different from the notion used in Exc. 2.8.22; the latter notion corresponds to what is called the *strong* order of convergence.

trajectories by a factor of 100, so that the total computational effort required is increased by a factor of one thousand.

Simply increasing the number of sample points in order to get better accuracy may therefore not be a practically feasible solution. As an alternative, one may try to reduce the variance of estimates. There are various ways of doing this; the ones that are most frequently used in financial applications are discussed below.

7.2 Variance reduction

As discussed above, the Monte Carlo method can be slow depending on the required level of accuracy and the nature of the option to be priced. Considerable savings may be achieved if instead of EX we can compute EY where Y is a random variable that has the same expectation as X but a smaller variance. More generally, we may also change the probability measure and compute $E^Q Y$ instead of $E^P X$, where the measure Q and the random variable Y are such that $E^Q Y$ equals $E^P X$ and $E^Q(Y - E^Q Y)^2$ is less than $E^P(X - E^P X)^2$. Such replacement strategies are known as *variance reduction methods*. We discuss here some of the best-known variance reduction methods and the way that these may be applied in option pricing applications.

7.2.1 Control variates

The general method of control variates proceeds as follows. Suppose that the quantity that we want to compute is EX, where X is a random variable. Let Y be a random variable on the same sample space with the following properties: (i) Y is (preferably strongly) correlated to X, and (ii) we know EY. The variable Y is called the "control variate". Suppose, for the moment, that we also know the correlation coefficient ρ_{XY} between X and Y and the variances σ_X^2 and σ_Y^2. Define now a new random variable Z by

$$Z = X - \alpha(Y - m) \tag{7.11}$$

where $m = EY$ (supposed known) and α is any chosen constant. Then we have $EZ = EX$, and so to compute EX we can also take samples from Z. While the means of the random variables Z and X are the same, their variances may well be different, since

$$\operatorname{var} Z = \operatorname{var} X + \alpha^2 \operatorname{var} Y - 2\alpha \operatorname{cov}(X, Y).$$

If we choose

$$\alpha = \frac{\operatorname{cov}(X, Y)}{\operatorname{var} Y} = \rho_{XY} \frac{\sigma_X}{\sigma_Y} \tag{7.12}$$

then

$$\operatorname{var} Z = (1 - \rho_{XY}^2) \operatorname{var} X \tag{7.13}$$

so that there can be a substantial reduction of the variance if the correlation between X and Y is sufficiently strong. For instance, to get a reduction of the variance by a factor of 4 (which halves the confidence interval; this can be achieved alternatively by using four times as many sample points), one needs a correlation coefficient of $\frac{1}{2}\sqrt{3} = 0.87$.

The optimal coefficient α as given in (7.12) is expressed in terms of the statistical quantities $\operatorname{cov}(X, Y)$ and $\operatorname{var} Y$. Since the computation is based on the presumption that the random variable X is sufficiently complicated so that its expectation has to be found by a numerical method, it is unlikely in applications that any statistical quantities related to X would be available analytically. It is possible that $\operatorname{var}(Y)$ could be obtained by analytic methods, but the computation may be cumbersome. Instead one can work with estimates of $\operatorname{cov}(X, Y)$ and $\operatorname{var} Y$ on the basis of their sample equivalents. The Monte Carlo estimate

$$\mathrm{MCE} = \frac{1}{n}\sum_{i=1}^{n}\big(x_i - \alpha(y_i - m)\big) \tag{7.14}$$

then depends on the samples (x_i, y_i) in a complicated way, so that the expression (7.2) for the variance of this estimate is no longer valid. It can be shown, however, that the error is small in large samples.[4] To avoid the issue, one might estimate $\operatorname{cov}(X, Y)$ and $\operatorname{var} Y$ from a *pilot sample* that is independent from the sample used for Monte Carlo computation of the actual quantity of interest.

The method of control variates can be used in applications to option pricing for instance when a given problem is a minor modification of one that is analytically solvable. For instance, let European options C_1 and C_2 with the same time of expiry T be defined by payoff functions $F_1(x)$ and $F_2(x)$, and suppose that an analytic solution formula is known for the latter but not for the former. If the functions F_1 and F_2 are close to each other, it is reasonable to expect that the random variables $F_1(X_T)$ and $F_2(X_T)$ are strongly correlated. Moreover, we know $E^{\mathbb{Q}}F_2(X_T)$ from the analytical pricing formula that is available for the second option. Consequently, the method of control variates can be applied. In this way, explicit solution formulas become useful even in situations where the assumptions that underlie the analytic solution are not exactly satisfied.

Another way in which the control variate method can be applied in finance is the following. Consider an option pricing problem formulated in terms of an equivalent

[4] In contrast, if the quantity m in (7.14) would be replaced by its sample equivalent $\frac{1}{n}\sum_{i=1}^{n} y_i$, then the error that is made by using the formula (7.2) to determine the confidence interval is large even in large samples. When the confidence interval is computed correctly, it turns out that application of (7.14) with the sample equivalent of m does not produce any advantage with respect to the "raw" Monte Carlo estimate. This is to be expected, since the strength of the control variate method is based on the fact that some additional knowledge is used, namely the exact value of EY.

martingale measure. From the general theory, we know that the relation

$$\frac{V_T}{N_T} - \frac{V_0}{N_0} = \int_0^T \phi(t, X_t)^\top \, d\frac{Y_t}{N_t} \tag{7.15}$$

holds, where V is a replicating portfolio for a given European option and $\phi = \phi(t, x)$ is the corresponding hedge strategy. The hedge strategy can be computed if the pricing function is known, but of course here we are considering the situation where the pricing function is what we want to compute, so we cannot assume that it is available. However, there may be some approximate hedge available, perhaps from an analytical solution of a related problem, or from some plausible rule of trading. Let $\tilde{\phi}(t, x)$ denote the approximate hedge. We know that

$$E^{\mathbb{Q}} \int_0^T \tilde{\phi}(t, X_t)^\top \, d\frac{Y_t}{N_t} = 0$$

because the relative price process Y_t/N_t is a martingale under $\mathbb{Q}$. Also, if the approximate hedge does reasonably well, the variance of the difference

$$\frac{F(X_T)}{N_T} - \int_0^T \tilde{\phi}(t, X_t)^\top \, d\frac{Y_t}{N_t}$$

is small. (Indeed, if the hedge would be perfect, as in (7.15), the variance of the difference would be zero.) Therefore, the integral can be used as a control variate. Of course, in the actual implementation, the integral must be approximated by a finite sum. Assuming that state trajectories are simulated on a time grid $0 = t_0, \ldots, t_k = T$, we would estimate the option value (relative to the numéraire) by

$$\frac{C_0}{N_0} \simeq \frac{1}{n} \left[\frac{F(x_T^i)}{\pi_N(T, x_T^i)} - \sum_{j=1}^{k-1} \tilde{\phi}(t_j, x_{t_j}^i)^\top \left[\frac{\pi_Y(t_{j+1}, x_{t_j+1}^i)}{\pi_N(t_{j+1}, x_{t_j+1}^i)} - \frac{\pi_Y(t_j, x_{t_j}^i)}{\pi_N(t_j, x_{t_j}^i)} \right] \right]. \tag{7.16}$$

The discussion above has concentrated on the use of a single control variate, but one can also use several control variates together. Let these variables be denoted by $Y_1, \ldots, Y_n$, with $EY_i = m_i$ for $i = 1, \ldots, n$. The control variate estimate is

$$Z = X - \sum_{i=1}^{n} \alpha_i (Y_i - m_i). \tag{7.17}$$

The value of the parameter vector $\alpha := [\alpha_1 \cdots \alpha_n]^\top$ that minimizes the variance of the control variate estimate is found as the solution of the matrix-vector equation

$$\Sigma_{YY} \alpha = \Sigma_{XY} \tag{7.18a}$$

where the matrix Σ_{YY} and the vector Σ_{XY} are defined by

$$(\Sigma_{YY})_{ij} = \mathrm{cov}(Y_i, Y_j) \quad (i = 1, \ldots, n;\ j = 1, \ldots, n) \tag{7.18b}$$

$$(\Sigma_{XY})_i = \mathrm{cov}(X, Y_i) \quad (i = 1, \ldots, n). \tag{7.18c}$$

Alternatively, it may be noted that the variance of Z is is equal to the minimum of $E[(Z - \alpha_0)^2]$ across all possible values of α_0, so that the parameters α_i that achieve the minimal variance of Z can also be found from the solution of the optimization problem

$$E\big[(X - \alpha_0 \mathbb{1} - \textstyle\sum_{i=1}^n \alpha_i Y_i)^2\big] \quad \to \quad \min. \tag{7.19}$$

This is a least-squares minimization problem. In particular, when the quantities $\mathrm{cov}(X, Y_i)$ are not available analytically, as is likely in applications, and the quantities $\mathrm{cov}(Y_i, Y_j)$ are also not available or found too cumbersome to compute, then estimates of the optimal parameter values α_i can be obtained from a standard linear regression of sample outcomes of X against sample outcomes of the variables Y_i and a vector of constants. As in the case of a single control variate, a pilot sample may be used for this purpose in order to avoid a possible source of bias in the estimated confidence interval. The reduction of the variance will become larger when X is approximated more accurately by a linear combination of control variates. For instance, in a situation in which the variable of interest is generated by Brownian motion on an interval $[0, T]$, one might use polynomials in W_T, making use of the fact that the moments of W_T are known exactly.

7.2.2 Importance sampling

Under some adverse circumstances, the Monte Carlo method may perform very badly, and, what is worse, may fail to indicate that it does not do well. To give a somewhat artificial example of such a situation, consider the computation of $Ef(Z)$ when Z is the uniform distribution on the interval $[0, 1]$ and f is the function defined by

$$f(x) = \begin{cases} \frac{1}{\varepsilon} & \text{for } 0 \le x \le \varepsilon \\ 0 & \text{for } \varepsilon < x \le 1. \end{cases}$$

where $\varepsilon < 1$ is a constant. The exact value of the expectation is 1. However, if for instance ε is equal to 10^{-6} and the number of trials used for the Monte Carlo computation is 10 000, the probability is quite high that all samples taken will return the value 0. The result obtained from the computation is then simply 0, and the estimated confidence interval only contains this point. In such a case, due to the limited number of samples taken, the Monte Carlo method does not "see" the small region of the outcome space where very large outcomes occur. As a result,

the method "believes" that the function to be integrated is identically zero, and it reports a completely wrong answer, not just in terms of the point estimate that is produced but also in terms of the estimated confidence interval. For a given value of ε, the correct answer will be still be retrieved *asymptotically* as predicted by the theory, but the asymptotic properties may only take substantial effect at numbers of trials that are not typically used in practice and that might be infeasible in terms of computation time.

In a more general setting, assume that we are aiming to compute a quantity of the form $E[X]$ where X is a random variable that can only take a finite number of values, say $\{x_1, \ldots, x_n\}$. Then

$$EX = \sum_{i=1}^{n} x_i P(X = x_i). \tag{7.20}$$

Suppose first that all x_i's are nonnegative. The expectation EX is then a sum of nonnegative terms which are typically of different size. The accuracy by which EX is computed depends mainly on how accurately the largest of those terms are calculated. Each of the terms is computed as the product of the value x_i and the probability $P(X = x_i)$. It may happen that, for some value of i, the probability $P(X = x_i)$ is very small, but nevertheless the term $x_i P(X = x_i)$ contributes substantially to the sum in (7.20), because x_i is relatively large. Since the probability $P(X = x_i)$ is small, its value is not determined accurately in a Monte Carlo experiment of practical size. The error in $P(X = x_i)$ is transmitted to the product $x_i P(X = x_i)$, and since this term is a substantial contributor to the expectation as a whole, the accuracy of the computation of EX is affected badly. However, the inaccuracy may not be reflected in the standard confidence interval of the Monte Carlo procedure. The use of the confidence interval is based on convergence to normality as is guaranteed by the central limit theorem, but in a situation in which the expectation of the random variable under consideration is determined to a large extent by very high outcomes in rare events, the asymptotic regime may not have set in yet, even at quite high sample sizes. If the computation of EX is inaccurate for this reason, then the computation of $\mathrm{std}(X)$ is likely to be inaccurate as well, so that, as seen in the example above, it may well happen that the computed confidence interval is small whereas the error in the computed expectation is large.

Further aggravation of computational problems can arise when some of the x_i's are positive and some are negative. If the sum of the positive terms is approximately equal in absolute value to the sum of the negative terms, then the expectation EX is close to zero, and even small inaccuracies in either the sum of the positive terms or the sum of the negative terms can cause a large relative error in the computed value of EX. Computational problems of this type can be extremely

hard to solve accurately. Fortunately, in many applications in finance (in particular option pricing) the integrand is nonnegative. Also, in cases where the computational challenge is due to near-cancellation of positive and negative terms but not so much to large outcomes in rare events, the confidence interval would be fairly reliable, so that the user is appropriately alerted; note that, in situation where EX is small relative to the expectations of the positive and negative parts of X, the variance is largely determined by the second moment, which is an expectation of a nonnegative random variable so that the near-cancellation problem does not arise.

Situations in which the price of a financial contract is determined to a large extent by outcomes in rare cases occur quite frequently. One may think for instance of credit insurance. Another typical example would be a far out-of-the-money put option, i.e. a put with a strike that is much lower than the current value of the underlying. In such cases, it is advisable to modify the sampling density in such a way that more samples fall into the regions that are important for the computation of the expectation, while making sure to correct for the change of density by introducing an appropriate factor that multiplies the values. The technique that aims for such modifications is called *importance sampling*. In general, it works as follows. Suppose that the random variable X is obtained as a function of another random variable (or random vector) that can be simulated; say $X = f(Z)$. The expectation $E[f(Z)]$ can be written as an integral by making use of the notation

$$E[f(Z)] = \int f \, dP$$

where, for any measurable subset A of the outcome space $\mathbb{R}^n$ of Z, $P(A)$ is the probability of a sample from the distribution of Z to fall in A. Suppose now that $\theta(z)$ is a positive function defined on the outcome space $\mathbb{R}^n$ of the vector random variable Z, and moreover assume that θ satisfies $E[\theta(Z)] = 1$. Then the prescription

$$Q(A) = P(A)\, E[\theta(Z) \mid Z \in A]$$

defines a new probability measure Q; in fact, θ is the Radon-Nikodym derivative of Q with respect to P. If the measure P has a density $\phi(z)$, then the density of Q is $\theta(z)\phi(z)$. More generally we can write $dQ = \theta \, dP$. Consequently,

$$\int f \, dP = \int \frac{f}{\theta} \, \theta \, dP = \int \frac{f}{\theta} \, dQ.$$

The expression on the right hand side can be computed by the Monte Carlo method if we are able to generate samples of Z according to the probability measure Q.

When samples $z_1, \ldots, z_n$ have been generated, the quantity

$$\frac{1}{n}\sum_{i=1}^{n}\frac{f(z_i)}{\theta(z_i)}$$

is a point estimate of $E^P[f(Z)] = E^Q[f(Z)/\theta(Z)]$. Dividing the values $f(z_i)$ by $\theta(z_i)$ provides compensation for the fact that the probability of drawing z_i under $\mathbb{Q}$ is $\theta(z_i)$ times larger than the probability of drawing z_i under $\mathbb{P}$.[5] An estimated confidence interval can be obtained from the variance of the samples $f(z_i)/\theta(z_i)$, $i = 1 \ldots, n$. The exact value of the variance is

$$\begin{aligned}\mathrm{var}^Q(f(Z)/\theta(Z)) &= E^Q\left[(f(Z)/\theta(Z))^2\right] - \left(E^Q[f(Z)/\theta(Z)]\right)^2 \\ &= E^P\left[f^2(Z)/\theta(Z)\right] - \left(E^P[f(Z)]\right)^2. \end{aligned} \tag{7.21}$$

One sees from this that the change of measure together with the corresponding compensation in the function values does not affect the expectation of $f(Z)$, as intended, but does affect the variance. Importance sampling is therefore not only used as a way of avoiding the problems with relatively large outcomes in rare events that have been described above; it can be applied as a general variance reduction technique.

An important consideration in practice is that it must be feasible to sample from the modified density of the variable Z. In applications, therefore usually first the new sampling density $\psi(z)$ is selected in such a way that sampling from this density is feasible, and then the corresponding Radon-Nikodym derivative is defined by $\theta(z) = \psi(z)/\phi(z)$. Numerical considerations impose certain limits on the extent to which densities can be modified; if the densities ψ and ϕ differ too much, then for some outcomes z_i the quotient $\psi(z_i)/\phi(z_i)$ may become hard to compute for standard computational software.

As an illustration of importance sampling, consider a one-dimensional example. Let us assume that we want to compute $\zeta = E[f(Z)]$ where f is the characteristic function of the interval $[2, 4]$, and Z is the standard normal variable. In other words, we want to compute the probability of a standard normal variable to take a value between 2 and 4. This quantity is readily available from tables of the normal distribution ($\zeta = \Phi(4) - \Phi(2) = 0.0227$) and even if that would not be so, the Monte Carlo method would not be the best method to use for this particular computational

[5] This is a somewhat loose statement. More precisely, the probability under $\mathbb{Q}$ to draw a sample in a ball of radius ε around z_i is, up to first order in ε, equal to $\theta(z_i)$ times the probability under $\mathbb{P}$ to draw a sample in the same ball.

problem; the example is just meant to be illustrative. We have

$$\zeta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{2 \leq x \leq 4} \, e^{-\frac{1}{2}x^2} \, dx.$$

The standard normal distribution generate most of its outcomes outside the interval $[2, 4]$. There would be a better match between the function and the distribution if we would sample from the normal distribution $N(3, 1)$; moreover, this is a distribution from which we can indeed easily draw samples. The density of the distribution $N(3, 1)$ is

$$\frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(x-3)^2} \, .$$

Therefore, let us define

$$\theta(x) = \frac{\frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(x-3)^2}}{\frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}x^2}} = e^{3x - \frac{9}{2}} \, .$$

We can write

$$\zeta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{2 \leq x \leq 4} \, e^{-3x + \frac{9}{2}} \, e^{-\frac{1}{2}(x-3)^2} \, dx.$$

Therefore, an alternative way to compute the quantity of interest is to generate samples z_i from the $N(3, 1)$ distribution and to compute the corresponding values

$$1_{2 \leq z_i \leq 4} \, e^{-3z_i + \frac{9}{2}} \, .$$

Taking the average of these values leads to an estimate for ζ. The variance per sample according to the original distribution is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1_{2 \leq x \leq 4}\right)^2 e^{-\frac{1}{2}x^2} \, dx - \zeta^2 = \zeta - \zeta^2 = 0.0222.$$

To compute the exact variance of each sample obtained from the distribution $N(3, 1)$ as described above, note that

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1_{2 \leq x \leq 4} \, e^{-3x + \frac{9}{2}}\right)^2 e^{-\frac{1}{2}(x-3)^2} \, dx \\
&\quad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{2 \leq x \leq 4} \, e^{-6x+9} \, e^{-\frac{1}{2}(x-3)^2} \, dx \\
&\quad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{2 \leq x \leq 4} \, e^{-\frac{1}{2}(x^2 - 3x + \frac{9}{2})} \, dx \\
&\quad = \frac{e^9}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1_{2 \leq x \leq 4} \, e^{-\frac{1}{2}(x+3)^2} \, dx = e^9(\Phi(7) - \Phi(5)).
\end{aligned}$$

It follows that the modified variance is

$$e^9(\Phi(7) - \Phi(5)) - \zeta^2 = 0.0018.$$

By application of importance sampling, the width of the confidence interval obtained from a given number of Monte Carlo is samples is reduced by the factor $\sqrt{0.0018/0.0222} = 0.285$. Approximately the same reduction would be achieved by a 12-fold increase of the number of samples; so it can be said that the use of importance sampling in this case reduces computation time (for a given level of accuracy) by a factor of 12.

For the application of the importance sampling method, one has to be able to draw samples from the new distribution Q, and one needs to be able to compute the value of the Radon-Nikodym derivative dQ/dP at every point in the outcome space. In the context of SDEs, Girsanov's theorem (Thm. 2.7.1) can be put to good use. The new measure is effectively defined by a change of drift, and the equation (2.92) keeps track of the RN derivative. If the drift is changed by a constant, then the SDE (2.92) does not need to be simulated, because in that case we are dealing with a geometric Brownian motion and the required value can be expressed directly in terms of the value W_T of the driving Brownian motion at time T.

7.2.3 Antithetic variables

If the random vector Z follows a multivariate normal distribution with zero mean, then the distribution of Z is the same as the distribution of $-Z$. It follows that, when f is any function such that $E[f(Z)]$ is defined, the equality $E[f(Z)] = E[f(-Z)]$ holds. When $z_1, \ldots, z_n$ are independent draws from the distribution of Z, we may therefore use for instance

$$\frac{1}{2n}\sum_{i=1}^{n} f(z_i) + \frac{1}{2n}\sum_{i=1}^{n} f(-z_i) \tag{7.22}$$

as an estimate of $E[f(Z)]$. The estimate is unbiased. To determine its variance, look at the above expression as a realization of the random variable

$$\hat{\zeta} := \frac{1}{n}\sum_{i=1}^{n} \tfrac{1}{2}(f(Z_i) + f(-Z_i))$$

in which the Z_i's are independent random variables all having the same distribution as Z. Define $\bar{X}_i = \frac{1}{2}(f(Z_i) + f(-Z_i))$; then the $\bar{X}_i$'s are independent and identically distributed, so that $\mathrm{var}(\hat{\zeta}) = \frac{1}{n}\mathrm{var}(\bar{X})$, where $\bar{X}$ denotes a random variable with

the same distribution as all $\bar{X}_i$'s. Moreover we have

$$\begin{aligned}\operatorname{var}(\bar{X}) &= \tfrac{1}{4}\big(\operatorname{var}(f(Z)) + 2\operatorname{cov}(f(Z), f(-Z)) + \operatorname{var}(f(-Z))\big)\\ &= \tfrac{1}{2}\big(\operatorname{var}(f(Z)) + \operatorname{cov}(f(Z), f(-Z))\big)\end{aligned}$$

because $\operatorname{var}(f(Z)) = \operatorname{var}(f(-Z))$. Overall, we find

$$\operatorname{var}(\hat{\zeta}) = \frac{1}{2n}\big(\operatorname{var}(f(Z)) + \operatorname{cov}(f(Z), f(-Z))\big). \tag{7.23}$$

If we assume that, in computing a value $f(z)$, the lion's part of the work is in the application of the function f rather than in obtaining z as a draw from the random variable Z, and if we also assume that no shortcuts are available which make it easy to compute $f(-z)$ once $f(z)$ has been computed, then the amount of computational effort in obtaining the values $f(z_1), \dots, f(z_n), f(-z_1), \dots, f(-z_n)$ is about the same as the effort in obtaining $2n$ values $f(z_1), \dots, f(z_{2n})$. The variance in (7.23) should then be compared with the variance that would be obtained from $2n$ replications, namely $\frac{1}{2n}\operatorname{var}(f(Z))$. It follows that the estimate (7.22) is an improvement on the "raw" Monte Carlo estimate when $f(Z)$ and $f(-Z)$ are negatively correlated. The improvement can be substantial if there is a strong negative correlation. The method described here is called the method of *antithetic variables*.

7.3 Price sensitivities (the Greeks)

The term "Greek" is used in finance to refer to derivatives (in the mathematical sense of the word). In particular, derivatives of option prices with respect to various parameters/variables are used. The term "Greeks" covers both first-order and higher-order derivatives. Examples are:

delta	$\dfrac{\partial C}{\partial S}$	gamma	$\dfrac{\partial^2 C}{\partial S^2}$	vega	$\dfrac{\partial C}{\partial \sigma}$
theta	$\dfrac{\partial C}{\partial t}$	vanna	$\dfrac{\partial^2 C}{\partial S \partial \sigma}$	vomma	$\dfrac{\partial^2 C}{\partial \sigma^2}$

Greeks are indicators of (local) sensitivity of an option price or the result of a portfolio strategy with respect to a given variable or parameter. This information is important for the computation of hedging strategies (for instance delta hedging), but also for risk management, optimization, and robustness analysis.

The most common way of computing sensitivity of a quantity computed by Monte Carlo is the *bump and reprice* method, also known as the *finite-difference* method. The problem is to compute

$$\frac{d}{d\theta} E[X(\theta)]$$

where θ is a parameter of interest. Let the standard MC estimate based on n samples be denoted by $\bar{X}_n$. The one-sided finite-difference estimate of the derivative is

$$\frac{\bar{X}_n(\theta + \Delta\theta) - \bar{X}_n(\theta)}{\Delta\theta}$$

where $\Delta\theta$ is a small step. An alternative is the two-sided (a.k.a. *central-difference*) estimate:

$$\frac{\bar{X}_n(\theta + \Delta\theta) - \bar{X}_n(\theta - \Delta\theta)}{2\Delta\theta}.$$

This requires an extra set of simulations but is often more accurate, since the Taylor series expansion shows that

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x) + O\big(\Delta x\big)$$

$$\frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} = f'(x) + O\big((\Delta x)^2\big).$$

The same observation was used in the construction of finite-difference methods of PDEs in Chapter 6. Both the one-sided and the two-sided estimator incorporate a *bias* which is due to the fact that we are using a finite-difference approximation. To determine the *variance*, assume first that estimates $\bar{X}_n(\theta)$, $\bar{X}_n(\theta+h)$, and $\bar{X}_n(\theta-h)$ are obtained from *independent* simulations. The variance of $X(\theta + \Delta\theta)$ is in cases of interest close to the variance of $X(\theta)$ when $\Delta\theta$ is small, so that we can write

$$\operatorname{var}\big((\bar{X}_n(\theta + h) - \bar{X}_n(\theta))/h\big) \simeq \frac{2}{nh^2} \operatorname{var}\big(X(\theta)\big).$$

A similar expression holds for the two-sided difference. To have a small variance for the estimator, the quantity nh^2 must be large; in particular is is not advisable to take a small step h when the number of samples n is moderate. For more specific advice, consider what is needed to minimize the RMSE. In the case of the one-sided estimate, minimization of the RMSE requires that the step size h for a given number of simulations n is selected such that

$$c_1 h^2 + \frac{c_2}{nh^2} \to \min.$$

This leads to $h \propto n^{-1/4}$, and

$$\text{RMSE} = O\big(n^{-1/4}\big)$$

for the one-sided estimate with independent sampling. In the case of the two-sided

estimate with independent sampling, we get

$$\text{RMSE} = O\big(n^{-1/3}\big).$$

This is still a rather slow convergence.

It is not a great restriction of generality to assume that the parameter-dependent quantity of interest $X(\theta)$ can be written in the form

$$X(\theta) = f(Z, \theta) \tag{7.24}$$

where f is a given function and Z is an underlying (possibly multivariate) random variable. Indeed, the variables that we study by means of simulations are always generated in this way.[6] It is then possible to use *common random numbers* in the simulation of $X(\theta)$, $X(\theta + \Delta\theta)$, and $X(\theta - \Delta\theta)$; this means that samples of $X(\theta)$ and of $X(\theta \pm \Delta\theta)$ are all generated from the same set of samples for Z. Correspondingly, the random variables $X(\theta)$ and $X(\theta \pm \Delta\theta)$ are not independent, so that the expressions for the variance of the one-sided and the two-sided finite difference are different from the corresponding expressions in the case of independence. If the function $f(z, \theta)$ in (7.24) is differentiable with respect to the parameter θ for all values of z, then

$$X(\theta + \Delta\theta) - X(\theta) = f(Z, \theta + \Delta\theta) - f(Z, \theta)\big) = \frac{\partial f}{\partial \theta}(Z, \theta)\Delta\theta + o\big(\Delta\theta\big). \tag{7.25}$$

This means that, for small $\Delta\theta$, the variance of the one-sided finite difference $(X(\theta + \Delta\theta) - X(\theta))/\Delta\theta$ is approximately equal to the variance of the random variable $(\partial f/\partial\theta)(Z, \theta)$. The same holds for the variance of the two-sided difference. Assuming that the variance of $(\partial f/\partial\theta)(Z, \theta)$ is finite, it follows that the mean square errors for the one-sided and the two-sided difference respectively are of the form

$$\text{MSE}_1 = c_1\Delta\theta^2 + \frac{c_2}{n}\,, \qquad \text{MSE}_2 = c_3\Delta\theta^4 + \frac{c_4}{n}\,. \tag{7.26}$$

The bias is purely controlled by the step size $\Delta\theta$ and the variance by the number of samples n. So we can give $\Delta\theta$ an arbitrarily small value, and the convergence rate is $O(n^{-1/2})$ just as in the case of estimation of the value itself.

The analysis above is based on the assumption that the function $f(z, \theta)$ is differentiable with respect to θ for all z. In applications, this assumption is frequently not satisfied, due to the fact that payoff function often have kinks. For instance, suppose that we are interested in applying a Monte Carlo method to find the vega (sensitivity with respect to volatility) of a call option in the Black-Scholes model. The time-0 value of the option can be written as $e^{-rT}E[f(Z, \sigma)]$ where the function

[6] Time discretization with a fixed time step is taken for granted here.

f is given by

$$f(z,\sigma) = \max\bigl(S_0 \exp\bigl((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,z\bigr) - K,\, 0\bigr). \tag{7.27}$$

This function is not differentiable as a function of the parameter σ at the point z for which the two arguments of the max operator are equal to each other. In terms of the analysis above, this means that, for every choice of $\Delta\theta$, there will be outcomes of Z such that the expression (7.25) is not valid due to lack of differentiability of the function f as a function of the parameter θ on the interval from θ to $\theta + \Delta\theta$. However, the probability of these outcomes is approximately proportional to $\Delta\theta$, and the difference $X(\theta+\Delta\theta) - X(\theta)$ is in these cases still of order $O(\Delta\theta)$ due to the fact that the function f in (7.27) is Lipschitz[7] continuous. The conclusion (7.26) is therefore still valid.

The application of the bump-and-reprice method to the case of the computation of the vega of a call option in the BS model is shown in Code Example 7.1. While in this case simulation could be based on the known solution of the geometric Brownian motion, such a solution might not be available in other models, and so the code uses time stepping instead. To verify that the bump-and-reprice method indeed leads to the correct answer, a comparison is made with the analytic formula for the vega of a call option in the BS model, which is available from (3.86).

It was argued above that the "bump and reprice" method for computing the sensitivity of option values can be used with arbitrarily small step size $\Delta\theta$ when the payoff function is Lipschitz continuous. In cases in which the payoff function is discontinuous, however, the relations (7.26) cannot be maintained. For instance, if one wants to compute the vega of a digital option by the bump-and-reprice method, then one has to face the fact that the variance of $X(\theta + \Delta\theta) - X(\theta)$ is $O(\Delta\theta)$ rather than $O(\Delta\theta^2)$, since there is an $O(\Delta\theta)$ probability of Z falling into the region where the option payoff is 0 according to the parameter value θ and 1 according to the parameter value $\theta + \Delta\theta$, or vice versa. As a consequence, the variance of the difference quotient $(X(\theta + \Delta\theta) - X(\theta))/\Delta\theta$ is $O(\Delta\theta^{-1})$, so that the mean square error is of the form

$$\mathrm{MSE} = c_1\Delta\theta^2 + \frac{c_2}{n\Delta\theta}\,.$$

In this case, the mean square error will be badly affected if the step $\Delta\theta$ is made too small. Instead, to optimize the speed of convergence one should take $\Delta\theta \propto n^{-1/3}$, which leads to the convergence rate $n^{-1/3}$ for the RMSE. This can be improved to $n^{-2/5}$ by using the two-sided difference, but the fact remains that the step $\Delta\theta$ cannot be made arbitrarily small.

[7] Rudolf Lipschitz (1832–1903), German mathematician. A real-valued function is said to be Lipschitz continuous if there exists a constant c such that $|f(x) - f(y)| \le c|x - y|$ for all x and y.

```
r = 0.03; sigma = 0.20; S0 = 100; K = 100; T = 1;
h = 0.01; nsteps = 50; nsamples = 10^5; dt = T/nsteps;
S=S0; Sb = S0; % initialization
for i = 1:nsteps
    dW  = sqrt(dt)*randn(nsamples,1);
    dS  = r*S*dt + sigma*S.*dW;
    S   = S + dS;
    dSb = r*Sb*dt + (sigma+h)*Sb.*dW; % bumped version
    Sb  = Sb + dSb;
end
C  = exp(-r*T)*max(S-K,0);
Cb = exp(-r*T)*max(Sb-K,0);
vega_est = mean((Cb-C)/h);
vega_std = std((Cb-C)/h)/sqrt(nsamples);
disp(['vega bump&repr: ' num2str(vega_est) ...
      ' +/- ' num2str(1.96*vega_std)])
% also compute exact vega in BS model for comparison
d1      = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));
vega_ex = S0*normpdf(d1)*sqrt(T);
disp(['vega exact:     ' num2str(vega_ex)])
```

Code Example 7.1: "Bump and reprice" method for computation of sensitivity of the option price with respect to the volatility of the underlying asset. The size of the bump is indicated by `h`.

In fact, in the case of the digital option it can be seen directly that making $\Delta\theta$ very small is not a good idea. In a finite sample, there is (with probability 1) a positive minimum bump size that is required to move a sample outcome from 0 to 1 or vice versa. If the step $\Delta\theta$ is taken smaller than this minimum, then the difference between the option payoffs and their bumped versions is 0 in all sample points, so that the resulting estimate for the vega would be 0 as well.[8] This is not the correct answer.

In cases in which it is possible to make the step size $\Delta\theta$ arbitrarily small, the fact that this is so suggests that it actually may be feasible to take the limit. In other words, the derivative of $E[f(Z,\theta)]$ might computed on basis of the formula

$$\frac{d}{d\theta}E[f(Z,\theta)] = E\Big[\frac{\partial f}{\partial\theta}(Z,\theta)\Big]. \tag{7.28}$$

Looking at the expectation as an integral, this interchange of differentiation and expectation can be viewed as an application of the Leibniz rule for differentiation of integrals. The right hand side in (7.28) can be computed by Monte Carlo, if one has a way of obtaining samples from $(\partial f/\partial\theta)(Z,\theta)$. In applications, the variable of interest (such as a payoff) is often constructed as the result of applying a function to a stochastic variable, say $X_T(\theta)$, which itself is constructed on the basis of a

[8] Moreover, there would be no indication from the confidence interval that the answer is wrong, because the estimated standard deviation is 0 as well.

```
---
    DS = 0; % S0 does not depend on sigma
---
        dDS = r*DS*dt + S.*dW + sigma*DS.*dW; % diff dS wrt sigma
        DS  = DS + dDS;
---
    DC = exp(-r*T)*(S>K).*DS; % differentiate payoff wrt sigma
    vega_est = mean(DC);
    vega_std = std(DC)/sqrt(nsamples);
---
```

Code Example 7.2: Modification of code in Code Example 7.1 to implement the pathwise method for computation of sensitivity of the option price with respect to the volatility of the underlying asset. The variables `Sb` and `Cb` are no longer needed.

sequence of steps of the form

$$X_{t_{k+1}}(\theta) = X_{t_k}(\theta) + \mu_X(t, X_{t_k}(\theta), \theta)\,\Delta t + \sigma_X(t, X_{t_k}(\theta), \theta)\sqrt{\Delta t}\,Z_k \tag{7.29}$$

starting from an initial condition $X_0(\theta)$; the functions $\mu_X = \mu_X(t, x, \theta)$ and $\sigma_X = \sigma_X(t, x, \theta)$ are given, and the variables Z_k are drawn independently from the standard normal distribution. Taking partial derivative with respect to θ in the above, one finds (writing DX_t to indicate the derivative with respect to θ)

$$\begin{aligned} DX_{t_{k+1}}(\theta) = DX_{t_k}(\theta) &+ \left[\frac{\partial \mu_X}{\partial x}(t, X_{t_k}(\theta), \theta) DX_t(\theta) + \frac{\partial \mu_X}{\partial \theta}(t, X_{t_k}(\theta), \theta)\right]\Delta t \\ &+ \left[\frac{\partial \sigma_X}{\partial x}(t, X_{t_k}(\theta), \theta) DX_t(\theta) + \frac{\partial \sigma_X}{\partial \theta}(t, X_{t_k}(\theta), \theta)\right] Z_k. \end{aligned} \tag{7.30}$$

Starting the iteration from $DX_0 = (\partial X_0/\partial\theta)(\theta)$, one finds in this way simulated values of DX_T. The required partial derivatives of the model functions μ_X and σ_X should be available analytically; this is frequently the case in practice. If also the partial derivative of the payoff function with respect to the state variables can be computed, one finds in this way simulated values of the random variable $(\partial f/\partial\theta)(Z, \theta)$ which appears at the right hand side of (7.28). The estimate of the sensitivity with respect to θ is subsequently obtained by computing the average of these values. At the expense of having to compute partial derivatives of the model functions and the payoff function, this eliminates the bias that is inherent in finite differencing. The technique shown above can be described as "pathwise differentiation"; therefore, the method that is based on this is called the *pathwise method*. Code Example 7.2 shows an implementation of the pathwise method, in the form of modifications with respect to Code Example 7.1.

In cases in which the payoff function is discontinuous, the pathwise method is likely to fail. Indeed, in the case of the vega of a digital option, the Leibniz rule does

not apply since $(\partial f/\partial\sigma)(Z,\sigma) = 0$ with probability 1, so that $E[(\partial f/\partial\sigma)(Z,\sigma)] = 0$, whereas the partial derivative of $E[f(Z,\sigma)]$ with respect to σ is nonzero. This is in line with the earlier observation concerning what happens when the step size in the bump-and-reprice method, as applied to computing the sensitivity of a digital option, is made too small.

An alternative method for computing sensitivities is available in cases where the payoff depends on a random variable, say S_T, whose density function is known explicitly. In many cases, the density function is a smooth function of parameters of interest, which gives reason to be optimistic concerning the validity of the interchange of differentiation and integration in the expressions below, even when the payoff function F is not smooth:

$$\begin{aligned}\frac{d}{d\theta}\,E[X(\theta)] &= \frac{d}{d\theta}\int F(s)\,g(s,\theta)\,ds = \int F(s)\,\frac{\partial g}{\partial\theta}(s,\theta)\,ds \\ &= \int F(s)\,\frac{(\partial g/\partial\theta)(s,\theta)}{g(s,\theta)}\,g(s,\theta)\,ds = E\left[F(S_T)\,\frac{(\partial g/\partial\theta)(S_T,\theta)}{g(S,\theta)}\right] \\ &= E\left[F(S_T)\,\frac{\partial\,\log\,g}{\partial\theta}(S_T,\theta)\right]. \end{aligned} \tag{7.31}$$

The partial derivative of the log density with respect to the parameter of interest is called the *score function* in statistics. If this can be computed, then (7.31) offers a possible way of computing the sensitivity by Monte Carlo sampling. This method is called the *likelihood ratio method* or the *score function method*.

As an example, consider (again) the Black-Scholes model. The distribution of S_T is given by

$$S_T = S_0 \exp\left((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,Z\right), \quad Z \sim N(0,1). \tag{7.32}$$

The density (say, $\psi(s)$) of S_T can be found by computing $P(S_T \le s)$ for given $s \in \mathbb{R}$ and differentiating with respect to s:

$$\psi(s) = \frac{1}{s\sigma\sqrt{T}}\,\phi\left(\frac{\log(s/S_0) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right).$$

Take for instance S_0 as the parameter of interest. We have

$$\log\,\psi(s,S_0) = -\frac{\left(\log(s/S_0) - (r - \frac{1}{2}\sigma^2)T\right)^2}{2\sigma^2 T} + \cdots$$

where the dots indicate terms that do not depend on S_0. The score function is

$$\frac{\partial}{\partial S_0} \log \psi(s, S_0) = \frac{\log(s/S_0) - (r - \frac{1}{2}\sigma^2)T}{\sigma^2 T\, S_0}\,.$$

The delta of an option with payoff $F(S_T)$ can now be computed using

$$\frac{\partial}{\partial S_0} E[F(S_T)] = E\left[F(S_T)\,\frac{\log(S_T/S_0) - (r - \frac{1}{2}\sigma^2)T}{\sigma^2 T\, S_0}\right] = E\left[F(S_T)\,\frac{Z}{\sigma\sqrt{T}\, S_0}\right]$$

where Z is as in (7.32).

The discussion above has concentrated on the calculation of first-order derivatives. There is also interest in second-order derivatives, however, such as gamma (the second derivative of the option price with respect to the underlying). The typical finite-difference estimator is

$$\frac{\bar{X}_n(\theta + \Delta\theta) - 2\bar{X}_n(\theta) + \bar{X}_n(\theta - \Delta\theta)}{\Delta\theta^2}.$$

The bias is $O(\Delta\theta^2)$ as indicated by (6.7). In the same way as discussed in the case of first-order derivatives, the variance of the second-order difference quotient tends to a finite limit when $X(\theta) = f(Z, \theta)$ and f is sufficiently smooth. Since we are dealing with second derivatives here, the smoothness requirements on f are more strict and often not met in practice. In case the variance of the second-order difference quotient is $O(\Delta\theta)$, we obtain

$$\mathrm{MSE} = c_1 \Delta\theta^4 + \frac{c_2}{n\Delta\theta}\,.$$

To optimize the convergence speed, one should take $\Delta\theta \propto n^{-1/5}$; the RMSE then converges at rate $O\big(n^{-2/5}\big)$. In the likelihood ratio method, the smoothness of the density function as a function of its parameters may make it possible to differentiate twice under the integral sign. As noted above, the applicability of this method depends on the availability of the density of the underlying at the time of expiry in explicit form.

7.4 Least-squares Monte Carlo

In applications of the Monte Carlo method within a continuous-time setting, the typical situation is that the value of the variable of interest at a given future time T can be computed as a function of the state variables at time T. Values of the state variables at time T can be simulated by means of a time-stepping procedure on the basis of the stochastic differential equations which are specified in the model that is being used. For instance, to price a European option within the context of

a stochastic volatility model, one would use the model equations to generate joint scenarios for the price of the underlying and the volatility to produce a large sample of possible values of the underlying at the time of expiry T, and then the given option payoff function would be used to compute corresponding option payoffs; the final estimate is subsequently obtained by averaging and discounting. The step from values of the state variables to option payoffs is made by means of the payoff function, which is typically given in an analytic form such as $\max(S_T - K, 0)$ for a call option. However, in practice situations frequently arise in which the values of the variable of interest cannot be computed easily from the corresponding values of the state variables, but instead must themselves be obtained by a Monte Carlo procedure. We then speak of *nested* Monte Carlo. Here are two examples.

Example 7.4.1 To meet regulatory requirements, an insurance company wants to compute the 99.5% quantile of the distribution of losses on its portfolio on a one-year horizon. Part of the portfolio consists of life insurance policies which include profit sharing optionalities that are too complicated to be valued analytically. The company can compute an approximation of the required quantile by a two-step procedure, as follows. First, generate a large number of scenarios under the real-world measure to find a set of possible values for relevant state variables one year from now. Then, for each of the possible values of the state vector that have been generated in the first step, start a new Monte Carlo simulation (under a suitable risk-adjusted measure) to determine the corresponding value of the products in the portfolio.

Example 7.4.2 A Bermudan put option expires in two years and has three early exercise opportunities, at six months, one year, and eighteen months from now. These points in time will be denoted by t_1, t_2, and t_3, and the final time of expiry will be denoted by T. To value the option, we work backwards in time. At the time of expiry, if the option has not been exercised before, its value is given by the standard formula $C_T = \max(K - S_T, 0)$, where C_T represents the option value at time T, S_T is the value of the underlying at time T, and K is the strike. At time t_3, if the option has not been exercised before, we must make a decision whether or not to exercise. The option should be exercised if the value of immediate exercise, which is $\max(K - S_{t_3})$, is larger than the value of the option when it is not exercised (the "continuation value"). At time t_3, the continuation value is equal to the value of a standard European put option that matures at time T, since there are no early exercise opportunities anymore after time t_3. Under the assumptions that the underlying follows a geometric Brownian motion and that interest rates are constant, the continuation value is therefore given by the Black-Scholes formula. Since that

value is always positive, we can write

$$C_{t_3} = \max\big(K - S_{t_3}, CV_3(S_{t_3})\big) \tag{7.33}$$

where CV_3 denotes the continuation value at time t_3, which depends on S_{t_3}. Likewise we have

$$C_{t_2} = \max\big(K - S_{t_2}, CV_2(S_{t_2})\big)$$

where $CV_2(S_{t_2})$ is the value of the option that matures at time t_3 with payoff given by the formula (7.33). Since the payoff formula is quite complicated, there is no explicit formula; instead, a Monte Carlo method can be used to compute an approximation to the continuation value at time t_2 for each given value of S_{t_2}. The Monte Carlo method can be used as well to compute the continuation value at time t_1; however, for each scenario starting at time t_1, a new set of scenarios would need to be started at time t_2 in order to find the payoff value at time t_1 by the standard MC method. Continuing in this way, application of the regular MC method at to determine the contract's value at the time of initiation would mean to create a large number of scenarios from the given initial value S_0 which would each branch into a new set of scenarios at time t_1, followed by another branching at time t_2. If for instance 1000 scenarios are viewed as necessary to get reasonable accurate Monte Carlo values, then the total number of scenarios generated in this way would be 10^9. If there would be an exercise opportunity every month instead of every six months, as is quite standard, then the number of scenarios would swell to 10^{69}, rendering the proposed computational method totally infeasible.

The computational problems that arise for standard Monte Carlo in the two examples are due to *nesting*. In situations as in the second example, in which there is only one state variable, an effective solution to the problem is provided by the so called *finite difference method* of Chapter 6. However, this method quickly becomes more involved when the number of state variables increases, as would happen for instance in models with variable interest rates. A popular way to handle nested problems involving several state variables is the *Least-Squares Monte Carlo* (LSMC) method. The basic idea of the method can be described as follows.

Consider the problem of finding a conditional expectation $E[f(X_1) \mid X_0 = x]$, where f is a given continuous function, and X_1 is a random vector that is generated from X_0 for instance by means of a discretized stochastic differential equation, so that samples from the distribution of X_1 can be generated when X_0 is given. The conditional expectation depends on the input vector x, so that we are actually looking for a function of x. One way to find an approximation to that function is to construct a collection of grid points $x_1, \ldots, x_N$ in the vector space in which X_0 takes its values, and to compute for each of these points an approximation $\hat{y}_j$ to the

corresponding value of the conditional expectation $y_j = E[f(X_1) \mid X_0 = x_j]$. An explicit approximate expression for the conditional expectation as a function of the vector variable x can be obtained in the form

$$E[f(X_1) \mid X_0 = x] \approx \sum_{k=1}^{M} w_k F_k(x)$$

where the functions $F_1(x)$, ..., $F_M(x)$ are pre-selected *basis functions*, and where the weights $w_1, \ldots, w_M$ are found by a regression procedure:

$$\text{minimize} \quad \sum_{j=1}^{N} \|\hat{y}_j - \textstyle\sum_{k=1}^{M} w_k F(x_j)\|^2.$$

Whether a good approximation is obtained in this way depends on the choice of the grid points $x_1, \ldots, x_N$ as well as on the choice of the basis functions $F_1(x), \ldots, F_M(x)$. Also, there needs to be a proper balance between the number N of grid points, the number M of basis functions, and the number of Monte Carlo runs that are used at each grid point x_j to determine the approximate value $\hat{y}_j$. The grid points should provide a representative sampling of the relevant domain; a random sample (i.e. Monte Carlo) may be good enough in some cases, but in other cases a hand-picked set may be preferred. Popular choices for basis functions include multivariate polynomials and exponentially weighted versions of these. If some rough information is available about the shape of the conditional expectation function, this may be used to guide the choice of the basis functions.

The term "least-squares Monte Carlo" (LSMC) comes from the fact that the function that approximates the conditional expectation is obtained from a least-squares optimization problem. The idea behind the method is that, by restricting the solution to be a linear combination of given basis functions, it is possible to tolerate a fairly low level of accuracy in the approximate values $\hat{y}_j$, since neighboring points will correct each other. This suggests that a fairly good approximation of the conditional expectation may be obtained, even when the number of Monte Carlo runs at each individual grid point is not high.

In situations as in Example 7.4.2, the computation of conditional expectations is applied in a number of successive iteration steps. One then needs to watch out for possible build-up of errors. In the specific case of the example, the approximate option value at a given time in a given state is computed as the maximum of two quantities, namely the value of immediate exercise and the continuation value. While the former is computed precisely, the latter is computed with an error. When the two values are close to each other (i.e. near the exercise boundary), it may happen that the exercise value is erroneously taken to be the largest because the continuation value is underestimated, or vice versa the exercise value is taken to be the smallest

only because the continuation value is overestimated. While these errors are about equally likely to occur, the *effects* of the two errors are not symmetric. In the case of an overestimate, the full amount of the error will be transferred to the computed option value. When the continuation value is underestimated, however, the error is truncated, since in this case the option value is taken to be the value of immediate exercise, not the continuation value. Therefore there is an upward effect, which after a number of iteration steps may give rise to a substantial error. The issue is in particular important when there are many stages in which the maximum of the value of immediate exercise and the continuation value needs to be computed, such as when an American option is approximated by a Bermudan option with many early-exercise opportunities.

The upward effect can be suppressed by looking at the computational procedure as a way of computing an exercise strategy, rather than as a way of computing an approximate option value directly. Given an exercise strategy (that is, a specification, for each exercise opportunity date, of the *exercise region*, i.e., the set of states at which the option will be exercised), one can use standard Monte Carlo in order to compute the option value that is obtained from this exercise strategy. This value is a *lower* bound for the true Bermudan option value, because the chosen exercise strategy may not be the optimal one.

A basic version of the LSMC method for valuation of a Bermudan put option is shown in Code Example (7.3). Five exercise opportunity dates are assumed. In the case of the standard Black-Scholes model that is used in the example, LSMC is in fact not the best method for valuation of a Bermudan put, so the BS model is used in the example just for purposes of illustration. The basis function used in the example (1, S, S^2, and S^3) are also chosen just for purposes of illustration. The method starts by generating a set of scenarios starting from the initial value of the underlying S_0. Simulation takes place under the risk-neutral measure. The points visited by these scenarios at the exercise opportunity dates are stored in a matrix. For each exercise opportunity date, one obtains in this way a (random) grid. Because the grid points are obtained from scenarios, there is for each grid point j at time t_k a corresponding grid point j at time t_{k+1}, namely the one that is obtained from the same scenario.[9] After the grid points have been created in a forward pass, the LSMC method proceeds to a backward pass. As a preliminary estimate of the continuation value in grid point j at time t_k, the value of the option is taken in grid point j at time t_{k+1}, discounted from time t_{k+1} to time t_k. This can be viewed as a Monte Carlo estimate with sample size 1. Regression on a prespecified set of

[9] Another effect of generating grid points at the successive exercise opportunity dates from scenarios is that the grid is more concentrated around the initial value S_0 for low values of the discrete time index k, and spreads out more widely later; this is reminiscent of tree methods as discussed in Section 6.7.

```
T = 1; sigma = 0.2; S0 = 100; K = 100; r = 0.04;
EV = @(x) max(K-x,0); % value of immediate exercise
dt = 0.2; N = T/dt; M = 1000; Sm = zeros(M,N);
BF = @(x) [ones(size(x)) x x.^2 x.^3]; % basis functions
% --- forward pass ---
S = S0; % initialization
for k = 1:N
    Z = randn(M,1);
    S = S.*exp((r-0.5*sigma^2)*dt+sigma*sqrt(dt)*Z); % or Euler
    Sm(:,k) = S; % store the sample points
end
% --- backward pass ---
V = EV(S); % initialization
for k = N-1:-1:1       % counting down
    C  = exp(-r*dt)*V; % simple estimate of continuation value
    S  = Sm(:,k);      % sample points at stage k
    w  = BF(S)\C;      % regress C on BF(S)
    CV = @(x) BF(x)*w; % continuation value as a function
    V  = max(EV(S),CV(S)); % estimated option values
end
% --- final step ---
C0 = exp(-r*dt)*mean(V);
Ce = max(EV(S0),C0);
disp(['LSMC estimate: ' num2str(Ce)])
```

Code Example 7.3: First version of LSMC method for the valuation of a Bermudan put option. The sample points are generated from the analytic solution of the geometric Brownian motion, but this might also have been done by an Euler method. The chosen basis functions are 1, x, x^2, and x^3.

basis functions is subsequently used to allow the values at neighboring grid points to correct each other. The actual estimate of the continuation value in grid point j at time t_k is then obtained as the value at this grid point of the interpolation function obtained from the regression. The estimated option value is obtained by taking the maximum of the estimated continuation value and the value of immediate exercise. The backward pass ends at time t_1. Since there are no exercise opportunities between time 0 and time t_1, it is straightforward to compute the estimated option value at time 0 from the extimated option values at time t_1.

This basic version of the LSMC method is sensitive to the upward drift that has been discussed above. To counteract the effect, one can use the fact that the backward pass not only delivers estimated option values, but also estimated exercise regions. In the course of the procedure, for each time index k, the grid points at time t_k have been labeled as either "exercise" or "do not exercise". An exercise region could be constructed from this; for instance, the exercise region at time t_k could be defined as the set of all points in the state space for which the most nearby grid point is an exercise point. An unbiased estimate of the value of the option with this exercise region, which is a lower bound for the Bermudan option value, could

```
---
    Tx = N*ones(M,1); Sx = S; % initialize exercise data
---
       jx = EV(S) > CV(S); % exercise indices
       Tx(jx) = k; Sx(jx) = Sm(jx,k); % update exercise data
---
    C0 = mean(exp(-Tx*r*dt).*EV(Sx)); % discount wrt exercise time
---
```

Code Example 7.4: Modification of Code Example 7.3 in which the approximate option value at time 0 is computed from an exercise strategy. After completion of the backward pass, the j-th entry of the vector `Tx` contains the index of the exercise time in scenario j, and the j-th entry of `Sx` contains the value of the underlying at the exercise time. Therefore the j-th entry of `EV(Sx)` is the option payoff at the time of exercise in the j-th scenario. As an alternative to constructing `Sx` by initialization and updating, it is also possible to recover `Sx` after completion of the backward pass from the vector `Tx` and the forward pass data in the matrix `Sm` by the command `Sx = Sm(sub2ind(size(Sm),(1:M)',Tx))`.

then obtained by generating a fresh set of Monte Carlo scenarios. It is attractive however to use the set from the original forward pass of the method, since these scenarios pass through points that have already been labeled as exercise points or non-exercise points.[10] This method was proposed by Jacques Carrière in 1996.[11] An implementation is shown in Code Example 7.4, in the form of modifications with respect to the first version of the LSMC method. Two new vectors are introduced which keep track, for each scenario, at which time exercise takes place within this scenario, and what the value of the underlying is at that time. These vectors are initialized before the start of the backward pass and are updated within the loop. The continuation value at time 0 is then computed as the expected result[12] of the exercise strategy, with discounting that takes the time into account at which exercise takes place.

The idea of computing an approximate option value from an exercise strategy can also be applied *within* the recursion of the backward pass, rather than only at the end of it. This version was proposed by Francis Longstaff and Eduardo Schwartz in 2001.[13] The required modifications with respect to Code Example 7.3 are shown in Code Example 7.5. The expected result of the exercise strategy at a given sample point in scenario j is computed only from that scenario (it can be considered as a Monte Carlo estimate with sample size 1), so no averages need to

[10] The unbiasedness of the resulting estimate may be compromised in this way. Since the exercise strategy is tested on the same scenarios that have been used for its design, the results are likely to be too optimistic; in other words, an upward bias is expected.

[11] J.F. Carrière, "Valuation of the early-exercise price for options using simulations and nonparametric regression", *Insurance: Mathematics and Economics* 19 (1996), 19–30.

[12] Expectation is taken with respect to the risk-neutral measure here, since the scenario set in the forward pass is generated under this measure.

[13] F.A. Longstaff and E.S. Schwartz, "Valuing American options by simulation: A simple least-squares approach", *Review of Financial Studies* 14 (2001), 113–147.

```
---
    Tx = N*ones(M,1); Sx = S; % initialize exercise data
---
       jx = EV(S) > CV(S); % exercise indices
       Tx(jx) = k; Sx(jx) = Sm(jx,k); % update exercise data
       V = exp(-(Tx-k)*r*dt).*EV(Sx); % estimated option values
---
```

Code Example 7.5: Alternative modification of Code Example 7.3. Approximate option values are computed from an exercise strategy already within the loop, rather than only after completion of the backward pass.

```
---
    jitm = EV(S) > 0; % "in the money" indices
    w    = BF(S(jitm))\C(jitm);
---
```

Code Example 7.6: Additional modification of Code Example 7.3: regression only on data points where the value of immediate exercise is positive.

be computed within the loop. The final step is the same as in Code Example 7.3. A further modification that was also proposed by Longstaff and Schwartz is to carry out the regression step only taking into account sample points in which the option is "in the money", i.e. the value of immediate exercise is positive. The corresponding modification is shown in Code Example 7.6.

The LSMC method requires the choice of basis functions which are used to approximate functions expressing conditional expectations. Even when convergence can be proved as the number of basis functions tends to infinity, the accuracy of the method when a limited set of basis functions is used may depend strongly on whether these functions are chosen appropriately. In a given application it may be possible to make a reasoned choice on the basis of a more or less precise idea of the shape of the conditional expectations. The dependence of the LSMC method on basis functions also means that the method is not as insensitive to state space dimension as one might hope for in a Monte Carlo method. For instance, the number of independent polynomials up to the second degree is 10 if the number of variables is three, and is 21 in five variables. Nevertheless, the method remains understandably popular in cases where no suitable alternative is available.

7.5 Exercises

1. In this exercise we use a Monte Carlo method for computing the price of a European call option according to the standard Black-Scholes model, using Euler time stepping and using the bond as a numéraire. Assume the following data: interest rate $r = 0.04$, volatility $\sigma = 0.2$, period $T = 1$, current asset price $S_0 = 100$, exercise price $K = 110$. With these parameters, generate 10^4 simulation runs of

the Black-Scholes model. Since we want to determine expectations with respect to the equivalent martingale measure $\mathbb{Q}_B$ that corresponds to taking the bond as a numéraire, the simulations should be done under $\mathbb{Q}_B$ (cf. (3.81)). Take time steps of size 0.02. For each run, compute $e^{-rT}\max(S_T - K, 0)$. Compute the average value and the variance of the results; from these, obtain a point estimate of the option value and an approximate confidence interval. Compare the obtained estimate to the exact value (3.50).

2. In this exercise we use the Monte Carlo method to compute the value of an Asian option in the standard Black-Scholes model. The value of the option at the maturity date is defined by $\max(A - K, 0)$ where A is the arithmetic average of the price of the underlying over the lifetime of the option, and K is the strike. We will use the following parameter values: riskless rate of interest $r = 0.02$, volatility $\sigma = 0.2$, time of maturity $T = 1$, current value of the underlying $S_0 = 100$, strike $K = 100$.

a. Write a program that generates, for any n, a collection of n independent trajectories of the price of the underlying and the corresponding averages. Use 100 time steps and assume that the average price of the underlying over the lifetime of the option can be replaced by the average of the values at the discrete time points. Since we want to determine expectations with respect to the equivalent martingale measure $\mathbb{Q}_B$ that corresponds to taking the bond as a numéraire, the simulations should be done under $\mathbb{Q}_B$ (cf. (3.81)).

b. Using the script of part a. with $n = 10^5$, compute a Monte Carlo estimate of the value of the Asian option. Is the Asian option more or less expensive than its European counterpart? Determine an approximate confidence interval corresponding to your Monte Carlo estimate. How large should n be taken approximately to reduce the width of the confidence interval to 0.01? Ignore the time discretization error.

c. Using 1000 simulations, produce a scatter plot of the European option payoff at expiry versus the Asian option payoff at expiry. Determine an approximate value of the correlation coefficient of the two variables. How much reduction of variance do you expect to obtain from using the European option payoff as a control variate?

d. Recompute the Monte Carlo estimate, using $n = 10^5$ as before, but this time with the European option payoff as a control variate. Determine an approximate confidence interval for your estimate and compare the length of this interval to the one that you got from the "raw" Monte Carlo method.

e. Now, instead of using the European option as a single control variate, use the following four control variates: $W_{T/2}$, $W_{T/2}^2$, W_T and W_T^2. Estimate optimal values of the coefficients α_i from a regression based on a pilot sample of size 10^3. Recompute the Monte Carlo estimate and compare the improvement in the length of the

confidence interval that you get in this way to the improvement obtained in part d.

f. Add the European call option to the four control variates of part e. How much improvement does this bring?

g. Instead of determining estimates for the optimal coefficients α_i by regression as in part e., it is also possible to use (7.18) where the matrix Σ_{XX} can be computed analytically, although the vector Σ_{XY} still needs to be determined on the basis of a sample. Find the variance-covariance matrix for the four control variates of part d., and compute the coefficients α_i based on this and on a pilot sample of size 10^3 to compute the vector Σ_{XY}. Recompute the Monte Carlo estimate using the new coefficients. Do you see improvement with respect to part e.? Explain your findings.

3. Let $X_1, X_2, \ldots, X_k$ be a collection of random variables whose expectations are to be estimated on the basis of samples drawn from their joint distribution. Let Y be a linear combination of the variables X_j with coefficients c_j, i.e. $Y = \sum_{j=1}^k c_j X_j$, and suppose that EY is known analytically, so that Y can be used as a control variate for each of the X_i's. The control variate estimate for EX_j, based on a sample of size n from the joint distribution of $X_1, \ldots, X_k$, is

$$\hat{m}_j = \frac{1}{n}\sum_{i=1}^n x_j^i - \alpha_j\Big(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^k c_j x_j^i - EY\Big) \tag{7.34}$$

where x_j^i is the outcome for X_j in the i-th trial. Suppose that the coefficients α_i are determined by (7.12) where the variance and the covariance are replaced by their sample equivalents (which may be taken from a pilot sample).

a. Let the outcomes of the random variables X_j $(j = 1, \ldots, k)$ in the i-th trial be denoted by x_j^i $(j = 1, \ldots, k)$. Define the modified versions

$$\hat{x}_j^i = x_j - \alpha_j\left(\sum_{j=1}^k c_j x_j - EY\right).$$

Prove that these modified versions satisfy $\sum_{j=1}^k c_j \hat{x}_j^i = EY$. Derive that the estimates defined in (7.34) satisfy the same relationship:

$$\sum_{j=1}^k c_j \hat{m}_j = EY. \tag{7.35}$$

b. The current value of assets in a trust fund is 100. The assets are invested; we assume that the value of the assets follows a geometric Brownian motion with volatility $\sigma = 15\%$. The riskless interest rate is 2% and does not change through

time. Three parties are involved with the trust fund. These parties are called Beneficiary, Sponsor, and Charity. At the end of each year, during the coming 10 years, a benefit will be paid to the Beneficiary. If the value of the fund's assets at the end of a given year is 105 or more, then the Beneficiary receives 4% of the assets; if the value is less than 105 but still at least 95, then the Beneficiary receives 2%; if the value of the fund's assets is less than 95, then the benefit for that year is canceled. At the end of the period of 10 years, after the last payment to the Beneficiary has been made, the asset value will be restored to 100 in the following way. If the end-of-period asset value is less than 100, the difference is supplied by the Sponsor. On the other hand, if the asset value is more than 100, then the part of the asset value that exceeds 100 goes to Charity. Using a basic Monte Carlo procedure with 10^5 samples, determine the time-0 value of the policy to the Beneficiary and to Charity, and the time-0 value of the contribution of the Sponsor. Report the results together with a 95% confidence interval.

c. Argue that the sum of the time-0 values to the Beneficiary and to Charity, together with the time-0 value of the assets at termination of the policy (i.e. after restoration), must be equal to the sum of the initial value of the assets and the time-0 value of the contribution of the Sponsor ("law of conservation of value"). Verify whether this equality is satisfied, within the limits of Monte Carlo accuracy, by the outcomes that you found in part b.

d. Part c. indicates that a control variate technique as described in part a. can be applied. Determine coefficients α_i on the basis of a pilot sample of size 10^4, and then repeat the computation of part b., this time using the control variate. Do you obtain a reduction of variance? Verify that the estimated values satisfy the relation of part c. up to machine epsilon.

4. Consider a European option with payoff function $f(S_T)$ within the Black-Scholes model. In this model, S_T can be written as a function of a standard normal random variable Z, so that the price of the option can be written in the form $E[F(Z)]$ where $Z \sim N(0,1)$ and $F(z)$ is a given function. If an approximation of F can be given in the form $F(z) \approx \sum_{i=1}^{n} c_i \phi_i(z)$ where the ϕ_i's are functions such that $E[\phi_i(Z)]$ is known, then the random variable $\sum_{i=1}^{n} c_i \phi_i(Z)$ should be an effective control variate.

a. (Compare Exc. 4.5.8.) Consider a call option, so that $f(S_T) = \max(S_T - K, 0)$. Write a script to approximate the corresponding function $F(z)$ by a linear combination of the functions $\phi_i(z)$ $(i = 1, 2, \ldots, N)$ defined, for odd values of N, by $\phi_i(z) = \cos\left(\frac{1}{2}(i-1)az\right)$ $(i = 1, 3, \ldots, N)$ and $\phi_i(z) = \sin\left(\frac{i}{2}az\right)$ $(i = 2, 4, \ldots, N-1)$. Choose the constant a such that the basis function $\phi_2(z) = \sin(az)$ is increasing across an interval that is wide enough to contain practically all sample values in a draw from the standard normal distribution of size 10^5. Determine the coefficients

c_i similarly to Exc. 4.5.8.

b. Take $N = 7$. Draw a plot showing both $F(z)$ and the approximating function $\tilde{F}(z) := \sum_{i=1}^{N} c_i \phi_i(z)$. Also generate a scatter plot showing outcomes of $F(Z)$ and $\tilde{F}(Z)$ in a draw of 10^5 samples from the standard normal distribution. On the basis of this draw, determine the control variate estimate and the corresponding confidence interval.

c. Experiment with the value of N and the number of grid points used to determine the coefficients c_i to see if you can get improvements. Do you get good results when you take high values for both parameters?

5. In this exercise we consider a put option with strike $K = 75$ in a Black-Scholes model with $r = 0.04$, $\sigma = 0.2$, $T = 1$, and $S_0 = 100$. With these parameter values, the put is far out of the money. Let us ignore the fact that the contract can be priced analytically and instead compute the value on the basis of a Monte Carlo procedure based on time stepping.

a. Apply a standard Monte Carlo procedure with 250 time steps and 10 000 replications. Compute a point estimate and an approximate confidence interval. Verify that the confidence interval contains the exact value obtained from the Black-Scholes formula.

b. The situation calls for an application of importance sampling based on a Girsanov transformation. Since there is a positive payoff only in asset price trajectories that go down substantially, it should be advantageous to replace the drift parameter in the asset price dynamics from the risk-neutral value $r = 0.04$ to a large negative value. With such a value, run again the Monte Carlo method, taking care to multiply the result of each simulation by the appropriate factor as indicated by the change of measure. Calculate a point estimate and compute an approximate confidence interval for the estimated option value. Do you observe improvement with respect to the method of the previous part? Try a few different values of the drift parameter to see which one works best.

6. Redo Exc. 1, but now using the stock price S_t as a numéraire rather than the bond; cf. (3.82). How does the confidence interval that you obtain in this way compare to the confidence interval that you found in Exc. 1?

7. **a**. Let $f(x)$ be a continuous function defined on $[0, 1]$ that is continuously differentiable on the open interval $(0, 1)$. Suppose that Z is a standard normal variable. Prove that the random variable $\Phi(Z)$ follows the uniform distribution on the interval $[0, 1]$, and that the following relation holds:

$$E[f'(\Phi(Z))] = f(1) - f(0). \tag{7.36}$$

b. In the particular case in which $f(x) = \sqrt{x}$, compute the expectation in (7.36) by means of the Monte Carlo method, using 10^4 trials. Find a point estimate and determine an estimated 95% confidence interval. Does the confidence interval contain the true value?

c. Repeat the experiment of part b., but now using $f(x) = x^{0.1}$. Does the computed confidence interval contain the true value? Repeat the calculation a few times. Is the size of the confidence interval approximately the same in each case?

d. Draw a plot in which you show the density of the normal distribution as well as the function $f'(\Phi(z))\phi(z)$, for $f(x) = x^{0.5}$ and for $f(x) = x^{0.1}$. Explain the difference in the results that you obtained in part b. and in part c.

8. Consider a standard Black-Scholes model with $r = 0.02$, $\sigma = 0.2$, and current asset value $S_0 = 100$. Suppose we want to price a put option that has strike 100 and one year to maturity, making use of the Monte Carlo method with Euler time stepping and variance reduction by the method of antithetic variables. We take the bond as a numéraire so that simulations will be done under the corresponding equivalent martingale measure.

a. Generate 1000 approximate trajectories of the underlying asset S with time step 0.01, using 100 independent draws $z_1, \ldots, z_{100}$ from the standard normal distribution for each trajectory. Compute the put payoffs that arise in these scenarios, and take the average to obtain a first Monte Carlo estimate. Also compute an approximate confidence interval.

b. For each trajectory generated in part a., also generate its antithetic counterpart, which is obtained by replacing the 100 draws $z_1, \ldots, z_{100}$ by $-z_1, \ldots, -z_{100}$. In this way you obtain again 1000 trajectories. Compute the corresponding option payoffs, and use these to find a second Monte Carlo estimate and a corresponding approximate confidence interval.

c. Now compute a Monte Carlo estimate and a confidence interval based on all of the 2000 trajectories that you have generated. Compare the confidence interval that you find in this way to the two confidence intervals that you have obtained before, and which were each based on 1000 trajectories. Is the reduction more than would be expected from the fact that twice as many samples are used? Explain on the basis of the theory in Subsection 7.2.3.

d. Repeat the steps above, assuming now that the strike of the put option is 80 rather than 100. Can you explain why the method of antithetic variables is less effective in this case?

9. The payoff of an Asian call option with two sample points at times $T/2$ and T is defined by $\max(\frac{1}{2}(S_{T/2} + S_T) - K, 0)$ where S_t is the price of the underlying and K

is the strike. Take $T = 1$, $K = 100$, $S0 = 100$, and assume the Black-Scholes model with interest rate $r = 0.02$ and volatility $\sigma = 0.2$.

a. Compute the price of the option at time 0 in the following three ways.

(i) (*Numerical integration method.*) The price of the option is given by

$$C = e^{-rT} E^{\mathbb{Q}} \max\left(\tfrac{1}{2}(S_{T/2} + S_T) - K,\, 0\right)$$

which may be written more explicitly as

$$C = e^{-rT} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x,y) g(x,y)\, dx\, dy$$

where

$$F(x,y) = \max\left(\tfrac{1}{2}[\, S_0(\exp(a+bx))(1+\exp(a+by))] - K,\, 0\right)$$

with

$$a = (r - \tfrac{1}{2}\sigma^2)T/2, \quad b = \sigma\sqrt{T/2}$$

and

$$g(x,y) = \tfrac{1}{2\pi} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2}.$$

It may not be attractive to evaluate the integral analytically; however a numerical approximation of the integral may be computed using the fact that in general a double integral of the form $\int\int f(x,y)\, dx\, dy$ can be approximated by a double sum $\sum\sum f(x_i, y_i)\Delta x_i \Delta y_i$. Make sure to use a grid that covers all of the area in which the integrand differs substantially from 0.

(ii) (*Finite-difference method.*) At time $T/2$, the Asian option becomes equivalent to a European call option, since one may write

$$\max\left(\tfrac{1}{2}(S_{T/2} + S_T) - K,\, 0\right) = \tfrac{1}{2}\max\left(S_T - (2K - S_{T/2}),\, 0\right).$$

So the price of the Asian option at time 0 may be viewed as the price of a European option which matures at time $T/2$ with a payoff that depends on $S_{T/2}$; specifically, the payoff is equal to one-half the value of European call option with initial asset price $S_{T/2}$, strike $2K - S_{T/2}$, and time to maturity $T/2$. The latter option may be valued by means of the standard Black-Scholes formula. Given this payoff, the option price at time 0 can be calculated by means of a finite-difference method.

(iii) (*Monte Carlo method.*) The Asian option can also be priced on the basis of a Monte Carlo method. Generate a large number (for instance 10^5) of trajectories under the risk-neutral measure using an Euler method, compute the value of the Asian option for each trajectory, take the average and discount to 0 to find an estimate of the option price. Also compute a 95% confidence interval for the estimate.

b. Comment on the applicability of the three methods when the price of the underlying asset is not a geometric Brownian motion, but instead is described by a stochastic differential equation for which no explicit solution is available. Also comment on the feasibility of each of the methods when the underlying asset dynamics is geometric Brownian motion but the number of sampling times is, say, 10 or more.

This page intentionally left blank

Appendix A

Notes

Chapter 1. Section 1.1 is based on an article that I wrote for *Nekst*, the magazine of the student association Asset Econometrics[1] at Tilburg University. The article was published in the June 2011 issue of this journal. The main sources for the article are the books by Bernstein and Mehrling that are mentioned in the Preface.

Chapter 2. The non-stochastic introduction to Itô calculus in Section 2.4.2 is motivated by work of Hans Föllmer ("Calcul d'Itô sans probabilités", *Séminaire de probabilités (Strasbourg)*, tome 15 (1981), pp. 143–150).

Chapter 3. The information in footnote 12 is from "Arbitrage: Historical perspectives" by Geoffrey Poitras, in *Encyclopedia of Quantitative Finance* (Rama Cont, ed.), Wiley, 2010. The "variational" form of the Black-Scholes formula (i.e. the form in which it is written as the maximum of another function), as given in Exc. 7, can be found in Jean-Paul Décamps and Jean-Charles Rochet, "A variational approach for pricing options and corporate bonds", *Economic Theory* 9 (1997), 557–569.

Chapter 4. The solution method for the partial differential equation (4.27) that is shown in Section 4.2.1 follows the method described by Fourier in §364 of his book *Théorie analytique de la chaleur* (Analytical Theory of Heat) which appeared in 1822.[2] For the important step from (4.31) to (4.32), Fourier gives credit to Laplace.[3] The expression (4.34) can be made more general by replacing the sum by an integral. The function $c(\lambda)$ associated to a given function $\phi(y)$ by the equation

$$\phi(y) = \int_{-\infty}^{\infty} c(\lambda) e^{-\lambda y}\, d\lambda \tag{A.1}$$

[1] In this context, the term "Asset" stands for "Association of students studying economics in Tilburg".

[2] J. Fourier, *Théorie analytique de la chaleur*, Didot, Paris, 1822. Reprinted by Éditions Jacques Gabay, Paris, 1988.

[3] Pierre Simon, marquis de Laplace (1749–1827), French mathematician and astronomer.

(where the minus sign in the exponent is just a matter of convention) is called the two-sided *Laplace transform* of ϕ. The integral equation (A.1) was studied by Laplace in 1782.

For the development in Section 4.4.2, specifically the trick of differentiating (4.56), I have followed my own paper "Efficiency of institutional spending and investment rules", *Scandinavian Actuarial Journal* 2020(5), 454–476. The standard method for dealing with equations of the form (4.56) is to consider them as a special case of Kummer's[4] differential equation, which has a known solution in terms of confluent hypergeometric functions. For instance, this is the way it was done by Merton in his 1973 paper on option pricing ("Theory of rational option pricing", *Bell Journal of Economics and Management Science* 4, 141–183).

A general formula for option prices subject to stochastic interest rates was given by Merton in the same paper. The specific form taken by the formula in the case of a call option within the Vasicek model is given in a paper by Ramon Rabinovitch ("Pricing stock and bond options when the default-free rate is stochastic", *Journal of Financial and Quantitative Analysis* 24 (1989), 447–457). The derivation in Section 4.4.4 is different from the one in Rabinovitch's paper. A derivation similar to the one given here is in "A Black-Scholes like model with Vasicek interest rates" (working paper, Wirtschaftsuniversität Wien, 2007) by Zehra Eksi.

Exc. 8 has taken inspiration from the COS method that has been developed by Fang Fang and Kees Oosterlee. For much more information, see their paper "A novel pricing method for European options based on Fourier-cosine series expansions" (*SIAM Journal on Scientific Computing* 31 (2009), 826–848).

Chapter 5. At the end of Section 5.4, the fact is discussed that any interest model can be made to reproduce a currently observed term structure by adding a suitable deterministic function of time to the short rate. This was probably first observed by Philip Dybvig; see his contribution "Bond and bond option pricing based on the current term structure" to the book *Mathematics of Derivative Securities* (Cambridge University Press, 1997), edited by Michael Dempster and Stanley Pliska. Dybvig refers to the function $f(t)$ appearing in (5.31) as a "fudge factor" (p. 280 in the cited book).

Exercise 9 was inspired by a paper of Jens Christensen, Francis Diebold, and Glenn Rudebusch ("The affine arbitrage-free class of Nelson-Siegel term structure models", *Journal of Econometrics* 164 (2011), 4–20).

Chapter 7. For the discussion of the LSMC method, I have used a paper by Lars Stentoft ("Value function approximation or stopping time approximation: A comparison of two recent numerical methods for American option pricing using simulation and regression", *Journal of Computational Finance* 18 (2014), 56–120).

[4]Ernst Eduard Kummer (1810–1893), German mathematician.

Appendix B

Hints and answers for selected exercises

B.1 Exercises from Chapter 1

1. For the total variation part, consider partitions of the form $0 < 1/((N - \frac{1}{2})\pi) < 1/((N - \frac{3}{2})\pi) < \cdots < 1/(\frac{1}{2}\pi) < 1$.

2. **a**. Let f be a continuous function defined on a closed and bounded interval $[a, b]$. For a proof by contradiction, suppose that f is not uniformly continuous. Then there exists $\varepsilon > 0$ such that for all $k \in \mathbb{N}$ there exist x_k and y_k in $[a, b]$ such that $\lim_{k\to\infty} |x_k - y_k| = 0$, but $|f(x_k) - f(y_k)| > \varepsilon$. Since the sequence $(x_k)_{k=1,2,\ldots}$ is an infinite sequence within the closed and bounded interval $[a, b]$, there must be a sequence of increasing indices $k_1 < k_2 < \cdots$ such that the subsequence $(x_{k_j})_{j=1,2,\ldots}$ converges to a limit, say $\bar{x}$, in the interval $[a, b]$. Likewise, the sequence $(y_{k_j})_{j=1,2,\ldots}$ is an infinite sequence in $[a, b]$, so it must have a subsequence that converges to a limit, say $\bar{y}$, in $[a, b]$. The corresponding subsequence of $(x_{k_j})_{j=1,2,\ldots}$ converges to $\bar{x}$. To avoid further proliferation of subindices, conclude from this reasoning that without loss of generality it can be assumed that the original sequences (x_k) and (y_k) converge to $\bar{x}$ and $\bar{y}$ respectively. Because f is continuous, the sequences $(f(x_k))$ and $(f(y_k))$ converge to $f(\bar{x})$ and $f(\bar{y})$ respectively. From the assumption that $|f(x_k) - f(y_k)| > \varepsilon$ for all k, it follows that $|f(\bar{x}) - f(\bar{y})| \geq \varepsilon$. On the other hand, since $\lim_{k\to\infty} |x_k - y_k| = 0$, the two limit points $\bar{x}$ and $\bar{y}$ must be the same. We have a contradiction. (An alternative and shorter proof can be given by using the notion of compactness.)

b. Let M denote the total variation of the function g on the interval $[a, b]$. Choose $\varepsilon > 0$. By the fact that g is uniformly continuous on $[a, b]$, as shown in part a., there exists $\delta > 0$ such that $|g(x) - g(y)| < \varepsilon/M$ for all x and y in $[a, b]$ such that

$|x-y|<\delta$. For every partition Π with $|\Pi|<\delta$, we then have

$$\sum_{j=0}^{n}\big(g(x_{j+1})-g(x_j)\big)^2<\frac{\varepsilon}{M}\sum_{j=0}^{n}|g(x_{j+1})-g(x_j)|\le\varepsilon.$$

B.2 Exercises from Chapter 2

4. For $s>t$, we can write

$$\begin{aligned}E_tW_s^3&=E_t\big[(W_t+(W_s-W_t))^3\big]=\\&=E_t\big[W_t^3+3W_t^2(W_s-W_t)+3W_t(W_s-W_t)^2+(W_s-W_t)^3\big]=\\&=W_t^3+3W_t(s-t).\end{aligned}$$

Therefore,

$$E_tX_s=E_t\big[\tfrac{1}{3}W_s^3-sW_s\big]=\tfrac{1}{3}W_t^3+W_t(s-t)-sW_t=\tfrac{1}{3}W_t^3-tW_t=X_t.$$

Alternatively, compute the differential of X_t:

$$d(\tfrac{1}{3}W_t^3-tW_t)=W_t^2\,dW_t+W_t\,dt-W_t\,dt-t\,dW_t=(W_t^2-t)\,dW_t.$$

From the telescope rule it now follows that we can write

$$X_t=X_0+\int_0^t(W_s^2-s)\,dW_s$$

which shows that the process X_t is a martingale.

5. Compute $d[X,X]_t$ and use the telescope rule.

6. **a**. From the Itô formula, one has

$$d(\cos W_t)=-\sin W_t\,dW_t-\tfrac{1}{2}\cos W_t\,dt.$$

The telescope rule implies that

$$\cos W_t-\cos W_0=-\int_0^t\sin W_s\,dW_s-\tfrac{1}{2}\int_0^t\cos W_s\,ds.$$

Taking expectations, one finds

$$E[\cos W_t]-1=-\tfrac{1}{2}\int_0^t E[\cos W_s]\,ds.$$

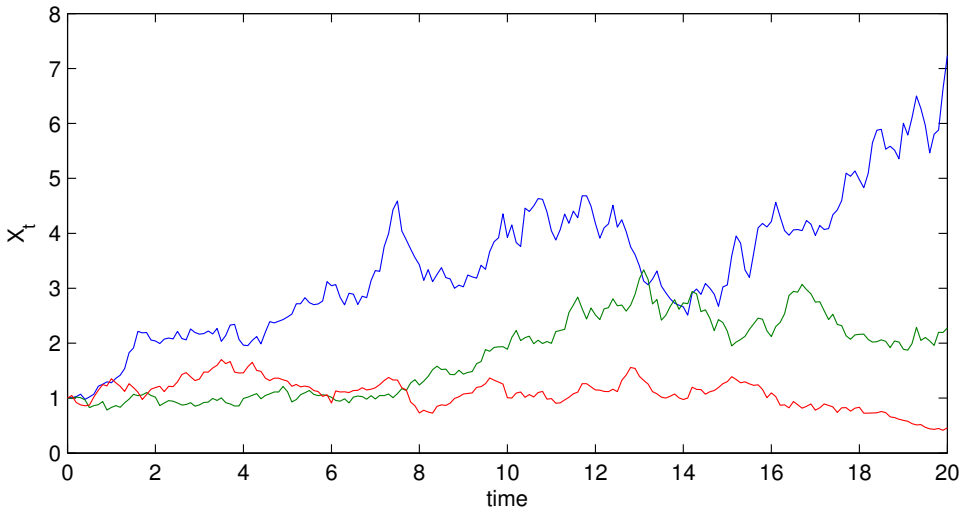

Figure B.1: Graphical output for Exc. 2.8.8a.

This shows that the function $x(t) := E[\cos W_t]$ satisfies the differential equation

$$\frac{dx}{dt}(t) = -\tfrac{1}{2}x(t).$$

Since $x(0) = 1$, it follows that $x(t) = \exp(-\frac{1}{2}t)$. Because the random variable aZ with $Z \sim N(0,1)$ has the same distribution as W_t with $t = a^2$, one obtains (2.99).

7. For part a., compute the cumulative distribution function of Z. For part b., note (for instance) that $P(X + Z = 0) = P(Y = -1) = \frac{1}{2}$. If X and Z would be jointly normally distributed, then $X + Z$ would also be normally distributed, and consequently the probability of the event $X + Z = 0$ would be zero.

8. **a**. The results are random, but qualitatively they should look as in Fig.,B.1.

b. Again the results are random, but less so than in the case of part a.; see Fig. B.2.

9. **b**. See Fig. B.3.

10. **b**. See Fig. B.4.

11. 1c, 2d, 3b, 4a.

12. 1d, 2a, 3b, 4c.

13. **a**. $dZ_t = 0$.

c. A script is shown in Code Example B.1; examples of graphical output are in Fig. B.5. Even with $\Delta t = 0.001$, the deviations due to the discretization error are still quite noticeable.

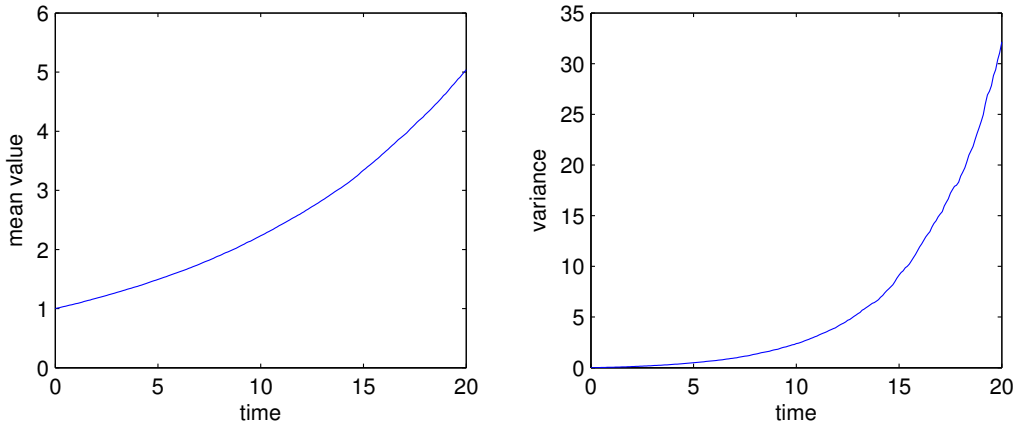

Figure B.2: Graphical output for Exc. 2.8.8b.

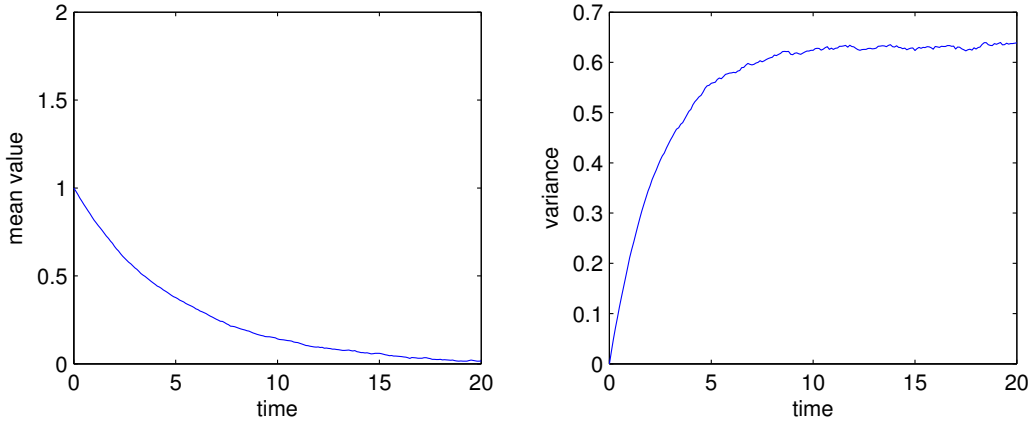

Figure B.3: Graphical output for Exc. 2.8.9b.

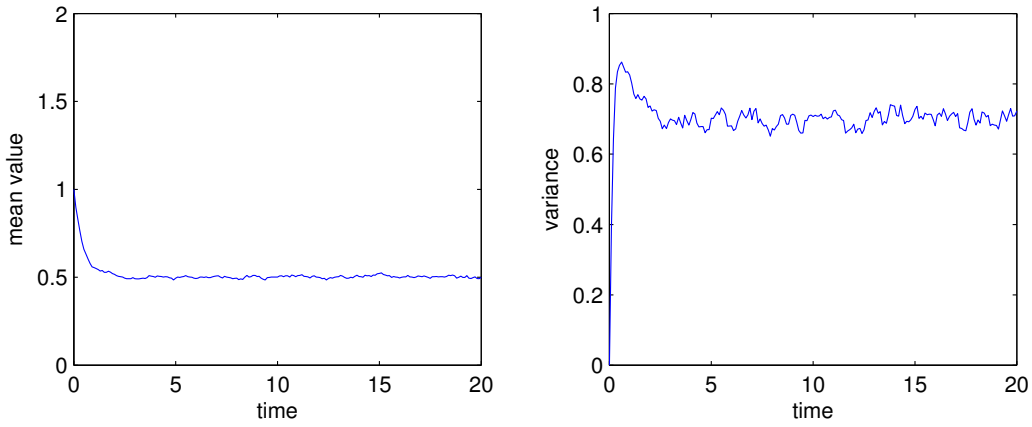

Figure B.4: Graphical output for Exc. 2.8.10b.

```
X0 = 1; Y0 = 0; dt = 0.001; T = 5; N = T/dt+1; % data
Xs = X0*ones(1,N); Ys = Y0*ones(1,N); % reservation of memory space
X = X0; Y = Y0; % initialization
for k = 1:N-1
    dW = sqrt(dt)*randn;
    dX = -0.5*X*dt - Y*dW; dY = -0.5*Y*dt + X*dW;
    X = X + dX; Y = Y + dY;
    Xs(k+1) = X; Ys(k+1) = Y;
end
plot(Xs,Ys); axis([-1.2 1.2 -1.2 1.2]); axis square
```

Code Example B.1: Code for Exc. 13.

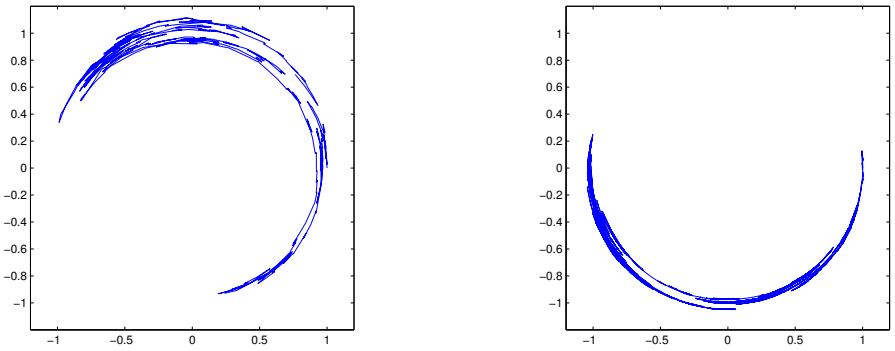

Figure B.5: Graphical output for Exc. 13. The left panel shows a simulation with $\Delta t = 0.01$, the right panel shows a simulation with $\Delta t = 0.001$.

15. **a**. From $d(tW_t) = W_t\,dt + t\,dW_t$ one finds by integration from 0 to 1:

$$1 \cdot W_1 - 0 \cdot W_0 = \int_0^1 W_t\,dt + \int_0^1 t\,dW_t$$

so that

$$X = \int_0^1 W_t\,dt = W_1 - \int_0^1 t\,dW_t = \int_0^1 dW_t - \int_0^1 t\,dW_t.$$

The two stochastic integrals with deterministic integrands on the right hand side are driven by the same Wiener process and are therefore jointly normally distributed. From (2.49) and (2.66) it follows that $EX = 0$ and

$$\mathrm{var}(X) = \int_0^1 dt - 2\int_0^1 t\,dt + \int_0^1 t^2\,dt = \tfrac{1}{3}.$$

b. Write $X_T = \int_0^T W_t\,dt$. The reasoning of part a. can be followed to conclude that $\mathrm{var}(X_T) = \frac{1}{3}T^3$. Alternatively, one can make use of Exercise 3 to derive that $X_T = T^{3/2}X_1$.

B.3 Exercises from Chapter 3

1. **b**. Suppose that condition (i) does not hold (i.e. there is no arbitrage). Then the subspace $\mathcal{V}$ defined by

$$\mathcal{V} = \left\{ \begin{bmatrix} \phi_0 S_u + \psi_0 B_u \\ \phi_0 S_d + \psi_0 B_d \end{bmatrix} \,\middle|\, \phi_0 S_0 + \psi_0 B_0 = 0 \right\}$$

intersects the nonnegative cone $\mathbb{R}^2_+$ only in 0. Since $\phi_0 S_0 + \psi_0 B_0 = 0$ holds if and only if $\psi_0 = -\phi_0 S_0/B_0$, the subspace $\mathcal{V}$ is generated by the vector $[S_u - B_u S_0/B_0 \;\; S_d - B_d S_0/B_0]^\top$. It follows from part a. that there exist positive constants y_1 and y_2 such that

$$y_1\Big(S_u - B_u \frac{S_0}{B_0}\Big) + y_2\Big(S_d - B_d \frac{S_0}{B_0}\Big) = 0.$$

Define $q_u = y_1 B_u/(y_1 B_u + y_2 B_d)$, $q_d = y_2 B_d/(y_1 B_u + y_2 B_d)$. Then q_u and q_d are both positive, $q_u + q_d = 1$, and $q_u S_u/B_u + q_d S_d/B_d = S_0/B_0$. Therefore, condition (ii) holds. Next, assume that condition (i) does hold, and that condition (ii) holds as well. It follows from (i) that there exists a number ϕ_0, necessarily nonzero, such that

$$\phi_0 \begin{bmatrix} S_u - B_u S_0/B_0 \\ S_d - B_d S_0/B_0 \end{bmatrix} \in \mathbb{R}^2_+ \setminus \{0\}.$$

Because both B_u and B_d are positive, this implies that also

$$\phi_0 \begin{bmatrix} S_u/Bu - S_0/B_0 \\ S_d/B_d - S_0/B_0 \end{bmatrix} \in \mathbb{R}^2_+ \setminus \{0\}.$$

On the other hand, condition (ii) implies that

$$q_u(S_u/B_u - S_0/B_0) + q_d(S_d/B_d - S_0/B_0) = 0$$

where both q_u and q_d are positive. From this it follows that $S_u/B_u - S_0/B_0$ and $S_d/B_d - S_0/B_0$ cannot be of the same sign, and so we have a contradiction.

2. **a**. The market is described in terms of one state variable (S_t), one driving Brownian motion, and two traded assets (S_t and C_t). We have

$$\begin{bmatrix} \sigma_Y & \pi_Y \end{bmatrix} = \begin{bmatrix} \sigma_S(t,S) & S \\ \sigma_C(t,S) & \pi_C(t,S) \end{bmatrix}$$

where Y_t is the vector of asset prices, and $\sigma_C = (\partial \pi_C/\partial S)\sigma_S$ by Itô's rule. The market is complete and arbitrage-free if and only if the above matrix is invertible

for all t and S, or in other words, if and only if

$$\pi_C(t,S) - S\,\frac{\partial \pi_C}{\partial S}(t,S) \neq 0 \quad \text{for all } t \text{ and } S$$

because the common factor $\sigma_S(t,S)$ is always positive. Since the function $\pi_C(t,S)/S$ is strictly increasing as a function of S, its partial derivative with respect to S is positive:

$$0 \;<\; \frac{\partial}{\partial S}\,\frac{\partial(\pi_C(t,S)/S)}{\partial S} = \frac{1}{S}\,\frac{\partial \pi_C(t,S)}{\partial S} - \frac{\pi_C(t,S)}{S^2}.$$

This implies the condition above (multiply by S^2).

b. The condition for the extended market to be arbitrage-free is that the equation

$$\begin{bmatrix} \mu_S \\ \mu_C \\ rB \end{bmatrix} = \begin{bmatrix} \sigma_S & S \\ \sigma_C & \pi_C \\ 0 & B \end{bmatrix} \begin{bmatrix} \lambda \\ \tilde{r} \end{bmatrix}$$

(with $B = B_0 e^{rt}$) admits a solution $(\lambda, \tilde{r})$. From the first and the third equation we get $\tilde{r} = r$ and $\lambda = (\mu_S - rS)/\sigma_S$. The condition to be fulfilled is therefore

$$\mu_C - r\pi_C = \sigma_C\,\frac{\mu_S - rS}{\sigma_S} = (\mu_S - rS)\frac{\partial \pi_C}{\partial S}\,.$$

With use of Itô's rule to expand μ_C, this condition may be written as follows in terms of the original problem data:

$$\frac{\partial \pi_C}{\partial t} + rS\,\frac{\partial \pi_C}{\partial S} + \tfrac{1}{2}\sigma_S^2\,\frac{\partial^2 \pi_C}{\partial S^2} = r\pi_C.$$

c. Apply the replication recipe:

$$\begin{bmatrix} \sigma_B & \pi_B \end{bmatrix} = \begin{bmatrix} \phi_S & \phi_C \end{bmatrix} \begin{bmatrix} \sigma_S & S \\ \sigma_C & \pi_C \end{bmatrix}.$$

Since $\sigma_B = 0$ and $\pi_B = B_0 e^{rt}$, we get

$$0 = \phi_S\sigma_S + \phi_C\sigma_C = \phi_S\sigma_S + \phi_C\frac{\partial \pi_C}{\partial S}\sigma_S$$

so that $\phi_S = -(\partial \pi_C/\partial S)\phi_C$, and

$$e^{rt}B_0 = \phi_S S + \phi_C\pi_C = \phi_C\big(-(\partial \pi_C/\partial S)S + \pi_C\big).$$

We find $\phi_C = e^{rt}B_0/\big(\pi_C - (\partial \pi_C/\partial S)S\big)$ and $\phi_S = -(\partial \pi_C/\partial S)\phi_C$.

7. **a**. We have

$$\frac{\partial f}{\partial x}(S_0, x) = S_0\phi(x) - e^{-rT}K\phi(x - \sigma\sqrt{T})\,.$$

This is equal to 0 when

$$\frac{S_0}{e^{-rT}K} = \frac{\phi(x - \sigma\sqrt{T})}{\phi(x)} = \exp\left(x\sigma\sqrt{T} - \tfrac{1}{2}\sigma^2 T\right).$$

From this one verifies that the derivative is 0 at $x = d_1$.

b. Since $C_0 = f(S_0, d_1)$, we have

$$\frac{\partial C_0}{\partial S_0} = \frac{\partial f}{\partial S_0}(S_0, d_1) + \frac{\partial f}{\partial x}(S_0, d_1)\,\frac{\partial d_1}{\partial S_0}(S_0) = \Phi(d_1)$$

because

$$\frac{\partial f}{\partial S_0}(S_0, x) = \Phi(x), \qquad \frac{\partial f}{\partial x}(S_0, x)\Big|_{x=d_1} = 0.$$

c. Write $z = \Phi^{-1}(y)$, and let the derivative of z with respect to y be denoted by z'. Since $\Phi(z) = y$ one finds by differentiation with respect to y that $\phi(z)z' = 1$, i.e. $z' = 1/\phi(z)$. Therefore

$$g'(y; a) = \phi(a + z)\,\frac{1}{\phi(z)} = \exp\left(-\tfrac{1}{2}a^2 - az\right).$$

Because $a > 0$ and y is an increasing function of z, it follows that $g'(y; a)$ is a decreasing function of y. Consequently, $g(y; a)$ is concave.

d. It was already shown in part a. that the function $x \mapsto f(S_0, x)$ has a stationary point at $x = d_1$. To show that we actually have a maximum, note that, if we define $y = \Phi(x - \sigma\sqrt{T})$, then

$$f(S_0, x) = S_0 g\left(y; \sigma\sqrt{T}\,\right) - ye^{-rT}K.$$

Since y and x are monotonically related, any extremum of the left hand side as a function of x is also be extremum of the right hand side as a function of y, and the two extrema are of the same type. Since the right hand side is concave by part c., only a single maximum can occur.

8. **a**. The condition for absence of arbitrage and completeness is that the matrix

$$\begin{bmatrix} \sigma_Y & \pi_Y \end{bmatrix} = \begin{bmatrix} \sigma S & 0 & S \\ \sigma_1 F & \sigma_2 F & F \\ 0 & 0 & B \end{bmatrix}$$

should be invertible. This is indeed the case, since the determinant of the matrix is

$\sigma S \cdot \sigma_2 F \cdot B$ and all factors are positive (for S_t and F_t this holds because they follow geometric Brownian motions).

b. The market prices of risk corresponding to the two Brownian motions can be computed from the equation $\mu_Y - r\pi_Y = \sigma_Y \lambda$:

$$\begin{bmatrix} (\mu - r)S \\ (\mu_1 - r)F \end{bmatrix} = \begin{bmatrix} \sigma S & 0 \\ \sigma_1 F & \sigma_2 F \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}.$$

If $\lambda_2 = 0$, then $(\mu - r)S = \sigma S \lambda_1$ as well as $(\mu_1 - r)F = \sigma_1 F \lambda_1$, so that $(\mu - r)/\sigma = (\mu_1 - r)/\sigma_1$, or in other words

$$\mu_1 - r = \frac{\sigma_1}{\sigma}(\mu - r).$$

Since according to economic theory the price of risk associated to a particular source of risk is determined by the correlation of this source to the wealth of the average investor, the assumption $\lambda_2 = 0$ could be justified when W_2 represents the part of the risk in the investment fund that is uncorrelated to the general market (idiosyncratic risk).

9. **b**. See the graphical output in Fig. B.6 (left panel). It appears that, when conditioning is done on the value at maturity of the underlying, the conditional mean is more or less constant, but the conditional variance is larger at points close to the strike. In scenarios in which the value of the underlying is close to strike when maturity approaches, the delta hedge becomes very sensitive to the value of the underlying; this causes a relatively large variance of the hedge error.

c. See the graphical output in Fig. B.6 (right panel). The plot indicates that, when realized volatility is less than the volatility that is assumed in the computation of the option price, the hedge error tends to be positive, which means that the hedger has a profit. On the other hand, if the realized volatility is higher, then the hedger experiences a loss. This is a natural relationship in view of the fact that the option price is increasing in volatility. The conditional variance, on the other hand, appears approximately constant when conditioning is done on realized volatility.

10. **b**. See the script in Code Example B.2. Graphical output is shown in Fig. B.7 and Fig. B.8. The standard deviation of the stop-loss strategy is about 6, while the delta hedge produces a standard deviation of approximately 2. These are nonnegligible numbers compared to the Black-Scholes price of the option, which is 9.93.

c. The standard deviation associated to the stop-loss strategy doesn't decrease much when the time step is reduced. In contrast, the standard deviation induced by the delta hedge is reduced each time; its value for $\Delta t = 10^{-4}$ is approximately 0.06.

d. In the limit, the scatter diagram for the delta hedge reduces to a straight line

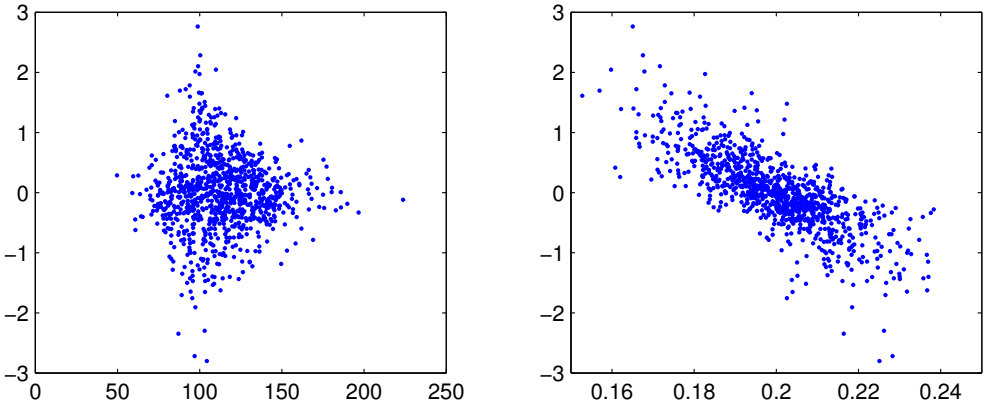

Figure B.6: Graphical output for Exc. 3.9. The left panel shows dependence of the hedge error on the value of the underlying at maturity, whereas the right panel shows dependence on the realized volatility.

```
mu = 0.12; sigma = 0.2; r = 0.04; T = 1; S0 = 100; K = 100;
Kh = exp(-r*T)*K; M = 1000; dt = 0.1;
t = 0; Vstl = 0; Vdel = 0; Sh = S0;
while t < T-0.5*dt
   phstl = Sh > Kh;
   phdel = normcdf((log(Sh/Kh)+0.5*sigma^2*(T-t))/(sigma*sqrt(T-t)));
   dW    = sqrt(dt)*randn(M,1);
   Shnw  = Sh.*exp((mu-r-0.5*sigma^2)*dt + sigma*dW);
   dVstl = phstl.*(Shnw-Sh);
   dVdel = phdel.*(Shnw-Sh);
   t = t + dt; Vstl = Vstl + dVstl; Vdel = Vdel + dVdel; Sh = Shnw;
end
CTh = max(Sh-Kh,0);
figure(1); plot(CTh,Vstl,'.')
xlabel('option payoff'); ylabel('sum of trading gains')
figure(2); plot(CTh,Vdel,'.')
xlabel('option payoff'); ylabel('sum of trading gains')
disp(['std from stop-loss hedge: ' num2str(std(CTh-Vstl))])
disp(['std from delta hedge:     ' num2str(std(CTh-Vdel))])
```

Code Example B.2: Code for Exc. 10.

with slope 1. The Black-Scholes price is equal to minus the intercept.

11. **a**. Let C_T^i denote the payoff at time T of a standard call option written on the asset S_t with strike K_i $(i = 1, 2, 3)$. We then have

$$C_T = \frac{L}{K_2 - K_1} C_T^1 - \left(\frac{L}{K_2 - K_1} + \frac{L}{K_3 - K_2} \right) C_T^2 + \frac{L}{K_3 - K_2} C_T^3.$$

In a similar way, any derivative whose payoff is a continuous piecewise linear function of S_T can be written as a linear combination of call options (and a constant payoff, if the payoff is not zero for low values of S_T).

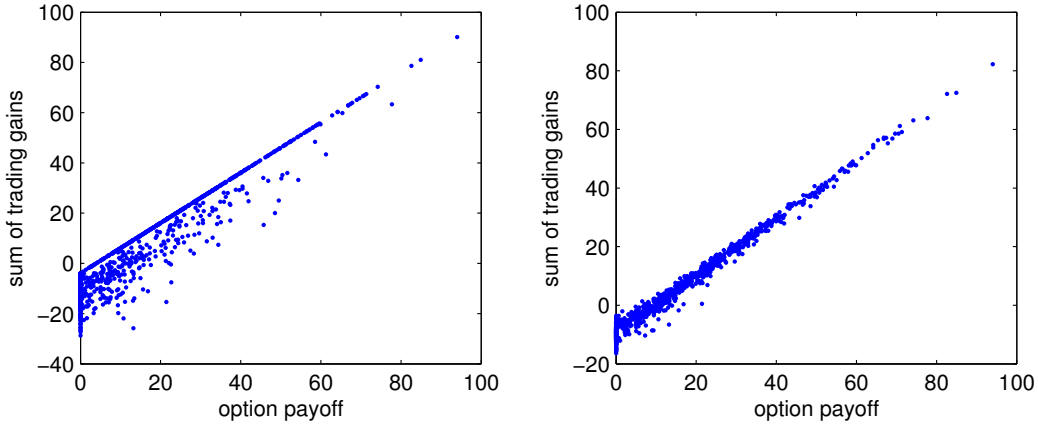

Figure B.7: Graphical output for Exc. 3.10. The left panel is for the stop-loss hedge, the right panel for the delta hedge. The plots have been created with time step $\Delta t = 0.1$.

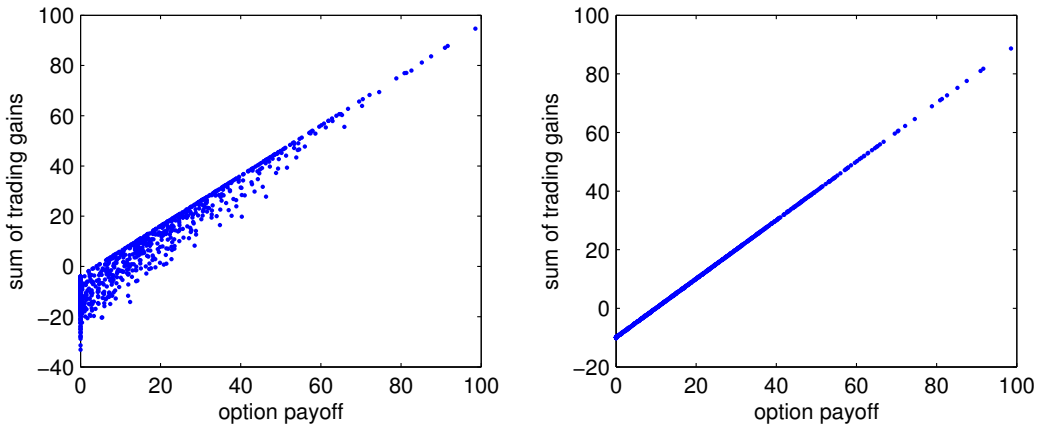

Figure B.8: Same as in Fig. B.7, but now with $\Delta t = 10^{-4}$.

b. Since it is given that H_T is obtained as the payoff of an optimal linear combination of available assets, the quantity $\mathrm{var}(C_T - \alpha H_T)$ must be minimized at $\alpha = 1$. Since $\mathrm{var}(C_T - \alpha H_T) = \mathrm{var}(C_T) - 2\alpha\,\mathrm{cov}(C_T, H_T) + \alpha^2\,\mathrm{var}(H_T)$, it follows from this that $\mathrm{cov}(C_T, H_T) = \mathrm{var}(H_T)$. Therefore,

$$\frac{\mathrm{var}(C_T - H_T)}{\mathrm{var}(C_T)} = 1 - \frac{\mathrm{cov}(C_T, H_T)}{\mathrm{var}(C_T)} = 1 - \frac{\mathrm{cov}(C_T, H_T)^2}{\mathrm{var}(C_T)\,\mathrm{var}(H_T)} = 1 - \rho^2.$$

The expression (3.99) follows from this. For HQ = 0.5 one needs $\rho = 0.87$; HQ = 0.8 requires $\rho = 0.98$. A plot of the hedge quality as a function of the correlation coefficient is shown in Fig. B.9. It appears that high-quality hedging is only possible if very high replication accuracy is achievable. Of course, this holds for dynamic hedging just as well as for static hedging.

14. **a**. The given model can be described with B^E, B^A, and Z as state variables,

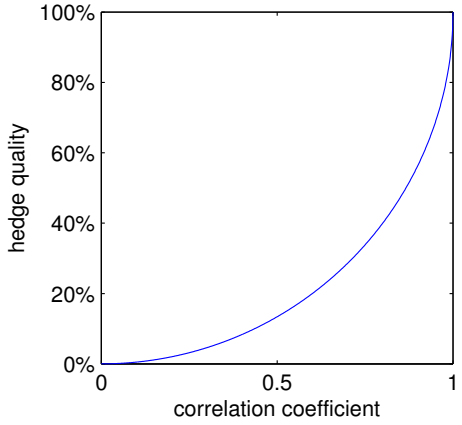

Figure B.9: Hedge quality as a function of correlation between product payoff and value of hedge portfolio at the time of maturity, under the assumption of optimality. The curve shown is a quartercircle.

and with $Y_1 := S$ and $Y_2 := B^E$ as traded assets. We have

$$dS = Z\,dB^A + B^A\,dZ = (r_A Z B^A + \mu Z B^A)\,dt + \sigma Z B^A\,dW$$

so that

$$[\sigma_Y \;\; \pi_Y] = \begin{bmatrix} \sigma Z B^A & Z B^A \\ 0 & B^E \end{bmatrix}.$$

From the fact that the above matrix is invertible (assuming $\sigma \neq 0$) for all relevant (i. e. positive) values of the state variables it follows that the market is complete. The risk-free rate of return r and the market price of risk λ are found by solving the Black-Scholes equation $\mu_Y = r\pi_Y + \sigma_Y\lambda$, which in this case is

$$\begin{bmatrix} (r_A + \mu) Z B^A \\ r_E B^E \end{bmatrix} = \begin{bmatrix} \sigma Z B^A & Z B^A \\ 0 & B^E \end{bmatrix} \begin{bmatrix} \lambda \\ r \end{bmatrix}$$

so that $r = r_E$ and $\lambda = (\mu + r_A - r_E)/\sigma$.

b. Under the equivalent martingale measure corresponding to the numéraire B^E, the relative price process S_t/B_t^E should be a martingale. Write

$$\begin{aligned} d\,\frac{S}{B^E} = d\left[Z\,\frac{B^A}{B^E}\right] &= Z(r_A - r_E)\,\frac{B^A}{B^E}\,dt + \frac{B^A}{B^E}(\mu Z\,dt + \sigma Z\,dW) \\ &= Z\,\frac{B^A}{B^E}\,((\mu + r_A - r_E)\,dt + \sigma\,dW) \\ &= Z\,\frac{B^A}{B^E}\,\sigma\,d\widetilde{W} \end{aligned}$$

where $\widetilde{W}$ satisfies

$$d\widetilde{W} = \frac{\mu + r_A - r_E}{\sigma}\, dt + dW.$$

The equivalent martingale measure is given, through Girsanov's theorem, as the measure under which the process $\widetilde{W}$ is a Brownian motion. For the exchange rate dynamics, one obtains

$$dZ = \mu Z\, dt + \sigma Z\, (d\widetilde{W} - \frac{\mu + r_A - r_E}{\sigma}\, dt) = (r_E - r_A) Z\, dt + \sigma Z\, d\widetilde{W}.$$

15. **c**. The number of steps required should not increase by more than one or two. Once it is close to the to be computed value, Newton's method is really fast.

B.4 Exercises from Chapter 4

1. The BS model under the measure $\mathbb{Q}_S$ is given by (see (4.41)

$$dS_t = (r + \sigma^2) S_t\, dt + \sigma S_t\, dW_t$$

where W_t is a Brownian motion under $\mathbb{Q}_S$. From the standard solution formula for geometric Brownian motion, it follows that

$$S_T = S_0 \exp\big((r + \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\, Z\big), \qquad Z \overset{\mathbb{Q}_S}{\sim} N(0,1).$$

We have $S_T > K$ in case $Z > -d_1$, with d_1 as in (3.50b). Let C_0 denote the time-0 price of the digital option. By the numéraire-dependent pricing formula, we have

$$\begin{aligned}
\frac{C_0}{S_0} &= E^{\mathbb{Q}_S}\left[\frac{\mathbb{1}_{S_T > K}}{S_T}\right] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-d_1}^{\infty} S_0^{-1} \exp\big(-(r + \tfrac{1}{2}\sigma^2)T - \sigma\sqrt{T}\, z\big) \exp(-\tfrac{1}{2} z^2)\, dz \\
&= \frac{e^{-rT}}{S_0\sqrt{2\pi}} \int_{-d_1}^{\infty} \exp\big(-\tfrac{1}{2}\big(z + \sigma\sqrt{T}\big)^2\big)\, dz \\
&= \frac{e^{-rT}}{S_0}\, \Phi(d_2).
\end{aligned}$$

From this it follows that $C_0 = e^{-rT}\Phi(d_2)$.

3. **a**. First rewrite the model under the risk-neutral measure associated to the money market account M. The relative drift of all traded assets is then equal to the short rate. Therefore the model becomes

$$dS_t = r S_t\, dt + \sigma S_t\, dW_{1,t}^M$$

$$dF_t = rF_t\,dt + \sigma_1 F_t\,dW^M_{1,t} + \sigma_2 F_t\,dW^M_{2,t}$$
$$dB_t = rB_t\,dt$$

where $W^M_{1,t}$ and $W^M_{2,t}$ are Brownian motions under the risk-neutral measure. Next, from the change-of-numéraire formula

$$\mu^N_C = \mu^M_C + \sigma_C\,\frac{\sigma_N^\top}{\pi_N}$$

we get in this case ($N = S$):

$$\mu^S_S = rS + \begin{bmatrix}\sigma S & 0\end{bmatrix}\begin{bmatrix}\sigma S\\ 0\end{bmatrix}\frac{1}{S} = (r+\sigma^2)S$$

$$\mu^S_F = rF + \begin{bmatrix}\sigma_1 F & \sigma_2 F\end{bmatrix}\begin{bmatrix}\sigma S\\ 0\end{bmatrix}\frac{1}{S} = (r+\sigma\sigma_1)F.$$

The model becomes

$$dS_t = (r+\sigma^2)S_t\,dt + \sigma S_t\,dW^S_{1,t}$$
$$dF_t = (r+\sigma\sigma_1)F_t\,dt + \sigma_1 F_t\,dW_{1,t} + \sigma_2 F_t\,dW^S_{2,t}$$
$$dB_t = rB_t\,dt$$

where $W^S_{1,t}$ and $W^S_{2,t}$ are Brownian motions under the equivalent martingale measure that corresponds to taking the stock price S_t as the numéraire.

b. Let C_t denote the contract value at time t. By the numéraire-dependent pricing formula:

$$\frac{C_0}{S_0} = E^{\mathbb{Q}_S}\left[\frac{C_T}{S_T}\right] = E^{\mathbb{Q}_S}\big[\max(1, F_T/S_T)\big].$$

From Itô's rule, we get:

$$d(\log S_t) = (r + \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dW^S_{1,t}$$
$$d(\log F_t) = \big(r + \sigma\sigma_1 - \tfrac{1}{2}(\sigma_1^2+\sigma_2^2)\big)\,dt + \sigma_1\,dW^S_{1,t} + \sigma_2\,dW^S_{2,t}.$$

The distribution of F_T/S_T under $\mathbb{Q}_S$ is therefore given by

$$F_T/S_T = (F_0/S_0)\exp\Big(\big(\sigma_1\sigma - \tfrac{1}{2}(\sigma^2+\sigma_1^2+\sigma_2^2)\big)T + \sqrt{(\sigma_1-\sigma)^2+\sigma_2^2}\,\sqrt{T}\,Z\Big)$$

where Z is standard normal. The value of the contract at time 0 can now be written as (defining $\bar\sigma := \sqrt{(\sigma_1-\sigma)^2+\sigma_2^2}$)

$$C_0 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\max\left(S_0, F_0\exp(-\tfrac{1}{2}\bar\sigma^2 T + \bar\sigma\sqrt{T}\,z)\right)\exp(-\tfrac{1}{2}z^2)\,dz.$$

```
r = 0.04; sigma = 0.2; T = 1; S0 = 100; K = 100;
F = @(z) max(S0*exp((r-0.5*sigma^2)*T + sigma*sqrt(T)*z)-K,0);
% --- part a. ---
N = 5; Ze = 4; a = pi/(2*Ze); [G,g] = BF(a,N); % basis functions
Zt = -Ze:Ze; coeffs = BM(G,Zt)\F(Zt)'; % regression
Zm = -6:0.1:6; figure(1); plot(Zm,F(Zm),Zm,BM(G,Zm)*coeffs,'--')
approxval = g*coeffs;
d1 = @(S0) (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));          % \
d2 = @(S0) (log(S0/K)+(r-0.5*sigma^2)*T)/(sigma*sqrt(T));          %  BS
exactval = @(S0) S0.*normcdf(d1(S0))-exp(-r*T)*K*normcdf(d2(S0)); % /
disp(['approximate value (N=' num2str(N) '): ' num2str(approxval)])
disp(['exact value:              ' num2str(exactval(S0))])
% --- part b. ---
fn = @(eta) g * RM(a*log(eta)/sigma*sqrt(T),N) * coeffs;
etas = 0.6:0.01:1.4;
approxvals = zeros(size(etas)); % reserve memory space
for k = 1:length(etas)
    approxvals(k) = fn(etas(k));
end
figure(2); plot(etas*S0,exactval(etas*S0),etas*(S0),approxvals,'--')
```

Code Example B.3: Code for Exc. 8. The code uses auxiliary functions `BF`, `BM` and `RM` as shown in Code Examples B.4, B.5 and B.6.

Define d_1 and d_2 by

$$d_1 = \frac{\log(F_0/S_0) + \frac{1}{2}\bar{\sigma}^2 T}{\bar{\sigma}\sqrt{T}}, \qquad d_2 = \frac{\log(F_0/S_0) - \frac{1}{2}\bar{\sigma}^2 T}{\bar{\sigma}\sqrt{T}}.$$

Then we can write

$$\begin{aligned} C_0 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-d_2} S_0 \exp(-\tfrac{1}{2}z^2)\, dz + \frac{1}{\sqrt{2\pi}} \int_{-d_2}^{\infty} F_0 \exp\big(\tfrac{1}{2}(z-\bar{\sigma})^2\big)\, dz \\ &= S_0\Phi(-d_2) + F_0\Phi(d_1). \end{aligned}$$

In particular, if $F_0 = S_0$, then $d_2 = -d_1$ and the value of the contract is given by

$$C_0 = 2\Phi\big(\bar{\sigma}\sqrt{T}\big)S_0.$$

8. A possible code for parts a. and b. is shown in Code Example B.3. There is no claim that the computational parameters used in the example code are optimal. Graphical output is shown in Fig. B.10.

```
function [ G, g ] = BF(a,N)
%BF Basis functions and corresponding expectations. N must be odd.
G = cell(1,N); g = zeros(1,N);
G{1} = @(Z) 1; g(1) = 1;
for k = 1:(N-1)/2
    G{2*k}   = @(Z) sin(a*k*Z);
    G{2*k+1} = @(Z) cos(a*k*Z);
    g(2*k+1) = exp(-0.5*a^2*k^2);
end
```

Code Example B.4: Auxiliary function for code shown in Code Example B.3.

```
function bmatrix = BM(G,Z)
%BM Basis functions evaluated in points in Z.
bmatrix = zeros(length(Z),length(G));
for k = 1:length(G)
    bmatrix(:,k) = G{k}(Z);
end
```

Code Example B.5: Auxiliary function for code shown in Code Example B.3.

```
function R = RM(x,N)
%RM Rotation matrix of size N. N must be odd.
R = zeros(N);
R(1,1) = 1;
for k = 1:(N-1)/2
    R(2*k,2*k)     =  cos(k*x);
    R(2*k,2*k+1)   = -sin(k*x);
    R(2*k+1,2*k)   =  sin(k*x);
    R(2*k+1,2*k+1) =  cos(k*x);
end
```

Code Example B.6: Auxiliary function for code shown in Code Example B.3.

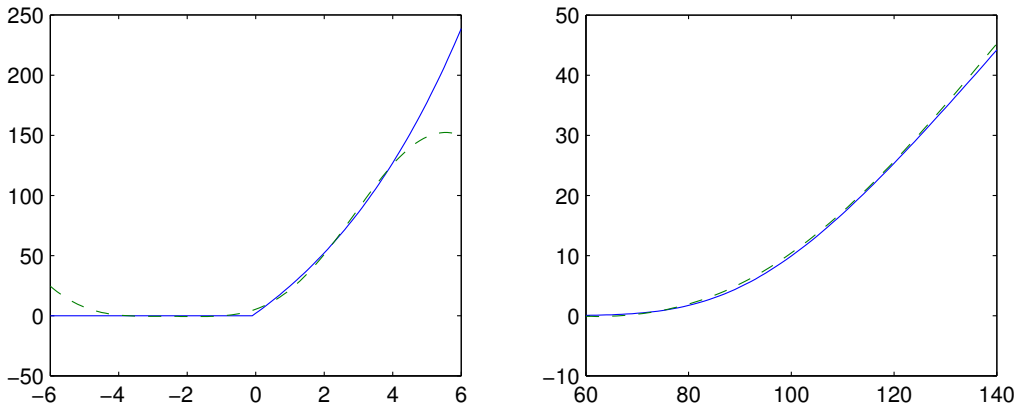

Figure B.10: Graphical output for Exc. 8. The left panel shows the approximation of the function $F(z)$ (payoff function in terms of a standard normal variable) by a linear combination of five basis functions. The right panel shows the corresponding approximation of the option value at time 0 as a function of the time-0 value of the underlying. In both panels, the drawn curve represents the exact values, and the dashed curve is the approximation.

B.5 Exercises from Chapter 5

1. Differentiate both sides of the equality

$$\int_0^T P(t)\,dt = \frac{1-P(T)}{S(T)}$$

with respect to T.

3. **a**. From the balancing rule (5.85) it follows that the amount received by mr. Balding at time T_1 is

$$b^1_{T_1} = \alpha_{T_1} b^1_0 = \frac{A_{T_1}}{P_{T_1}(T_1)b^1_0 + P_{T_1}(T_2)b^2_0}\, b^1_0 = \frac{A_{T_1}}{1+\eta P_0(T_1)P_{T_1}(T_2)/P_0(T_2)} \tag{B.1}$$

where $\eta := A^2_0/A^1_0$. Since the process $P_t(T_2)/P_t(T_1)$ is a martingale under the T_1-forward measure, we have

$$\frac{P_0(T_2)}{P_0(T_1)} = E^{\mathbb{Q}_{T_1}} \frac{P_{T_1}(T_2)}{P_{T_1}(T_1)} = E^{\mathbb{Q}_{T_1}} P_{T_1}(T_2). \tag{B.2}$$

In particular, if $P_{T_1}(T_2)$ is deterministic as seen from time 0, then $P_{T_1}(T_2) = P_0(T_2)/P_0(T_1)$ (as can also be inferred from a direct arbitrage argument). The expression (B.1) can then be rewritten as $b^1_{T_1} = A_{T_1}/(1+\eta)$. Since the time-0 value of the amount A_{T_1} paid at time T_1 is A_0, it follows that the time-0 value of mr. Balding's benefit is $A_0/(1+\eta) = A^1_0$.

b. On the basis of the assumption of independence and (B.1), the time-0 value of the amount received by mr. Balding is

$$\begin{aligned} P(0,T_1)b^1_{T_1} &= P(0,T_1)E^{\mathbb{Q}_{T_1}}\left[\frac{A_{T_1}}{1+\eta P_0(T_1)P_{T_1}(T_2)/P_0(T_2)}\right] \\ &= P(0,T_1)E^{\mathbb{Q}_{T_1}}\left[A_{T_1}\right]E^{\mathbb{Q}_{T_1}}\left[\frac{1}{1+\eta P_0(T_1)P_{T_1}(T_2)/P_0(T_2)}\right] \\ &= A_0 E^{\mathbb{Q}_{T_1}}\left[\frac{1}{1+\eta X}\right] \end{aligned}$$

where $X := P_0(T_1)P_{T_1}(T_2)/P_0(T_2)$ is a positive random variable that satisfies $E^{\mathbb{Q}_{T_1}} X = 1$ by (B.2). The function $f(x) = 1/(1+\eta x)$ is strictly convex for $x > 0$ and $\eta > 0$. It follows that

$$E^{\mathbb{Q}_{T_1}}\left[\frac{1}{1+\eta X}\right] > \frac{1}{1+\eta E^{\mathbb{Q}_{T_1}}[X]} = \frac{1}{1+\eta}$$

where the strict inequality is due to the assumption that X is not degenerate. Consequently, the time-0 value of mr. Balding's benefit is larger than $A_0/(1+\eta) = A_0^1$.

4. **a**. The assumption that the short rate is constant in time under the real-world measure implies that it is also constant in time under the risk-neutral measure. The statement therefore follows from (5.23) and (5.2).

b. From the relation $P_t(T) = \exp(-r_t(T-t))$, it follows by Itô's formula that

$$dP_t(T) = \exp(-r_t(T-t))[-(T-t)\,dr_t + r_t\,dt] + \tfrac{1}{2}\exp(-r_t(T-t))(T-t)^2\sigma_r\sigma_r^\top$$

which implies

$$\mu_T = \pi_T\big[r_t - (T-t)\mu_r + \tfrac{1}{2}(T-t)^2\sigma_r\sigma_r^\top\big].$$

The no-arbitrage condition $\mu_T = r\pi_T$ therefore implies

$$-(T-t)\mu_r + \tfrac{1}{2}(T-t)^2\sigma_r\sigma_r^\top = 0 \quad \Rightarrow \quad \mu_r = \tfrac{1}{2}(T-t)\sigma_r\sigma_r^\top.$$

Since this relation must hold for all T and $t \le T$ whereas μ_r and σ_r do not depend on T, it follows that $\sigma_r\sigma_r^\top = 0$ and $\mu_r = 0$. Because σ_r is a row vector, we have $\sigma_r\sigma_r^\top = \sum_{i=1}^k (\sigma_r)_i^2$ so that from $\sigma_r\sigma_r^\top = 0$ it follows that in fact $\sigma_r = 0$. Consequently, r_t is constant.

5. Suppose for instance that $R_{T_1}(T_2) < F_0(T_1, T_2)$. At time 0, buy a bond with face value $P_0(T_1)/P_0(T_2)$ that matures at time T_2, and sell a bond with face value 1 that matures at time T_1. The total cash outlay for these two transactions is 0. At time T_1, the value of the second bond is 1, and the value of the first bond according to the prescribed scenario is

$$\begin{aligned} P_{T_1}(T_2)\,\frac{P_0(T_1)}{P_0(T_2)} &= \exp\big(-R_{T_1}(T_2)(T_2-T_1)\big)\,\frac{P_0(T_1)}{P_0(T_2)} \\ &> \exp\big(-F_0(T_1,T_2)(T_2-T_1)\big)\,\frac{P_0(T_1)}{P_0(T_2)} = 1 \end{aligned}$$

where in the last step the definition of the forward rate is used. Meeting the obligation from the second bond and cashing in the proceeds from selling the second bond at time T_1 leaves an immediate profit. If $R_{T_1}(T_2) > F_0(T_1, T_2)$, then an arbitrage is constructed in a similar way, replacing "buy" by "sell" and vice versa.

To show that no arbitrage is possible in case the prescribed scenario is such that $R_{T_1}(T_2) = F_0(T_1, T_2)$ for all $T_2 \ge T_1$, consider the deterministic model in which the evolution of the short rate is prescribed by $r_t = F_0(t)$, where $F_0(T)$ is the instantaneous forward rate that holds in the market at time 0. From (5.23) it

follows that the arbitrage-free bond prices in this model are given by

$$P_t(T) = \exp\Big(-\int_t^T F_0(s)\,ds\Big) = \exp\Big(\int_t^T \frac{d}{ds}\log P(s)\,ds\Big)$$
$$= \exp\big(\log P_0(T) - \log P_0(t)\big) = \frac{P_0(T)}{P_0(t)}\,.$$

Since $P_0(0) = 1$, this shows in particular that the bond prices from the model are consistent with currently observed bond prices. The yield curve at time t is given by

$$R_t(T) = \frac{-1}{T-t}\log P_t(T) = -\frac{\log P_0(T) - \log P_0(t)}{T-t} = \frac{TR_0(T) - tR_0(t)}{T-t} = F_0(t,T)$$

where (5.6) is used in the last step. This evolution of the yield curve does not admit arbitrage, since it is derived from an arbitrage-free model.

10. **a**. Since at time t the portfolio that constitutes the constant-maturity bond consists fully of bonds with maturity $t+T$, and the portfolio is self-financing, the volatility of the constant-maturity bond at time t is equal to the volatility of the bond with maturity $t+T$ times the number of those bonds held in the portfolio. Therefore,

$$dV_t = h(t,X_t)V_t\,dt + \frac{\partial \pi_{t+T}}{\partial x}(t,X_t)\sigma_X(t,X_t)\,\frac{1}{\pi_{t+T}(t,X_t)}V_t\,dW_t \tag{B.3}$$

where the drift term follows from (3.80).

b. Using the Vasicek bond pricing formula as given in (4.66), one finds from (B.3)

$$dV_t = r_tV_t\,dt - \frac{1-e^{-aT}}{a}\sigma V_t\,dW_t \tag{B.4}$$

where W_t is a Brownian motion under the risk-neutral measure and r_t follows (4.64). If the price of risk λ in the Vasicek model is constant, then the corresponding model under the real-world measure $\mathbb{P}$ is

$$dV_t = \Big(r_t - \frac{1-e^{-aT}}{a}\sigma\lambda\Big)V_t\,dt - \frac{1-e^{-aT}}{a}\sigma V_t\,dW_t$$

where now W_t denotes a $\mathbb{P}$-Brownian motion.

B.6 Exercises from Chapter 6

1. In four decimals, the smallest (i.e. most negative) eigenvalue of the matrix D, for matrix size 100, is -3.9990, and the largest is -0.0010. Code Example B.7

```
N = 720;
D = -2*eye(N) + diag(ones(1,N-1),-1) + diag(ones(1,N-1),1);
[V,~] = eig(D);
for i = 1:N
    plot(V(:,i)); pause(0.1)
end
```

Code Example B.7: Approximate eigenfunctions of the second order differential operator. The effect of the presentation relies on Matlab's habit of presenting the eigenvalues in order of magnitude.

shows all of the eigenvectors for matrix size 720, in rapid succession. The sinusoidal appearance is related to the fact that the solutions of the differential equation $y''(x) = \lambda y(x)$, for $\lambda < 0$, are of the form $y(x) = a\sin\bigl(\sqrt{-\lambda}x + b\bigr)$ where a and b are arbitrary constants.

2. Taking $n = 4$ as an example, one can write

$$x^\top Dx = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$= -2x_1^2 + 2x_1x_2 - 2x_2^2 + 2x_2x_3 - 2x_3^2 + 2x_3x_4 - 2x_4^2$$
$$= -x_1^2 - (x_1 - x_2)^2 - (x_2 - x_3)^2 - (x_3 - x_4)^2 - x_4^2.$$

It is seen that $x^\top Dx < 0$ for all $x \in \mathbb{R}^4$ with $x \neq 0$. The corresponding statement for general n follows in the same way. Consequently, the matrix is D is negative definite, so that all eigenvalues must be less than 0. By similar reasoning one finds that $D+4I$ is positive definite; this means that all eigenvalues of $D+4I$ are positive, or in other words that all eigenvalues of D are larger than -4. The second claim follows by noting that, for $x = \mathbb{1}$, $x^\top Dx/x^\top x = -2/\sqrt{n}$ and $x^\top(D+4I)x/x^\top x = 2/\sqrt{n}$. (By using suitable other vectors than $x = \mathbb{1}$, sharper bounds can be obtained.)

B.7 Exercises from Chapter 7

2. **b**. The averaging that takes place in the Asian option effectively lowers the volatility, which suggest that the price of the Asian option should be less than that of its European counterpart. Indeed, the value of the Asian option is found to be approximately 5.04, as opposed to the European option value which is 8.92. The width of the confidence interval with $n = 10^5$ is approximately 0.1; to reduce this to 0.01 (not taking into account the time discretization error), one needs approximately 100 times more simulations.

```
r = 0.02; sigma = 0.2; T = 1; S0 = 100; K = 100; N = 100; dt=T/N;
% --- pilot sample ---
M  = 10^3; S = S0; W = 0; Ac = S0;
for k = 1:N
    dW = sqrt(dt)*randn(M,1); dS = r*S*dt + sigma*S.*dW;
    S = S + dS; W = W + dW; Ac = Ac + S;
    if k == N/2; hW = W; end
end
A = Ac/(N+1); F = exp(-r*T)*max(A-K,0);
coef  = [ones(size(W)) hW hW.^2 W W.^2 ]\F; % regression
coeff = coef(2:end);
% --- actual sample ---
M = 10^5; S = S0; W = 0; Ac = S0;
for k = 1:N
    dW = sqrt(dt)*randn(M,1); dS = r*S*dt + sigma*S.*dW;
    S = S + dS; W  = W + dW; Ac = Ac + S;
    if k == N/2; hW = W; end
end
A = Ac/(N+1); Fnc = exp(-r*T)*max(A-K,0);
ECV = [ 0 0.5*T 0 T ]; % expectations of control variates
Fcv = Fnc - [hW hW.^2 W W.^2 ]*coeff + ECV*coeff;
disp(['raw MC: ' num2str(mean(Fnc)) ' +/- ' ...
       num2str(1.96*std(Fnc)/sqrt(M))])
disp(['with control variates: ' num2str(mean(Fcv)) ' +/- ' ...
       num2str(1.96*std(Fcv)/sqrt(M))])
```

Code Example B.8: Script for Exc. 7.5.2.e.

c. The correlation coefficient is approximately 0.83. The use of the European option as a control variate should therefore reduce the confidence interval by a factor close to 2.

e. See the script in Code Example B.8. The confidence interval is reduced by approximately a factor 3 compared to Monte Carlo without control variates.

f. Hardly any improvement.

g. The variance-covariance matrix is

$$\Sigma_{YY} = \begin{bmatrix} \frac{1}{2}T & 0 & \frac{1}{2}T & 0 \\ 0 & \frac{1}{2}T^2 & 0 & \frac{1}{2}T^2 \\ \frac{1}{2}T & 0 & T & 0 \\ 0 & \frac{1}{2}T^2 & 0 & 2T^2 \end{bmatrix}.$$

However, the effort that goes into computing this matrix does not appear to pay off. Note that the variance as a function of the coefficients is quadratic and so is rather flat near the optimal values; therefore the coefficients can still produce a value close to the minimum even when they are not estimated very accurately. Moreover, only the matrix Σ_{YY} is known exactly; the vector Σ_{XY} still needs to be estimated.

3. **a**. The coefficients α_j $(j = 1, \ldots, k)$ are defined by

$$\alpha_j = \frac{\sum_{i=1}^n (x_i^j - \bar{x}^j)(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{B.5}$$

where

$$\bar{x}^j = \frac{1}{n}\sum_{i=1}^n x_i^j, \qquad y_i = \sum_{j=1}^k c_j x_i^j, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^n y_i. \tag{B.6}$$

Whether the draws x_j^i that are used in (B.5) are from a pilot sample or not, the relations (B.6) imply that $\sum_{j=1}^k c_j(x_i^j - \bar{x}^j) = y_i - \bar{y}$, and hence that the coefficients α_j that are defined by (B.5) satisfy $\sum_{j=1}^k c_j \alpha_j = 1$. Consequently,

$$\sum_{j=1}^k c_j \hat{x}_j^i = \sum_{j=1}^k c_j x_j^i - \sum_{j=1}^k c_j \alpha_j \left(\sum_{j=1}^k c_j x_j^i - EY \right) = EY.$$

Since this is a linear relation which holds for all samples separately, the same relation holds as well for the sample averages, as stated in (7.35).

b. A typical answer is: Beneficiary 20.55 ± 0.10, Charity 10.21 ± 0.14, Sponsor 12.48 ± 0.09.

c. No value at time 0 is created or destroyed by the formulation of the fund policy; the policy just determines the distribution of value between the parties involved and the fund's final value. Therefore the time-0 values of the payments received (or paid, when occurring with a negative sign), plus the time-0 value of the fund's assets at time T, must sum to the initial value of the fund. To verify that this equation is satisfied within the limits of Monte Carlo accuracy, form a Monte Carlo estimate of the random variable that is formed by taking the the cumulative benefits received by the Beneficiary plus the payment received by Charity minus the contribution of the Sponsor (all discounted to time 0), and check whether the confidence interval contains $S_0 - e^{-rT}L$ with $S_0 = 100$, $r = 0.02$, $T = 10$, $L = 100$.

d. A typical answer is: Beneficiary 20.50 ± 0.05, Charity 10.15 ± 0.07, Sponsor 12.52 ± 0.06. The use of the control variate results in a substantial reduction of each of the three confidence intervals. Moreover, the point estimates are now such that the relation of part c. is satisfied. For a script, see Code Example B.9.

4. See the script in Code Example B.11. Graphical output is shown in Fig. B.11. In part c., keep an eye on the size of the regression coefficients, in order to avoid situations in which a very large number (regression coefficient) is multiplied by a very small number (expectation associated to the corresponding basis function). Such situations are numerically sensitive and may lead to large errors, in spite of the correctness of the calculation if it could be carried out in infinite precision.

```
% pilot sample
Mp = 10^4; [ CB PC SC Y ~ ] = FundSim(Mp);
CVM  = cov([Y CB PC SC]);
a_CB = CVM(1,2)/CVM(1,1);
a_PC = CVM(1,3)/CVM(1,1);
a_SC = CVM(1,4)/CVM(1,1); % or use a_SC = a_CB + a_PC - 1
% actual calculation
Mc = 10^5; [ CB PC SC Y EY ] = FundSim(Mc);
CBcv = CB - a_CB*(Y - EY);
PCcv = PC - a_PC*(Y - EY);
SCcv = SC - a_SC*(Y - EY);
disp(['beneficiary: ' num2str(mean(CBcv)) ...
      ' +/- ' num2str(1.96*std(CBcv)/sqrt(Mc))])
disp(['charity:     ' num2str(mean(PCcv)) ...
      ' +/- ' num2str(1.96*std(PCcv)/sqrt(Mc))])
disp(['sponsor:     ' num2str(mean(SCcv)) ...
      ' +/- ' num2str(1.96*std(SCcv)/sqrt(Mc))])
```

Code Example B.9: Script for Exc. 7.5.3.a. For the function `FundSim`, see Code Example B.10.

```
function [ CB PC SC Y EY ] = FundSim(M)
%FundSim Fund simulation results in BS model
r = 0.02; sigma = 0.15; S0 = 100; L = 100; T = 10; dt = 1;
ben = @(S) (S>=95).*(S<105).*0.02.*S + (S>=105).*0.04.*S;
S = S0; B = 1; CB = 0; % cumulative benefits
for t = 0:dt:T-dt
    S       = S.*exp((r-0.5*sigma^2)*dt+sigma*sqrt(dt)*randn(M,1));
    B       = B*exp(r*dt);
    benefit = ben(S);
    S       = S - benefit;
    CB      = CB + benefit/B; % discount to time 0
end
SC = max(L-S,0)/B; % sponsor contribution, discounted to time 0
PC = max(S-L,0)/B; % payment to charity, discounted to time 0
Y  = CB + PC - SC;
EY = S0 - exp(-r*T)*L;
```

Code Example B.10: Function used in Code Example B.9.

7. a. The random variable $\Phi(Z)$ takes values in $[0,1]$, and for $0 \le z \le 1$ we have $P(\Phi(Z) \le z) = P(Z \le \Phi^{-1}(z)) = \Phi(\Phi^{-1}(z)) = z$. Furthermore,

$$E[f'(\Phi(Z))] = \int_{-\infty}^{\infty} f'(\Phi(z))\phi(z)\,dz = \int_0^1 f'(y)\,dy = f(1) - f(0)$$

where the substitution $y = \Phi(z)$ has been used.

d. See Fig. 7. In the case of $f(x) = x^{0.1}$, a large part of the contribution to $E[f'(\Phi(Z))]$ comes from samples in a region that is very far in the tail of the standard normal distribution.

```
r = 0.02; sigma = 0.2; T = 1; S0 = 100; K = 100;
ST = @(z) S0*exp((r-0.5*sigma^2)*T + sigma*sqrt(T)*z);
F  = @(z) max(ST(z)-K,0); % payoff
N = 7; Ze = 5; npts = 11; a = pi/(2*Ze); [G,g] = BF(a,N);
Zt = linspace(-Ze,Ze,npts); coeffs = BM(G,Zt)\F(Zt)'; % regression
M = 10^5; Zm = randn(M,1);
V = exp(-r*T)*(F(Zm) - BM(G,Zm)*coeffs + g*coeffs);
disp(['N = ' num2str(N) ', npts = ' num2str(npts)])
disp(['value = ' num2str(mean(V)) ' +/- ' num2str(1.96*std(V)/sqrt(M))])
Zx = linspace(-1.5*Ze,1.5*Ze);
figure(1); plot(Zx,F(Zx),Zx,BM(G,Zx)*coeffs,'--')
figure(2); plot(F(Zm),BM(G,Zm)*coeffs,'.',[0 150],[0 150])
xlabel('payoff'); ylabel('control variate')
```

Code Example B.11: Script for Exc. 7.5.4. For the auxiliary functions `BF` and `BM` that are used in the code, see code examples B.4 and B.5.

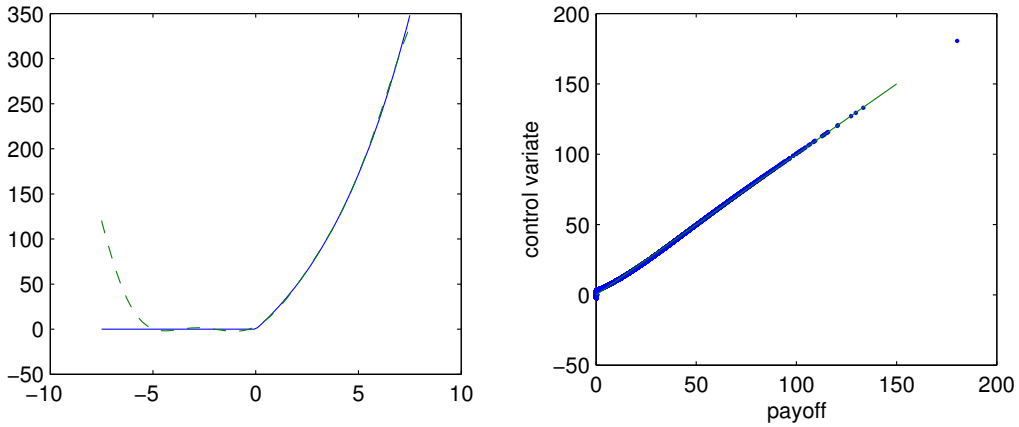

Figure B.11: Graphical output for Exc. 7.5.4.

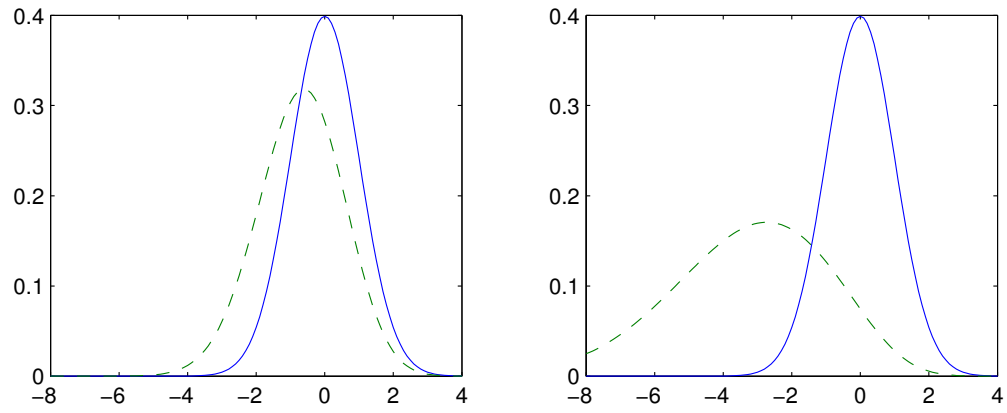

Figure B.12: Graphical output for Exc. 7.5.7.d. The left panel is for $f(x) = x^{0.5}$, the right panel is for $f(x) = x^{0.1}$. In both panels, the drawn curve represents the density $\phi(z)$ of the standard normal distribution, and the dashed curve represents the function $f'(\Phi(z))\phi(z)$.

Appendix C

Memorable formulas

This is a list of formulas that are candidates for commitment to memory. The following notational conventions are used:

C	price of a single asset
Y	vector of asset prices
V	portfolio value
ϕ	vector of portfolio holdings
M	money market account
N	general numéraire
X	general random variable (not necessarily an asset price)
Z	martingale.

C.1 Financial Models

$$\frac{C_0}{N_0} = E^{\mathbb{Q}_N}\left[\frac{C_T}{N_T}\right] \tag{NDPF}$$

$$\mu_C - r\pi_C = \sigma_C\lambda \tag{BSE}$$

$$\begin{cases} V = \phi^\top Y \\ dV = \phi^\top dY \end{cases} \tag{SFP}$$

$$\begin{bmatrix} \sigma_C & \pi_C \end{bmatrix} = \phi^\top \begin{bmatrix} \sigma_Y & \pi_Y \end{bmatrix} \tag{RR}$$

$$\mu^N_{C/N} = 0 \tag{FTAP}$$

$$dM_t = r_t M_t \, dt \tag{MMA}$$

$$dW^M = \lambda \, dt + dW^{\mathbb{P}} \tag{MPoR}$$

$$dW^N = dW^M - \frac{\sigma_N^\top}{\pi_N} \, dt \tag{CoN}$$

$$\mu_X^N = \mu_X^M + \sigma_X \frac{\sigma_N^\top}{\pi_N} \tag{CoD}$$

$$\mu_C^M = r\pi_C \tag{DURNM}$$

$$\mu_C^N = r\pi_C + \sigma_C \frac{\sigma_N^\top}{\pi_N} \tag{DUGN}$$

C.2 Stochastic Calculus

$$X_T - X_0 = \int_0^T dX_t \tag{TR}$$

$$d(f(X)) = f'(X) \, dX + \tfrac{1}{2} f''(X) \, d[X, X] \tag{IF}$$

$$d(f(t, X)) = \frac{\partial f}{\partial t} \, dt + \frac{\partial f}{\partial x} \, dX + \tfrac{1}{2} \frac{\partial^2 f}{\partial x^2} \, d[X, X] \tag{TIF}$$

$$d(f(X, Y)) = \frac{\partial f}{\partial x} \, dX + \frac{\partial f}{\partial y} \, dY + \tfrac{1}{2} \Big(\frac{\partial^2 f}{\partial x^2} \, d[X, X] + 2 \frac{\partial^2 f}{\partial x \partial y} \, d[X, Y] + \frac{\partial^2 f}{\partial y^2} \, d[Y, Y] \Big) \tag{BIF}$$

$$d[W, W] = dt \tag{QVBM}$$

$$d[X, X] = \sigma_X^2 \, dt \tag{QV}$$

$$d[X, Y] = \sigma_X \sigma_Y \rho \, dt \tag{QC}$$

$$E_{t_1}[Z_{t_2}] = Z_{t_1} \tag{MP}$$

$$E\left[\int_0^T X_t\, dZ_t\right] = 0 \tag{YCBTS}$$

$$\int_0^T f(t)\, dW_t \sim N\Big(0, \int_0^T f(t)^2\, dt\Big) \tag{SIDI}$$

$$\begin{cases} d\theta \;=\; -\theta\, \lambda^\top dW \\ d\widetilde{W} \;=\; \lambda\, dt + dW \end{cases} \tag{GT}$$

C.3 Stochastic Differential Equations

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t \quad \Rightarrow \quad X_T = X_0 \exp\left((\mu - \tfrac{1}{2}\sigma^2)T + \sigma W_T\right) \tag{GBM}$$

$$dX_t = -aX_t\, dt + \sigma\, dW_t \quad \Rightarrow \quad X_T = e^{-aT} X_0 + \int_0^T e^{-a(T-t)} \sigma\, dW_t \tag{LSDE}$$

C.4 Term Structure

$$P(T) = E^{\mathbb{Q}_M}\left[\exp\left(-\int_0^T r_t\, dt\right)\right] \tag{BPF}$$

$$R(T) = -\frac{1}{T}\, \log\, P(T) \tag{DF2YC}$$

$$F(T) = -\frac{d}{dT}\, \log\, P(T) \tag{DF2FC}$$

C.5 Key to acronyms

BIF	Bivariate Itô Formula
BPF	Bond Pricing Formula
BSE	Black-Scholes Equation (assets only)
CoD	Change of Drift (all variables)
CoN	Change of Numéraire
DF2FC	Discount Factor to Forward Curve
DF2YC	Discount Factor to Yield Curve
DUGN	Drift Under General Numéraire (assets only)
DURNM	Drift Under Risk-Neutral Measure (assets only)
FTAP	Fundamental Theorem of Asset Pricing (assets only)
GBM	Geometric Brownian Motion
GT	Girsanov Theorem
IF	Itô Formula
LSDE	Linear Stochastic Differential Equation
MMA	Money Market Account
MP	Martingale Property
MPoR	Market Price of Risk
NDPF	Numéraire-Dependent Pricing Formula
QC	Quadratic Covariation
QV	Quadratic Variation
QVBM	Quadratic Variation of Brownian Motion
RR	Replication Recipe
SIDI	Stochastic Integral with Deterministic Integrand
SFP	Self-Financing Portfolio
TIF	Time-dependent Itô Formula
TR	Telescope Rule
YCBTS	You Can't Beat The System

Appendix D

Notation

General notational convention: dependence on time is indicated by a subscript in the case of stochastic process (as in X_t), and by round brackets in the case of deterministic functions (as in $f(t)$).

$\mathbb{1}$	vector with entries all equal to 1
$\mathbb{1}_A$	random variable that takes the value 1 when event A occurs, and which is 0 otherwise
$B(n,p)$	binomial distribution
B_t	value of bond (riskless savings account) at time t
C_t	option price at time t
d	differential; parameter in binomial tree model representing a "down" move
d_1	parameter in option pricing formulas
d_2	parameter in option pricing formulas
E	expectation of a random variable
E_t	conditional expectation given information up to time t
$E^{\mathbb{Q}}$	expectation under the measure $\mathbb{Q}$
$H(t)$	variance-covariance matrix for the value taken by a given stochastic process at time t (used in particular for Gaussian processes)
I	identity matrix
J	number of successes in a repeated Bernoulli trial
k	number of Brownian motions in a state space model; also in use as a discrete time index, and as generic natural number
m	number of assets; also in use as a generic natural number
M	generic matrix; in numerical algorithms, also used as a number
M_t	money market account
n	number of state variables; also in use as a generic natural number

$\mathbb{N}$	set of natural numbers $\{1, 2, \dots\}$
$N(\mu, \sigma^2)$	normal distribution
N_t	value of numéraire at time t
o	order symbol: $f(x) = o(g(x))$ $(x \to a)$ means that $f(x)/g(x)$ tends to 0 as x tends to a
O	order symbol: $f(x) = O(g(x))$ $(x \to a)$ means that $f(x)/g(x)$ remains bounded as x tends to a
P	generic probability measure
$\mathbb{P}$	real-world probability measure
Q	generic probability measure
$\mathbb{Q}$	pricing measure
r	constant interest rate
r_t	short rate
$\mathbb{R}$	set of real numbers
S_t	value of a single asset at time t
t	continuous time
T	final time, time of expiry
u	parameter in binomial tree model representing "up" move
V_t	portfolio value at time t
W_t	standard scalar or vector Brownian motion; lies in $\mathbb{R}^k$
X_t	vector of state variables at time t; lies in $\mathbb{R}^n$. Also in use as generic stochastic process
Y_t	vector of asset prices at time t; lies in $\mathbb{R}^m$. Also in use as generic stochastic process
$\mathbb{Z}$	set of integers $\{\dots, -1, 0, 1, \dots\}$
cov	covariance
std	standard deviation
tr	trace (sum of the diagonal elements of a matrix)
var	variance
$\Gamma(z)$	gamma function
$\Gamma(z, w)$	upper incomplete gamma function
Δ	forward difference
Δt	length of a (small) time interval
θ	parameter in time-stepping scheme; parameter in a pricing model
θ_t	Radon-Nikodym process
λ	market price of risk; also in use as eigenvalue
λ_t	market price of risk process

μ	relative drift
μ_X	drift of the vector of state variables
μ_Y	drift of the vector of asset prices
ν	combination of assets used in defining a numéraire
π	price; also in use as $3.14159\cdots$
Π	partition
ρ	correlation coefficient
σ	relative volatility; also used for standard deviation
Σ	variance-covariance matrix; also used as summation symbol
$\phi(x)$	standard normal density function
$\Phi(x)$	standard normal cumulative distribution function
ϕ_t	vector of portfolio holdings at time t; lies in $\mathbb{R}^m$
$[X,X]_t$	quadratic variation process corresponding to a given semimartingale X_t
$[X,Y]_t$	quadratic covariation process corresponding to given semimartingales X_t and Y_t
$:=$	left hand side is defined as right hand side
$=:$	right hand side is defined as left hand side
$\sim$	indicates the distribution of a random variable
$f(x)\big\vert_{x=a}$	function $f(x)$ evaluated at $x=a$; alternative notation for $f(a)$
$\propto$	proportionality symbol; $f \propto g$, or $f(x) \propto g(x)$, means that there is a constant c such that $f(x) = cg(x)$ for all x
$A^\top$	transpose of the matrix A; same notation is used for transposes of vectors
$\perp$	orthogonality: $x \perp y$, for $x, y \in \mathbb{R}^n$, means $x^\top y = 0$, i.e. $\sum_{i=1}^n x_i y_i = 0$
$\operatorname{colsp} A$	linear space spanned by the columns of the matrix A
$\lvert\Pi\rvert$	mesh of a partition Π (maximum distance of points in Π)

This page intentionally left blank

Appendix E

Matlab commands

Code examples in this book are stated in Matlab. To facilitate readability of these examples for non-Matlab users, a brief explanation of the most relevant commands is given below.

E.1 General features

Square brackets are used to form vectors or matrices. Round brackets are used for function arguments and for entries of vectors or matrices. Curly brackets are used for entries in cell arrays, which are matrices of objects that may be of any type.

It is a peculiarity of Matlab that commands write their output on the screen unless the command line is terminated by a semicolon `;`. The semicolon can also be used to place several commands on the same text line, and to separate rows in specifications of matrices. Elements of rows can be separated by commas or just by spaces. Single quotes `'` are used to form strings of alphanumeric symbols, as in `'hello world'`. They are also used to indicate transpose; `A'` is the transpose of the matrix `A`. The colon `:` is used to form sequences; for instance `0:3` is the same as `[0 1 2 3]`, and `1:0.25:2` is the same as `[1 1.25 1.5 1.75 2]`.

Many operations in Matlab can be carried out elementwise on vectors and matrices. This includes logical operations; for instance, if `x` is a vector of real numbers, then `y = x>0` defines a vector `y` whose entries are 1 or 0 depending on whether the corresponding entry in `x` is positive or not.

Matlab allows subscript indexing of matrices as well as linear indexing. In the latter, the entries of the matrix are numbered consecutively, columnwise. For instance, in a matrix `A` of size 3×5, `A(1,3)` is the same as `A(7)`. Logical indexing is allowed as well; for instance the command `x(x<0)=0` replaces all negative elements of the vector `x` by zeros.

The codeword `function` at the beginning of a Matlab file indicates a function file. Function files cannot be executed as such, but they can be called from a script file. The first line of the file specifies the name of the function, inputs, and outputs. If there are multiple output variables, their names are placed between square brackets.

E.2 Specific operations and commands

`*`	matrix multiplication and matrix-vector multiplication
`.*`	element-by-element multiplication of vectors or matrices of equal size
`/`	division
`./`	element-by-element division of vectors or matrices of equal size
`^`	power
`.^`	element-by-element power; for instance `[2 3].^2` is `[4 9]`
`\`	`x = A\b` solves $Ax = b$ in case A is invertible, and otherwise provides the least-squares solution; the operation can therefore also be used for linear regression
`@`	function specification; for instance `f = @(x) x.^2 + 1` defines `f` as the function $f(x) = x^2 + 1$, allowing this function to operate elementwise on vectors
`%`	comment
`~`	placeholder for unused output arguments
`axis`	define boundaries for horizontal and vertical axes in a plot; `axis square` makes the lengths of both axes equal
`cov(A)`	variance-covariance matrix of columns in matrix `A`
`diag(x)`	diagonal matrix with diagonal given by vector `x`
`diag(x,k)`	matrix of which the k-th diagonal above the main diagonal is given by the entries of `x`, and which otherwise contains zeros; k can be negative
`disp(str)`	print the string `str` to the screen
`eig(A)`	compute the eigenvalues of the matrix `A`; the form `[V,D]=eig(A)` computes also the eigenvectors, which are placed in the matrix `V`
`end`	final index of a vector
`eye(n)`	unit matrix of size $n \times n$
`length(x)`	length of the vector `x`
`linspace(a,b,n)`	generate an equally-spaced grid from a to b containing n points; default is $n = 100$
`loglog`	draw a loglog plot
`max`	elementwise maximum
`mean(X)`	average value of entries of `X`
`min`	elementwise minimum
`normcdf`	standard normal cumulative distribution function
`norminv`	inverse standard normal cumulative distribution function
`normpdf`	standard normal density function
`num2str`	convert integer or floating-point number to a string
`ones(n,m)`	$n \times m$ matrix whose entries are all 1

`pause(t)`	pause execution for `t` seconds
`plot(X,Y)`	plot vector `Y` against vector `X`; can also be used with multiple data vectors for the horizontal and vertical axes, for instance `plot(X1,Y1,X2,Y2)`
`randn(n,m)`	$n \times m$ matrix whose entries are independent standard normal random numbers
`std(X)`	standard deviation of entries of `X`
`size(A)`	size of the matrix `A`, shown as a vector
`sub2ind`	linear indices corresponding to given subscript indices, for a matrix of given size; the subscript indices are to be specified as vectors of equal length
`while`	repeat loop as long as condition is fulfilled
`xlabel`	label for horizontal axis in a plot
`ylabel`	label for vertical axis in a plot
`zeros(n,m)`	$n \times m$ matrix whose entries are all 0

The commands `normcdf`, `norminv` and `normpdf` are provided by the Statistics toolbox. In cases where this toolbox is not available, the following replacements can be used:

```
normcdf = @(x) 0.5 + 0.5*erf(x/sqrt(2))
norminv = @(x) sqrt(2)*erfinv(2*x-1)
normpdf = @(x) exp(-0.5*x.^2)/sqrt(2*pi)
```

This page intentionally left blank

Appendix F

An English-Dutch dictionary of mathematical finance and insurance

For the benefit of Dutch-speaking students, this small dictionary provides translations of some technical terms that arise in mathematical finance and insurance. Not for every English term there is a standard translation; nonstandard translations are indicated by [ns].

above par boven pari (koers hoger dan nominale waarde)
adapted aangepast (proces)
agent subject, actor [ns]
annualized return rendement op jaarbasis
arbitrage id.
arbitrage opportunity arbitragemogelijkheid
asset actief, bezitting, waardedrager [ns]
assets activa
at par à pari (koers gelijk aan nominale waarde)
at the money op de rand van het uitbetalingsgebied [ns]
auction veiling
basis point basispunt (1/100 van een procentpunt)
bear market neergaande markt
below par beneden pari (koers lager dan nominale waarde)
benchmark richtpunt [ns]
bond obligatie
borrow lenen (van); vgl. *lend*
boundary condition randvoorwaarde
Brownian motion Brownse beweging
bull market opgaande markt
calculus id.
call option kooprecht [ns]
chain rule kettingregel
collateral onderpand
commodities gebruiksgoederen
complete market volledige markt
contingent claim contract met toestandsafhankelijke uitbetaling [ns]
continuous continu
continuous-time in continue tijd, met continue tijdsparameter. *Continuous-time Markov process*: Markov-proces met continue tijdsparameter
contribution bijdrage; premie (voor een pensioenfonds)
convergence convergentie
convergence in quadratic mean convergentie in kwadratisch gemiddelde
convertible converteerbaar
corporate bond bedrijfsobligatie
corporate finance ondernemingsfinanciering
coupon id.; tussentijdse rentebetaling op een obligatielening
currency munt, geldeenheid, valuta
debt servicing afbetaling van schulden
default (zn.) faillissement; (ww.) ver-

zaken, niet nakomen
defined benefit plan uitkeringsregeling
defined contribution plan premieregeling
derivative derivaat, afgeleid instrument; afgeleide (van een functie)
difference equation differentievergelijking
differential equation differentiaalvergelijking
discount (zn.) korting, afslag; (ww.) verdisconteren. *Sell at a discount*: verkopen voor een prijs die lager ligt dan een gegeven standaardwaarde
discount rate verdisconteringsvoet
drift trend [ns]
endowment dotatie, ter beschikking gesteld kapitaal
endowment fund vermogensfonds
equation vergelijking
equilibrium evenwicht
equity aandelen
exchange rate wisselkoers
expectation verwachting (van een stochast)
expiry afloop (van een contract)
exposure gevoeligheid (voor een risicofactor)
filtration filtratie
finance financiering
finite difference method eindige differentiemethode
finite element method eindige elementenmethode
fixed-income market obligatiemarkt
funding ratio dekkingsgraad
future termijncontract
futures market termijnmarkt
gilt Britse staatsobligatie
government bond staatsobligatie
growth rate groeivoet
hedge afdekken (van risico's)
implied volatility geïmpliceerde volatiliteit
in the money in het uitbetalingsgebied [ns]
incomplete market onvolledige markt
increment id.; toename
independent onafhankelijk
inequality ongelijkheid
inflation inflatie
infinity oneindig
initial condition beginvoorwaarde
integer (zn.) geheel getal
interest rente
interest rate rentevoet
kurtosis id., dikstaartigheidsindex
law of large numbers wet van de grote aantallen
lend lenen (aan); vgl. *borrow*
liability verplichting
liabilities passiva, verplichtingen
long lang, aan de positieve balanszijde. *The party that is long the contract*: de houd(st)er van het contract
Markov chain Markovketen
Markovian process Markovproces, proces met de Markov-eigenschap
martingale martingaal
martingale measure martingaalmaat
mature (ww.) aflopen (van een contract)
maturity looptijd, eindvervaldag
median mediaan
moneyness uitbetalingspositie [ns]
mortgage hypotheek
nonlinear niet-lineair
nonnegative niet-negatief
normal distribution normale verdeling
number getal; nummer. *Real number*: reëel getal
numéraire id.
occupational fund bedrijfstakpensioenfonds
optimize optimaliseren
option optie
orthogonal orthogonaal, loodrecht; x *is orthogonal to* y: x staat loodrecht op y
out of the money buiten het uitbetalingsgebied [ns]
pay-as-you-go system omslagstelsel
payoff uitbetaling
pension plan pensioenregeling
policy ladder staffel
portfolio portefeuille
premium premie; opslag, toeslag. *Sell at a premium*: verkopen voor een prijs die hoger ligt dan een gegeven standaardwaarde
probability kans
probability distribution kansverdeling
price (zn.) prijs; (ww.) prijzen, waarderen
pricing waardering

pricing kernel weegfactor [ns]
principal hoofdsom (bij een lening)
put option verkooprecht [ns]
quadratic variation kwadratische variatie
random number toevalsgetal, toevalsgrootheid, stochastische grootheid
random variable stochastische variabele, stochast
rebalance herschikken (van een portefeuille)
redeem aflossen
redemption aflossing
reinsurance herverzekering
replication replicatie
return rendement
risk risico
riskless risicovrij, risicoloos
risk-neutral risiconeutraal
risky risicodragend; riskant
satiated verzadigd
security vermogenstitel
securitize verhandelbaar maken [ns]
self-financing zichzelf financierend [ns]
series reeks
series expansion reeksontwikkeling
short kort, aan de negatieve balanszijde. *The party that is short the contract*: de verlener van het contract.
single-period model één-periodemodel
skewness scheefheid
smile opkrullend verloop (van de geïmpliceerde volatiliteit als functie van de uitoefenprijs) [ns]
smirk scheef verloop (van de geïmpliceerde volatiliteit als functie van de uitoefenprijs) [ns]
solvency solvabiliteit
solvent solvabel
sovereign staatsobligatie
state toestand
state space toestandsruimte
stationary stationair
stock aandeel
strike uitoefenprijs
swap ruil, i.h.b. van betaling van vaste of variabele rente; renteruilovereenkomst [ns]
swaption renteruiloptie [ns]
tax code belastingwetgeving
test toets
term structure termijnstructuur
terminal condition eindvoorwaarde
time axis tijdas
time to maturity resterende looptijd
Treasury bill / note Amerikaanse staatsobligatie
underfunded in onderdekking
underlying onderliggende
unit of currency geldeenheid, munt
utility function nutsfunctie
value-weighted naar waarde gewogen
volatility volatiliteit
yield opbrengst, rendement, rente
yield curve rentekromme

This page intentionally left blank

Subject Index

Name Index

Johannes M. (Hans) Schumacher studied Mathematics at the Vrije Universiteit in Amsterdam and obtained the PhD degree from the same institution in 1981. He spent a postdoc year at the Laboratory for Information and Decision Sciences of MIT. After stints at the Department of Econometrics of Erasmus University, Rotterdam, and at the European Space Agency's research center ESTEC in Noordwijk, the Netherlands, he joined the Centre for Mathematics and Computer Science (CWI) in Amsterdam in 1984. As of 1987, he combined work at CWI with a part-time position as Professor of Mathematics in the Department of Econometrics and Operations Research of Tilburg University. He moved to Tilburg University on a full-time basis in 1999. Hans Schumacher has made contributions to mathematical system and control theory, to the theory of nonsmooth dynamical systems (in particular complementarity systems), and to mathematical finance and insurance. After his retirement in 2016, he remains active in research and holds a visiting position at the Amsterdam School of Economics, University of Amsterdam.